Martin Cripps; Godfrey Keller und Sven Rady: Strategic Experimentation with Exponential Bandits

Munich Discussion Paper No. 2003-2

Department of Economics
University of Munich

Volkswirtschaftliche Fakultät
Ludwig-Maximilians-Universität München

Strategic Experimentation with Exponential Bandits*

Martin Cripps† Godfrey Keller‡ Sven Rady§

This version: July 9, 2003

Abstract

This paper studies a game of strategic experimentation with two-armed bandits whose risky arm might yield a payoff only after some exponentially distributed random time. Because of free-riding, there is an inefficiently low level of experimentation in any equilibrium where the players use stationary Markovian strategies with posterior beliefs as the state variable. After characterizing the unique symmetric Markovian equilibrium of the game, which is in mixed strategies, we construct a variety of pure-strategy equilibria. There is no equilibrium where all players use simple cut-off strategies. Equilibria where players switch finitely often between the roles of experimenter and free-rider all lead to the same pattern of information acquisition; the efficiency of these equilibria depends on the way players share the burden of experimentation among them. In equilibria where players switch roles infinitely often, they can acquire an approximately efficient amount of information, but the rate at which it is acquired still remains inefficient; moreover, the expected payoff of an experimenter exhibits the novel feature that it rises as players become more pessimistic. Finally, over the range of beliefs where players use both arms a positive fraction of the time, the symmetric equilibrium is dominated by any asymmetric one in terms of aggregate payoffs.

Keywords: Strategic Experimentation, Two-Armed Bandit, Exponential Distribution, Bayesian Learning, Markov Perfect Equilibrium, Public Goods.

JEL Classification Numbers: C73, D83, H41, O32.

---

*Our thanks for very helpful discussions and suggestions are owed to Dirk Bergemann, Antoine Faure-Grimaud, Christian Gollier, Thomas Mariotti, Klaus Schmidt, Jeroen Swinkels and Achim Wambach, and to seminar participants at the IGIER Workshop in Economic Theory at Erasmus University Rotterdam, the London School of Economics, the Stanford GSB Brown Bag Lunch, the University of Munich, the University of Oxford, the University of Pennsylvania, the University of Warwick, the University of Wisconsin–Madison, the 2000 European Summer Symposium in Economic Theory in Gerzensee, the Eighth World Congress of the Econometric Society in Seattle, and the DET Workshop on Learning in Economics at the University of Copenhagen. We would like to thank the Financial Markets Group at the London School of Economics and the Studienzentrum Gerzensee for their hospitality.

†Olin School of Business, Washington University, One Brookings Drive, St. Louis MO 63130, USA.

‡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.

Introduction

In this paper we analyse a game of strategic experimentation based upon two-armed bandits with a safe arm that offers a deterministic flow payoff and a risky arm that *might* generate positive payoffs after some exponentially distributed random time. The players have replica two-armed bandits with all risky arms being of the same type (all good, or all bad), but with 'breakthroughs' occurring independently. Each player observes each other player's actions and payoffs, so information about the type of the risky arm is a public good. Thus, each player can choose to free-ride on the costly information acquisition of any other player. On the other hand, the information a player generates may encourage others to acquire more information in the future, which may counteract the temptation to free-ride.

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

With exponential bandits, news arrives only once and then resolves all uncertainty. Examples would be the occasional breakthrough in research and development, failure of some equipment or technology whose reliability is being tested, a completed research paper in a longer-term research agenda, or a crucial proof in a paper. For concreteness, we focus on a situation where this news is good, so beliefs gradually become less optimistic as long as no news arrives. This deterministic decay of the level of optimism entails that players' value functions are (closed-form) solutions to first-order differential equations. As a consequence, we are able to provide a relatively simple and tractable taxonomy of what is possible in our model of strategic experimentation.

Above all, there is of course the fundamental inefficiency of information acquisition because of free-riding. In the unique symmetric Markovian equilibrium of the game, which requires players to use time-slicing strategies that allocate a fraction of each period to either arm, the effect of free-riding is extreme insofar as the critical belief at which all players change irrevocably to the safe arm is the same as if there where only one player. This means that the total amount of experimentation with risky arms is independent of the number of players. In other words, there is no encouragement effect whereby the presence of other players encourages at least one of them to continue experimenting at beliefs somewhat more pessimistic than the single-agent cut-off belief. This effect was first analysed by Bolton and Harris (1999). Its absence in the exponential bandit framework is easy to explain. In fact, the encouragement effect rests on two conditions: the additional experimentation by one player must increase the likelihood that other players will experiment in the future, and this future experimentation must be valuable to the player who acted as a pioneer. With exponential bandits, however, the only way for this player to increase the likelihood that others will experiment is to have a breakthrough on his risky arm – but as such a breakthrough is fully revealing, he knows everything he needs to know from then on, and the additional 'experimentation' by the other players is of no value to him.

We show, however, that a different sort of encouragement is at work in all pure-strategy Markov equilibria, with players alternating between the roles of free-rider (playing safe) and pioneer (playing risky). The players generate the same *amount* of information at all pure-strategy Markovian equilibria if their strategies switch actions only a finite number of times, and this amount is again the same as the single-agent optimum. This result is driven by backward induction: with finite switching there is a last agent to engage in experimentation and this agent has no incentive to provide more information than would be optimal in the single-agent set-up. Although the amount of information acquired is constant over all pure-strategy Markovian equilibria with finite switching, the *rate* at which the information is acquired does vary. The more equitably the players share the burden of experimentation when it becomes costly (i.e., when ceasing to experiment would yield a higher short-term payoff), the longer they are able to maintain the maximal rate of information acquisition, and the more efficient is the equilibrium. The extreme equilibria where one player bears most of the costs of experimentation are the least efficient. We also show that (at least over the range of beliefs where players use both arms a positive fraction of the time) the symmetric equilibrium is less efficient than even the worst asymmetric Markovian equilibria.

Casual intuition might lead one to believe that the simplest two-agent pure-strategy equilibrium had one player ceasing to experiment when the cost of experimentation became significant and free-riding ever after. In fact no such equilibrium exists. In the simplest pure-strategy equilibrium of the two-player game, one player changes from experimentation to free-riding when beliefs hit a threshold, leaving her opponent to continue experimenting. Then, at a more pessimistic belief threshold, the two players exchange actions – the player who was experimenting free-rides and the player who was free-riding experiments – until all experimentation ceases at the lowest threshold for beliefs. Why do we observe such an equilibrium? In Markovian equilibria the players are not really choosing strategies to affect the amount of information acquired (in aggregate the same amount is always acquired) – but instead they are choosing strategies to adjust the rate at which information is acquired. The last player to experiment is obliged to do this at some cost to herself (and benefit to her opponent). Thus she is not in a hurry to find herself in this role and is willing to delay the time at which this phase of play arrives. Her opponent benefits from this phase of play and so is prepared to experiment in order to accelerate its arrival. Prior to this final phase, therefore, the player who must run the final leg is prepared to defer it by not experimenting herself, whilst the free-rider on the final leg is happy to carry the burden of the experimentation before it. Thus, somewhere between the optimistic threshold (where the number of experimenters drops from two to one) and the pessimistic threshold (where the number of experimenters drops from one to zero) there must be at least one other threshold where the players swap roles.

This simplest equilibrium can be elaborated on by many switches between the role of free-rider and experimenter. We give a complete characterization of when and how this can happen. As the players share the intermediate phase more equally the equilibrium becomes more efficient, because there is less of a temptation for either of

the players to free-ride before the last phase. Further, we indicate how these results can be generalized to the N-player case.

Our last result is to show that an approximately efficient amount of information can be acquired if we allow the players to use Markovian strategies that switch actions an infinite number of times during a finite time interval. If there is never a last period of experimentation for any player, each individual can be given an incentive to take turns in providing additional (smaller and smaller) amounts of experimentation. A level of experimentation which is approximately socially efficient can then be induced; the rate at which this information is acquired is, however, socially inefficient. In summary, while the Bolton-Harris encouragement effect is not present here, players still do encourage each other by taking turns in an incentive-compatible way.

The exponential model is a simple continuous-time analogue of the two-outcome bandit model in Rothschild (1974), the paper that started the economics literature on active Bayesian learning. Bergemann and Hege (1998, 2001) study models of financial contracting that embed an exponential bandit, but the emphasis of their analysis lies on the contractual relationship between a single experimenter and a financier, not on strategic experimentation itself. Malueg and Tsutsui (1997) analyze a model of a patent race with learning where the arrival time of the innovation is exponentially distributed given the stock of knowledge. This leads to the same structure of belief revisions as with exponential bandits, yet the nature of firms' interaction in their model is entirely different from the situation that we consider.

The paper closest to us is Bolton and Harris (1999). Their model of strategic experimentation is based upon two-armed bandits where the risky arm yields a flow payoff with Brownian noise. There, both good and bad news arrives continuously, and beliefs are continually adjusted up or down by infinitesimally small amounts. Owing to the technical complexity of the Brownian model, Bolton and Harris restrict themselves to studying symmetric equilibria. They prove existence and uniqueness, and show the presence of the encouragement effect described above. Except for this effect, the symmetric equilibrium of our game confirms all their findings in a mathematically much simpler framework. What is more, we are able to present pure-strategy asymmetric equilibria and show that they are more efficient than the symmetric one.[1]

Our paper is also related to a recent literature on the dynamic provision of public goods. Recall that an approximately efficient amount of information can be acquired if we allow the players to switch actions an infinite number of times during a finite time interval. To put this result in perspective, note that in a situation of strategic experimentation with observable actions and outcomes, the players are providing each other with a public good (information). The provision of this public good is irreversible and ultimately costly if the experiments are unsuccessful. Recent work on the dynamic provision of public goods has found that efficient provision is possible if the players make

[1]The construction of such equilibria is made easier by the absence of the encouragement effect. Bolton and Harris (2000) shut down the encouragement effect in their model by taking it to the undiscounted limit. Adding background information, they are then able to characterize all equilibria, both symmetric and asymmetric.

smaller and smaller contributions over time and there is no one player who is the last to contribute; see, for example, Admati and Perry (1991), Marx and Matthews (2000), or Lockwood and Thomas (2002). These models use (non-Markovian) trigger strategies to achieve efficiency. If a player deviates from the agreed path of contributions at any point in time, then no other player will make contributions to the public good in the future. Thus the players choose to continue to contribute to the public good because their net gain (from others' future contributions) outweighs their current cost of provision. Although our model is very different – time is continuous whereas actions are discrete – the same economic principle applies. Trigger strategies are unnecessary here because the beliefs encode the punishment. If a player does not perform an appropriate amount of experimentation, then her opponents' beliefs will not fall sufficiently for them to embark on their round of experimentation, and this hurts the deviating player.

The rest of the paper is organized as follows. Section 1 introduces the exponential bandit model. Section 2 characterizes the optimal strategy for a single player. Section 3 establishes the efficient benchmark where several players cooperate in order to maximize joint expected payoffs. Section 4 considers the strategic problem and shows that, because of free-riding, any Markov equilibrium of the game leads to inefficiently low levels of experimentation. Section 5 presents the unique symmetric Markov perfect equilibrium, which is in mixed strategies. Section 6 describes pure-strategy, and hence asymmetric, equilibria. Section 7 contains some concluding remarks. Some of the proofs are relegated to the appendix.

1 Exponential Bandits

The purpose of this section is to introduce continuous-time two-armed bandit problems where the time at which uncertainty is resolved obeys an exponential distribution. One arm S is 'safe' and yields a known deterministic flow payoff whenever it is played; the other arm R is 'risky' and can be either 'bad' or 'good'. If it is bad, then it always yields 0. If it is good, then it yields 0 until a 'breakthrough' occurs. This happens once the total time that the arm has been used reaches some random threshold that is exponentially distributed; once it happens, the arm yields lump-sum payoffs that are equivalent in expectation to a known positive flow payoff forever.

We assume that the agent strictly prefers R, if it is good, to S, and strictly prefers S to R, if it is bad, so she has a motive to experiment with the risky action in the hope of discovering that R is indeed good. The problem she faces, however, is that when she plays R she cannot immediately tell whether it is good or bad, because in either case she initially receives no payoff at all. The longer she waits without getting a payoff, the less optimistic she becomes and there will come a time when it is optimal for her to cut her losses and change irrevocably to S. Of course, if she does eventually receive a payoff then she becomes certain that R is good and she will continue with R forever.

More formally, time $t \in [0, \infty[$ is continuous, and the discount rate is $r > 0$. The flow payoff of the safe arm is s. A good risky arm produces lump-sum payoffs that arrive

according to a Poisson process with parameter λ. These lump-sums are drawn from a time-invariant distribution on $\mathbb{R}_{++}$ with mean h; in expectation, they are therefore equivalent to a constant flow payoff of $g = \lambda h$. We assume that $0 < s < g$. Note that the total time that a good risky arm must be used before it generates the first lump-sum payoff (the 'breakthrough') is exponentially distributed with parameter $\lambda > 0$, that is, with mean $1/\lambda$.

If k_t indicates the agent's choice at time t between S ($k_t = 0$) and R ($k_t = 1$), then her expected flow payoff at that time is $(1 - k_t)s + k_t\gamma$ with $\gamma \in \{0, g\}$. Starting with a prior belief p_0 that the risky arm is good, her overall objective is to choose a strategy $\{k_t\}_{t \geq 0}$ that maximizes

$$\mathrm{E}\left[\int_0^\infty r\, e^{-r\,t}\ [(1 - k_t)s + k_t\gamma]\ dt \,\middle|\, p_0\right],$$

which expresses the total payoff in per-period terms. Of course, this choice of strategy is subject to the constraint that the action taken at any time t be measurable with respect to the information available at that time.

Let p_t denote the subjective probability at time t that the agent assigns to the risky arm being good, so that gp_t is her current expectation of the flow payoff of R. By the Law of Iterated Expectations, we can rewrite the above payoff as

$$\mathrm{E}\left[\int_0^\infty r\, e^{-r\,t}\ [(1 - k_t)s + k_t g p_t]\ dt \,\middle|\, p_0\right].$$

This highlights the potential for beliefs to serve as a state variable.

Were an agent with current belief $p_t = p$ to act myopically, she would weigh the short-run payoff from playing the safe arm, s, against what she expects from playing the risky arm, gp. So the belief that makes her indifferent between these choices is $p^m = s/g$. For $p > p^m$ it is myopically optimal to play R; for $p < p^m$ it is myopically optimal to play S. As we shall see below, a forward-looking agent (who values information) continues to play R for some beliefs $p < p^m$, and is said to *experiment*.

We shall consider the cases where there is a single agent, where there are N agents playing cooperatively, and where there are N players who act strategically but use only Markovian strategies with the state variable being the belief p.

2 The Single-Agent Problem

When S is played over a period of time dt, the belief does not change. When R is played over a period of time dt, a breakthrough occurs with probability $\lambda\, dt$ if the risky arm is good,[2] and the posterior belief jumps to 1; no breakthrough occurs with probability $1 - \lambda\, dt$ if the risky arm is good, and with probability 1 if the risky arm is bad. If the agent starts with the belief p, plays R over a period of time dt and does

[2]This is up to terms of the order $o(dt)$, which we can ignore here and in what follows.

not achieve a breakthrough, then the updated belief at the end of that time period is

$$p + dp = \frac{p\,(1 - \lambda\,dt)}{1 - p + p\,(1 - \lambda\,dt)}$$

by Bayes' rule. Simplifying, we see that the belief changes by

$$dp = -\lambda p(1-p)\,dt\,.$$

We now derive the agent's Bellman equation. By the Principle of Optimality, the agent's value function satisfies

$$u(p) = \max_{k\in\{0,1\}} \left\{ r\left[(1-k)s + kgp\right]\,dt + e^{-r\,dt}\mathrm{E}\left[u(p+dp)\mid p\right]\right\}$$

where the first term is the expected current payoff and the second term is the discounted expected continuation payoff.

As to the expected continuation payoff, with subjective probability $pk\lambda\,dt$ a breakthrough occurs and the agent expects a flow payoff of g in the future; with probability $p(1-k\lambda\,dt) + (1-p) = 1 - pk\lambda\,dt$ there is no breakthrough and she expects $u(p) + u'(p)dp = u(p) - k\lambda p(1-p)u'(p)\,dt$.[3]

Using $1 - r\,dt$ to approximate $e^{-r\,dt}$, we see that her discounted expected continuation payoff is

$$(1 - r\,dt)\,\left\{u(p) + k\lambda p[g - u(p) - (1-p)u'(p)]\,dt\right\}$$

and so her expected total payoff is

$$u(p) + r\left\{(1-k)s + kgp + k\lambda p[g - u(p) - (1-p)u'(p)]/r - u(p)\right\}dt\,.$$

When this is maximized it equals $u(p)$. Simplifying and rearranging, we thus obtain the Bellman equation

$$u(p) = \max_{k\in\{0,1\}} \left\{(1-k)s + kgp + k\lambda p[g - u(p) - (1-p)u'(p)]/r\right\}\,.$$

Note that the maximand is linear in k, and the equation can be rewritten more succinctly as

$$u(p) = s + \max_{k\in\{0,1\}} k\left\{b(p,u) - c(p)\right\}\,,$$

where

$$c(p) = s - gp$$

and

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

[3]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 single-agent and cooperative cases, we will indeed see equilibria of the strategic experimentation game where a player's payoff function is not of class C^1.

Clearly, $c(p)$ is the opportunity cost of playing R; the other term, $b(p,u)$, is the (discounted) expected benefit of playing R, and has two parts: $\lambda p[g-u(p)]$ is the expected value of the jump to $u(1)=g$ should a breakthrough occur; $-\lambda p(1-p)u'(p)$ is the negative effect on the overall payoff should no breakthrough occur. The agent is indifferent between the two options when cost equals expected benefit, each option resulting in $u(p)=s$. Thus she is effectively unrestricted by the discrete nature of her choice; as usual, there is no scope for randomization in this single-agent decision problem.

So, when it is optimal to play S ($k^*=0$), $u(p)=s$ as one would expect; and when it is optimal to play R ($k^*=1$), u satisfies the first-order ODE

$$\lambda p(1-p)u'(p)+(r+\lambda p)u(p)=(r+\lambda)gp, \tag{1}$$

which has the solution

$$V_1(p)=gp+C\,(1-p)\,\Omega(p)^{\mu} \tag{2}$$

where

$$\Omega(p)=(1-p)/p \quad \text{and} \quad \mu=r/\lambda.$$

$\Omega(p)$ denotes the odds ratio at the belief p, and μ highlights the interplay between the discount rate and the expected delay before a breakthrough occurs. The first term, gp, is the expected payoff from committing to the risky arm, while the second term (the solution to the homogeneous equation) captures the option value of being able to change to the safe arm. At sufficiently optimistic beliefs, this option value is positive, implying a positive constant of integration C and a convex solution V_1.

Proposition 2.1 (Single-agent optimum) *In the single-agent problem, there is a cut-off belief p_1^* given by*

$$p_1^*=\frac{\mu s}{(\mu+1)(g-s)+\mu s}<p^m \tag{3}$$

such that below the cut-off it is optimal to play S and above it is optimal to play R. The value function V_1^* for the single-agent is given by

$$V_1^*(p)=gp+(s-gp_1^*)\left(\frac{1-p}{1-p_1^*}\right)\left(\frac{\Omega(p)}{\Omega(p_1^*)}\right)^{\mu} \tag{4}$$

when $p>p_1^*$, and $V_1^*(p)=s$ otherwise.

Proof: The expression for p_1^* and the constant of integration in (4) are obtained by imposing $V_1^*(p_1^*)=s$ (value matching) and $(V_1^*)'(p_1^*)=0$ (smooth pasting). Optimality follows by standard verification arguments. ■

The value function for a single agent is illustrated in Figure 1 – it is the lower of the two curves. (The solid kinked line is the payoff from the myopic strategy; the upper curve is relevant for the next section.) Note that an individual agent can never be forced to accept a worse payoff, since any player can always act unilaterally (see Lemma 4.2).

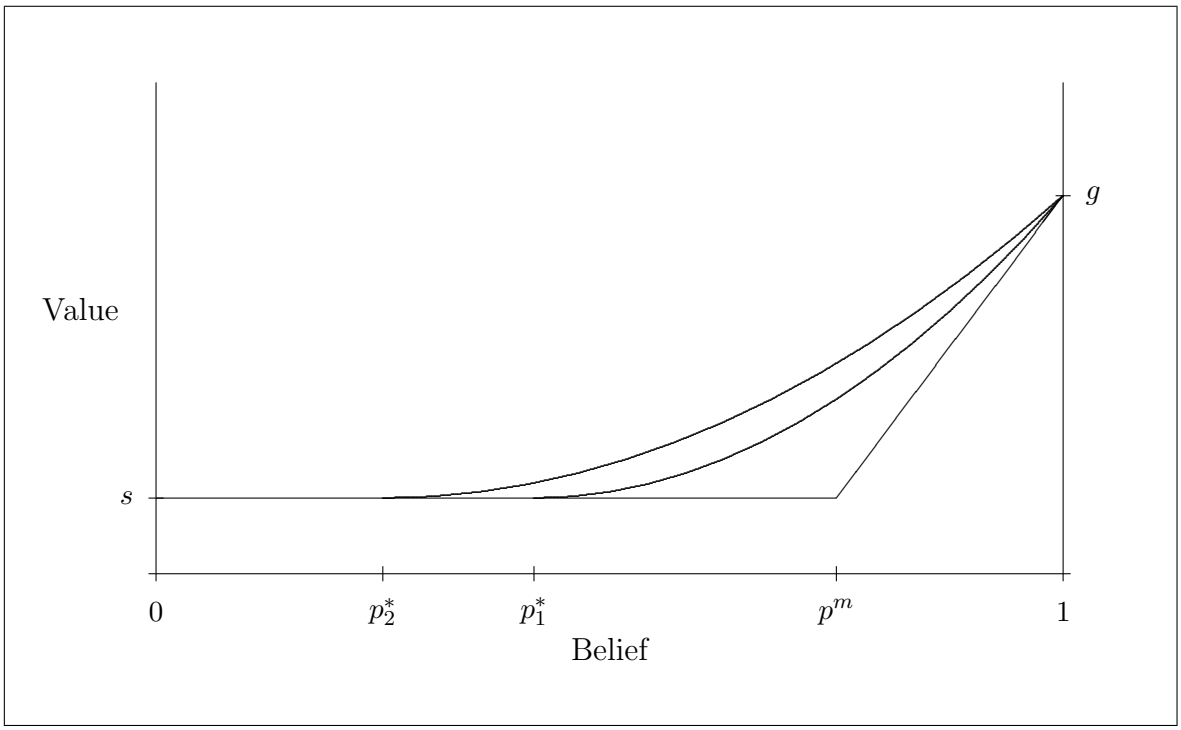

Figure 1: Payoffs for a myopic agent, a single agent, and two agents cooperating

This solution exhibits all of the familiar properties which were elegantly described in Rothschild (1974): the optimal strategy has a cut-off where the experimenter changes irrevocably from R to S; there are occasions where the experimenter makes 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 experimenter becomes more patient, and as the expected reward from the risky action increases.

3 The N-Agent Cooperative Problem

Now suppose that there are $N \geq 2$ identical agents (same prior belief, same discount rate), each with a replica two-armed bandit (same safe payoff, same flow payoff arriving according to independent and identical exponential distributions), who work cooperatively, i.e. want to maximize the *average* expected payoff. Information is public: the players can observe each other's actions and outcomes, so the players hold a common belief at each moment of time.

If K of them play R over a period of time dt, then, if a breakthrough occurs they all change to R else their belief decays K-times as fast. Whenever the risky arm is good, the probability of *none* of them achieving a breakthrough is $(1-\lambda\,dt)^K = 1-K\lambda\,dt$, the probability of *exactly one* of them achieving a breakthrough is $K\lambda\,dt(1-\lambda\,dt)^{K-1} =$

$K\lambda\,dt$, and the probability of *more than one* of them achieving a breakthrough is negligible.[4]

Lemma 3.1 *In the N-agent cooperative problem, it is optimal for* all *players to play R or for* none *of them to do so.*

PROOF: Let u be the value function for the cooperative problem, expressed as average payoff per player. When the current belief is p and the current choice is for K agents to play R, the average expected current payoff is $r\left[(1-\frac{K}{N})s+\frac{K}{N}gp\right]\,dt$. Paralleling the calculation for the single-agent problem, we see that the discounted expected continuation payoff is

$$(1-r\,dt)\,\{u(p)+K\lambda p[g-u(p)-(1-p)u'(p)]\,dt\}$$

and so the average expected total payoff is

$$u(p)+r\left\{(1-\tfrac{K}{N})s+\tfrac{K}{N}gp+K\lambda p[g-u(p)-(1-p)u'(p)]/r-u(p)\right\}dt\,.$$

Thus the value function satisfies the Bellman equation

$$u(p)=\max_{K\in\{0,1,\ldots,N\}}\left\{(1-\tfrac{K}{N})s+\tfrac{K}{N}gp+K\lambda p[g-u(p)-(1-p)u'(p)]/r\right\}\,,$$

or equivalently

$$u(p)=s+\max_{K\in\{0,1,\ldots,N\}}K\left\{b(p,u)-c(p)/N\right\}.$$

Once again, the maximand is linear in K, and the cooperative is indifferent between all levels of K when $c(p)/N$, the *shared* opportunity cost of playing R, equals $b(p,u)$, the *full* expected benefit, each of them resulting in $u(p)=s$. Thus at all beliefs $K^*=N$ or $K^*=0$ is optimal. ■

So, when it is optimal for them all to play S, $u(p)=s$ as usual; and when it is optimal for them all to play R, u satisfies

$$N\lambda p(1-p)u'(p)+(r+N\lambda p)u(p)=(r+N\lambda)gp \tag{5}$$

which is like equation (1) with λ replaced by $N\lambda$ (reflecting an N-times faster rate of information acquisition). This has the solution

$$V_N(p)=gp+C\,(1-p)\,\Omega(p)^{\mu/N}. \tag{6}$$

Proposition 3.1 (Cooperative solution) *In the N-agent cooperative problem, there is a cut-off belief p_N^* given by*

$$p_N^*=\frac{\mu s}{(\mu+N)(g-s)+\mu s}<p_1^* \tag{7}$$

[4]Again, we are ignoring terms of order $o(dt)$.

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

$$V_N^*(p) = gp + (s - gp_N^*)\left(\frac{1-p}{1-p_N^*}\right)\left(\frac{\Omega(p)}{\Omega(p_N^*)}\right)^{\mu/N} \tag{8}$$

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

PROOF: As Lemma 3.1 reduces the cooperative problem to a binary choice, the same arguments as in the proof of Proposition 2.1 apply. ■

The value function for either member of a two-agent cooperative is illustrated in Figure 1 – it is the upper of the two curves. The cut-off belief p_N^* is increasing in μ and decreasing in N, and it is straightforward to show that each player's payoff $V_N^*(p)$ increases in N over the range of beliefs where playing the risky arm is optimal.

Note that the *average* payoff of the players in *any* N-player problem can never be higher than $V_N^*(p)$, since the cooperative can always replicate their strategies (see Lemma 4.2). In this sense, the above proposition determines the *efficient* experimentation strategies for N players. More precisely, we can distinguish two aspects of efficiency here. Given a strategy profile $\{(k_{1,t}, \ldots, k_{N,t})\}_{t\geq 0}$ for the N players, the sum $K_t = \sum_{n=1}^N k_{n,t}$ measures how many risky arms are used at a given time t. We will call this number the *intensity* of experimentation. On the other hand, the integral $\int_0^\infty K_t\, dt$ measures how much the risky arms are used overall. We will call this number the *amount* of experimentation that is performed.

In fact, the amount of experimentation depends only on the initial belief and the belief at which all experimentation ceases, and is independent of the intensity.

Lemma 3.2 *If all experimentation ceases when the common belief decays to $p_c < p_0$, then the amount of experimentation performed is $\left(\ln \Omega(p_c) - \ln \Omega(p_0)\right)/\lambda$.*

PROOF: With an intensity of experimentation K_t, the change in the belief is given by $dp = -K_t \lambda p(1-p)\, dt$ when no success occurs. Thus

$$\int_0^\infty K_t\, dt = -\frac{1}{\lambda}\int_{p_0}^{p_c} \frac{dp}{p(1-p)} = \frac{1}{\lambda}\Big[\ln \Omega(p)\Big]_{p_0}^{p_c}.$$

■

With this lemma, Proposition 3.1 implies that the efficient amount of experimentation is $\left(\ln \Omega(p_N^*) - \ln \Omega(p_0)\right)/\lambda$. 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, the efficient intensity is maximal at early stages, and minimal later on.

As we shall see next, Markov equilibria of the N-player *strategic* problem are never efficient. Although it is possible to approach the efficient amount of experimentation in such an equilibrium, the intensity of experimentation will always be inefficient because of each player's incentive to free-ride on the efforts of the others.

4 The N-Player Strategic Problem

We continue to assume that the players have the same prior belief, the same discount rate, replica two-armed bandits, and that information is public. We consider stationary Markovian pure strategies with the common belief as the state variable.

Let $k_n \in \{0,1\}$ indicate the current choice of player n between S ($k_n = 0$) and R ($k_n = 1$); let $K = \sum_{n=1}^{N} k_n$ and $K_{\neg n} = K - k_n$, so that $K_{\neg n}$ summarizes the current choices of the other players. Taking into account the information generated if the other players play R, we see that player n's value function satisfies the Bellman equation

$$u_n(p) = \max_{k_n \in \{0,1\}} \left\{(1-k_n)s + k_n gp + (k_n + K_{\neg n})\lambda p[g - u_n(p) - (1-p)u_n'(p)]/r\right\} ,$$

or in terms of opportunity cost and expected benefit

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

Immediately we see that the best response, $k_n^*(p)$, is determined by comparing the opportunity cost of playing R with the expected *private* benefit:

$$k_n^*(p) \begin{cases} = 0 & \text{if } c(p) > b(p, u_n), \\ \in \{0,1\} & \text{if } c(p) = b(p, u_n), \\ = 1 & \text{if } c(p) < b(p, u_n). \end{cases} \tag{9}$$

If the best response is to play R ($k_n^* = 1$) then player n's value function u_n satisfies

$$K\lambda p(1-p)u'(p) + (r + K\lambda p)u(p) = (r + K\lambda)gp \tag{10}$$

with $K = K_{\neg n} + 1$.[5] The solution to (10) is

$$V_K(p) = gp + C\,(1-p)\,\Omega(p)^{\mu/K}. \tag{11}$$

If the best response is to free-ride by playing S ($k_n^* = 0$) then u_n satisfies

$$K\lambda p(1-p)u'(p) + (r + K\lambda p)u(p) = rs + K\lambda gp \tag{12}$$

with $K = K_{\neg n}$. The solution to (12) is

$$F_K(p) = s + \frac{K(g-s)}{\mu + K}p + C\,(1-p)\,\Omega(p)^{\mu/K}. \tag{13}$$

We note for future reference that when everyone is playing pure strategies, the average payoff satisfies

$$u(p) = s + Kb(p,u) - \tfrac{K}{N}c(p) \tag{14}$$

[5]Note that equation (10) for the strategic problem is the same ODE as that for the cooperative problem with K players; cf. equation (5). This is because the arrival probability of a breakthrough is the same in both situations.

whenever K players are experimenting and the remaining $N-K$ players are free-riding. (This corresponds to exactly K members of an N-player cooperative experimenting, and so (14) follows directly from the developments in the previous section.) Consequently, u satisfies a convex combination of (10) and (12) with weightings K/N and $(N-K)/N$, and the solution is the corresponding combination of (11) and (13).

Finally, using the indifference condition from (9) to substitute $c(p)$ for $b(p,u_n)$ in the Bellman equation, we see that for $K_{\neg n}>0$, player n is indifferent if and only if $u_n(p)=s+K_{\neg n}(s-gp)$. Note that

$$\mathcal{D}_K := \{(p,u)\in[0,1]\times I\!R_+ :\ u = s+K\,(s-gp)\}$$

is a diagonal line in the (p,u)-plane which cuts the safe payoff line $u=s$ at $p=p^m$, the myopic cut-off. If the graphs of $F_{K_{\neg n}}$ and $V_{K_{\neg n}+1}$ meet $\mathcal{D}_{K_{\neg n}}$ at the same belief p_c then $F'_{K_{\neg n}}(p_c)=V'_{K_{\neg n}+1}(p_c)$, which is a manifestation of the usual smooth-pasting property. The role of the diagonals becomes apparent in the following result, which analyses each player's best-response correspondence over the relevant range of pairs of beliefs and continuation payoffs.

Lemma 4.1 *Assume that K players play R and $N-K-1$ play S for all beliefs in an interval $]p_\ell,p_r]$ with $p_\ell<p_r$. Let the remaining player, n, have a continuation payoff of $u_n\geq V_1^*(p_\ell)$ at the belief p_ℓ. There is an interval $]p_\ell,p_\ell+\varepsilon]\subseteq]p_\ell,p_r]$ such that if (p_ℓ,u_n) lies above $\mathcal{D}_K$, then player n's best response on $]p_\ell,p_\ell+\varepsilon]$ is R; whereas if (p_ℓ,u_n) lies below $\mathcal{D}_K$, then player n's best response on $]p_\ell,p_\ell+\varepsilon]$ is S.*

PROOF: See the Appendix. ■

Note that this result holds even for $K=0$, $\mathcal{D}_0$ being the line $u=s$. The result is illustrated for $N=2$ in Figure 2 where the solid kinked line is the payoff from the myopic strategy, and the solid curve the payoff from the single-agent optimal strategy.

We now derive some bounds on the players' payoffs in *any* Markov perfect equilibrium.

Lemma 4.2 *In any Markov perfect equilibrium of the N-player strategic game, the average payoff can never exceed V_N^* and no individual payoff can fall below V_1^*.*

PROOF: The upper bound follows immediately from the fact that the cooperative solution maximizes the average payoff. As to the lower bound, we know that $V_1^*(p)=s+\max\{b(p,V_1^*)-c(p),0\}$ with $b(p,V_1^*)>0$, that player n's payoff satisfies $u_n(p)=s+K_{\neg n}\,b(p,u_n)+\max\{b(p,u_n)-c(p),0\}$, and that in any equilibrium $u_n(1)=V_1^*(1)=g$ and $u_n(0)=V_1^*(0)=s$. If u_n were to fall below V_1^*, there would have to be some belief p' such that $u_n(p')<V_1^*(p')$ and $u_n'(p')\leq(V_1^*)'(p')$, implying that $b(p',u_n)>b(p',V_1^*)$. Thus we would obtain the chain of inequalities

$$\begin{aligned}\max\{b(p',V_1^*)-c(p'),0\} &> K_{\neg n}\,b(p',u_n)+\max\{b(p',u_n)-c(p'),0\}\\ &\geq K_{\neg n}\,b(p',V_1^*)+\max\{b(p',V_1^*)-c(p'),0\}\\ &\geq \max\{b(p',V_1^*)-c(p'),0\},\end{aligned}$$

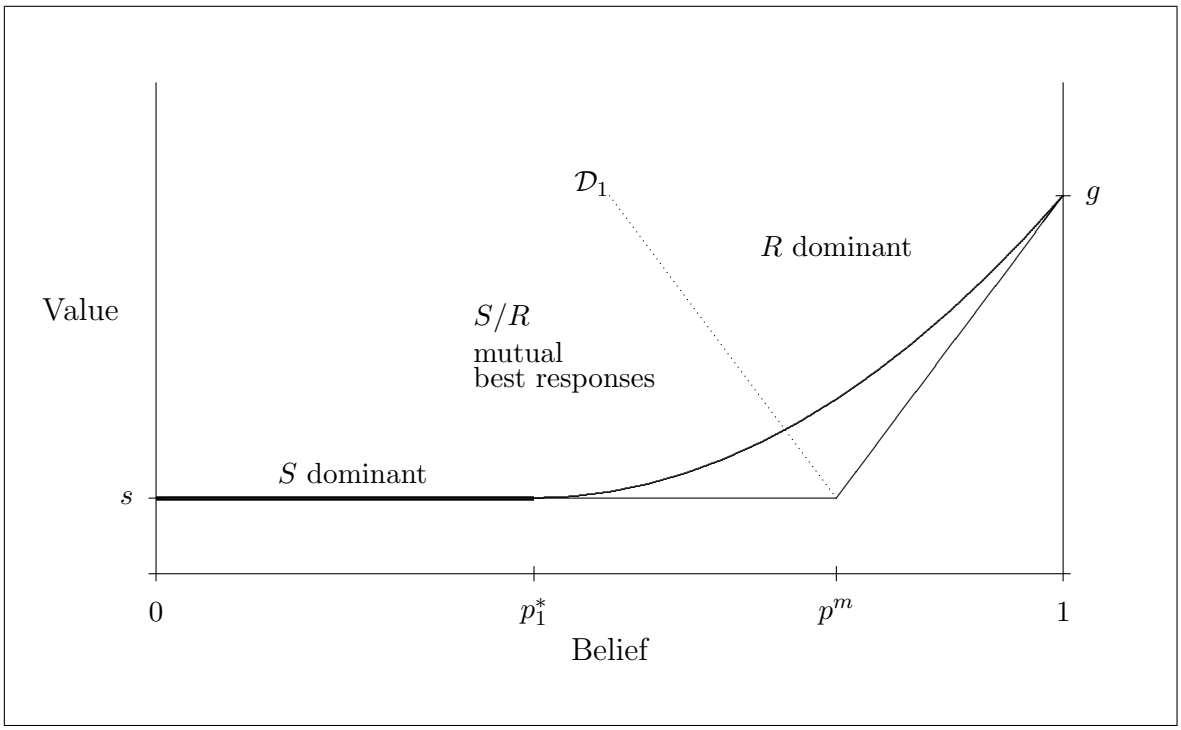

Figure 2: Characterization of best responses in the two-player case

which is a contradiction. ■

We can use the above results to show the following.

Proposition 4.1 (Inefficiency) *All Markov perfect equilibria of the N-player strategic game are inefficient.*

PROOF: Since the average payoff lies between V_1^* and V_N^* (Lemma 4.2), there must be some belief greater than p_N^* where the payoff of each player is below $\mathcal{D}_{N-1}$ and above $\mathcal{D}_0$, in which case the efficient strategies from Proposition 3.1 are *not* best responses (Lemma 4.1). ■

The intuition for this result is simple. Along the efficient experimentation path, the benefit of an additional experiment, $b(p, V_N^*)$, tends to $1/N$ of its opportunity cost, $c(p)/N$, as p approaches p_N^*. From the perspective of a self-interested player, therefore, the benefit of an additional experiment drops below the full opportunity cost, $c(p)$, so it becomes optimal to deviate from the efficient path by using S instead of R. Thus, the incentive to free-ride on the experimentation efforts of the other players makes it impossible to reach efficiency.

In the following two sections we turn to the question as to what *can* be achieved in Markov perfect equilibria. We shall consider symmetric mixed-strategy equilibria and asymmetric pure-strategy equilibria of the N-player game.

5 Symmetric Equilibrium

Since the efficient strategy profile is symmetric and Markovian with the belief as state variable, it is natural to ask what outcomes can be achieved in symmetric Markovian equilibria of the N-player game. We maintain the assumptions of the previous sections.

It is clear from Lemma 4.1 that there is no symmetric equilibrium in pure strategies. We therefore need to allow agents to use mixed strategies. Following Bolton and Harris (1999), we actually consider the time-division game in which agent n allocates a fraction κ_n of the current period $[t, t+dt[$ to R, and the remainder to S. In other words, κ_n is the derivative with respect to calendar time of the total time player n spends on R. We shall interpret this as the player using the mixed strategy that places probability κ_n on playing R, and the remainder on S.[6]

So, let $\kappa_n \in [0,1]$ indicate the current decision of player n, $K = \sum_{n=1}^{N} \kappa_n$, and $K_{\neg n} = K - \kappa_n$. Once again taking into account the information generated by the other players, we see that player n's value function satisfies the Bellman equation

$$u_n(p) = \max_{\kappa_n \in [0,1]} \left\{ (1-\kappa_n)s + \kappa_n gp + (\kappa_n + K_{\neg n})\lambda p \left[g - u_n(p) - (1-p)u_n'(p) \right] / r \right\} ,$$

or alternatively

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

Again the best response, $\kappa_n^*(p)$, is determined by comparing the opportunity cost of experimentation with the expected benefit:

$$\kappa_n^*(p) \begin{cases} = 0 & \text{if } c(p) > b(p, u_n), \\ \in [0,1] & \text{if } c(p) = b(p, u_n), \\ = 1 & \text{if } c(p) < b(p, u_n). \end{cases}$$

In any Markov perfect equilibrium of the time-division game, player n's value function will be defined piecewise: when all the time is devoted to S it satisfies equation (12) with $K = K_{\neg n}$ and is of the form $F_{K_{\neg n}}$; when all the time is devoted to R it satisfies equation (10) with $K = K_{\neg n} + 1$ and is of the form $V_{K_{\neg n}+1}$; and when the time is divided strictly between S and R it satisfies

$$\lambda p(1-p)u'(p) + \lambda p u(p) = (r+\lambda)gp - rs, \tag{15}$$

[6] One can make this interpretation mathematically precise by using a time-alternation approach to randomization in continuous time; see Harris (1993) or Keller and Rady (2003) for this approach in stochastic differential games.

which has the (strictly convex) solution

$$W(p) = s + (\mu + 1)(g - s) + \mu s(1 - p) \ln \Omega(p) + C\,(1 - p). \tag{16}$$

Note that, when $K_{\neg n} > 0$, the diagonal $\mathcal{D}_{K_{\neg n}}$ separates the region where player n uses the risky arm all the time from the region where he uses both arms a positive fraction of the time; and if the graphs of the relevant solutions W_n and $V_{K_{\neg n}+1}$ meet $\mathcal{D}_{K_{\neg n}}$ at the same belief p_c then $W_n'(p_c) = V_{K_{\neg n}+1}'(p_c)$. This implies that in a *symmetric* equilibrium, the point where the players change from using the risky arm all the time to using both arms a positive fraction of the time lies on the diagonal $\mathcal{D}_{N-1}$, and here we have smooth pasting of the relevant solutions W and V_N.

Our next result describes the unique symmetric Markov perfect equilibrium of the strategic experimentation game.

Proposition 5.1 (Symmetric equilibrium) *The N-player time-division game has a unique symmetric equilibrium in Markovian strategies with the common posterior belief as the state variable. In this equilibrium, all time is devoted to the safe arm at beliefs below the single-player cut-off p_1^*; all time is devoted to the risky arm at beliefs above a cut-off $p_N^\dagger > p_1^*$ solving*

$$(N-1)\left(\frac{1}{\Omega(p^m)} - \frac{1}{\Omega(p_N^\dagger)}\right) = (\mu+1)\left[\frac{1}{1-p_N^\dagger} - \frac{1}{1-p_1^*} - \frac{1}{\Omega(p_1^*)}\ln\left(\frac{\Omega(p_1^*)}{\Omega(p_N^\dagger)}\right)\right];$$

and a positive fraction of time is devoted to each arm at beliefs strictly between p_1^* and $p_N^\dagger$. The fraction of time that each player allocates to the risky arm at such a belief is

$$\kappa_N^\dagger(p) = \frac{1}{N-1}\,\frac{W^\dagger(p) - s}{s - gp} \tag{17}$$

with

$$W^\dagger(p) = s + \mu s\left[\Omega(p_1^*)\left(1 - \frac{1-p}{1-p_1^*}\right) - (1-p)\,\ln\left(\frac{\Omega(p_1^*)}{\Omega(p)}\right)\right], \tag{18}$$

which is each player's value function on $[p_1^*, p_N^\dagger]$ and satisfies $W^\dagger(p_1^*) = s$, $(W^\dagger)'(p_1^*) = 0$. Below p_1^* the value function equals s, and above $p_N^\dagger$ it is given by $V_N(p)$ from equation (6) with $V_N(p_N^\dagger) = W^\dagger(p_N^\dagger)$.

Proof: Consider a function W that solves (15) below and to the left of $\mathcal{D}_{N-1}$. Lemma 4.2 implies that for W to be part of a common equilibrium value function, it cannot reach the level s to the right of p_1^* (else it would fall below V_1^*), and also that it cannot stay above the level s on $[p_N^*, p_1^*]$ (else it would lie above V_N^* in some interval of beliefs). So in a symmetric equilibrium there must be a belief in $[p_N^*, p_1^*]$ where the relevant function W assumes the value s. Let p_c be the largest such belief, so that $W'(p_c) \geq 0$. By (15), this implies $p_c \geq p_1^*$, so the only possibility is that the solution to (15) has $p_c = p_1^*$, which implies $W'(p_c) = 0$.

Using $W(p_1^*) = s$ in equation (16) determines the constant of integration C, giving the expression (18) for the value function over the range where both arms are used

for a positive fraction of time. Given this function, the expression (17) for the share of time $\kappa_N^\dagger$ allocated to R follows from the Bellman equation; more precisely, we use the indifference condition to substitute $c(p)$ for $b(p, u_n)$ and then exploit the fact that $K_{\neg n} = (N-1)\kappa_N^\dagger$ by symmetry. As $W^\dagger$ is strictly convex, $\kappa_N^\dagger$ is strictly increasing to $+\infty$ as $p \uparrow p^m$. Thus there is a unique cut-off $p_N^\dagger < p^m$ where $\kappa_N^\dagger(p_N^\dagger) = 1$. Simplifying $W^\dagger(p_N^\dagger) - s = (N-1)(s - p_N^\dagger g)$ gives the equation satisfied by $p_N^\dagger$. ■

Several points are noteworthy. First, the lower cut-off belief at which all experimentation in the symmetric MPE stops does not depend on the number of players; by Lemma 3.2, this means that the same amount of experimentation is performed, no matter how many players participate. This is very strong evidence of the free-rider effect at work here.

Second, the lower cut-off equals the optimal cut-off from the single-player problem. While it is clear that experimentation in a symmetric equilibrium cannot stop at a belief above the single-agent cut-off, it is remarkable that experimentation does not extend at all to beliefs below that cut-off. As pointed out in the introduction, this means that we do not have the *encouragement effect* of Bolton and Harris (1999). In their model, an agent who on his own would be indifferent between the two actions, strictly prefers the risky action when other players are present. In fact, his own experimentation may produce favourable information that will make everybody more optimistic, and thus encourage the other players to perform some more experimentation themselves from which the first player will eventually benefit. Note that it is crucial for this argument that after a favourable experimentation outcome of one's own, there still be something to learn from other players' future outcomes. This is obviously not the case here because the first breakthrough resolves all uncertainty.

Third, the expected equilibrium payoff that each player obtains at beliefs where both arms are used a positive fraction of time does not depend on the number of players either; see equation (18). The reason for this is that the relevant ODE, equation (16), is just the indifference condition of a *single* player, and that the boundary condition at the lower cut-off belief is the same for any number of players. Put differently, the combined intensity of experimentation by $N-1$ players when both arms are used a positive fraction of time is independent of the total number of players, N; see equation (17). Over that range of beliefs, therefore, a player's best response and payoff do not depend on N either. What does depend on N is the upper cut-off belief, of course: with more players, the temptation to free-ride becomes stronger, and $p_N^\dagger$ increases. Formally, this is most easily seen from equation (17): given that $\kappa_N^\dagger(p_N^\dagger) = 1$ and $W^\dagger$ is a strictly increasing function, $p_N^\dagger$ must increase with N. Informally, the indifference diagonal $\mathcal{D}_{N-1}$ rotates clockwise as N increases. From this, it is straightforward to show that each player's payoff is strictly increasing in N over the range of beliefs where the risky arm is used all the time.

Fourth, not only is the amount of experimentation inefficiently low (as can be seen from the lower cut-off being above the cooperative cut-off, and Lemma 3.2) and the intensity of experimentation inefficiently low (at any belief between p_N^* and $p_N^\dagger$ there is strictly too little use of risky arms), but the acquisition of information is slowed down

so severely that the equilibrium amount of experimentation is not even performed in finite time – as the following result shows, the players never actually stop allocating at least some of their time to playing the risky arm.[7]

Corollary 5.1 *Starting from a prior belief above the single-agent cut-off p_1^*, the players' common posterior belief never reaches this cut-off in the symmetric Markov perfect equilibrium.*

PROOF: Close to the right of p_1^*, the dynamics of the belief p given no breakthrough are

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

(If a breakthrough occurs, the statement of the corollary is trivially true.) As $W^\dagger(p_1^*) = s$, $(W^\dagger)'(p_1^*) = 0$ and $(W^\dagger)''(p_1^*) > 0$, we can find a positive constant c such that

$$\lambda \frac{N}{N-1} \frac{W^\dagger(p) - s}{s - gp} p(1-p) < c\,(p - p_1^*)^2$$

in a neighbourhood of p_1^*. Starting from an initial belief $p_0 > p_1^*$ in this neighbourhood, consider the dynamics

$$dp = -c\,(p - p_1^*)^2\, dt.$$

The solution of these dynamics with initial value p_0 is

$$p_t = p_1^* + \frac{1}{ct + (p_0 - p_1^*)^{-1}}.$$

Obviously, this solution does not reach p_1^* in finite time. Since the modified dynamics have a higher rate of decrease than the original ones, this result carries over to the true evolution of beliefs. ■

This is the same result as Bolton and Harris (1999) obtain for the symmetric equilibrium of their game. In both instances, the amount of time spent experimenting falls to zero quadratically at the cut-off belief where experimentation stops forever, as indicated in Figure 3. This is because the equilibrium value function reaches the level s smoothly at the cut-off belief where all experimentation stops. The following simple argument shows why there must be smooth pasting at this cut-off. Suppose we had a symmetric equilibrium where the common payoff function u hits the level s at the belief $\tilde{p}$ with slope $u'(\tilde{p}+) > 0$. At beliefs immediately to the right of $\tilde{p}$, we would then have $b(p,u) = c(p)$, implying $\lambda\tilde{p}[g-s]/r = c(\tilde{p}) + \lambda\tilde{p}(1-\tilde{p})u'(\tilde{p}+)/r > c(\tilde{p})$ by continuity. Immediately to the left of $\tilde{p}$, the fact that $u'(p) = 0$ would then imply $b(p,u) = \lambda p[g-s]/r > c(p)$, so there would be an incentive to deviate from S to R.

[7] To some readers, this phenomenon might be familiar from the production of joint research papers. Once the initial enthusiasm has waned, each co-author might spend less and less time working on the paper, without actually withdrawing completely – and the paper might never be put out of its misery.

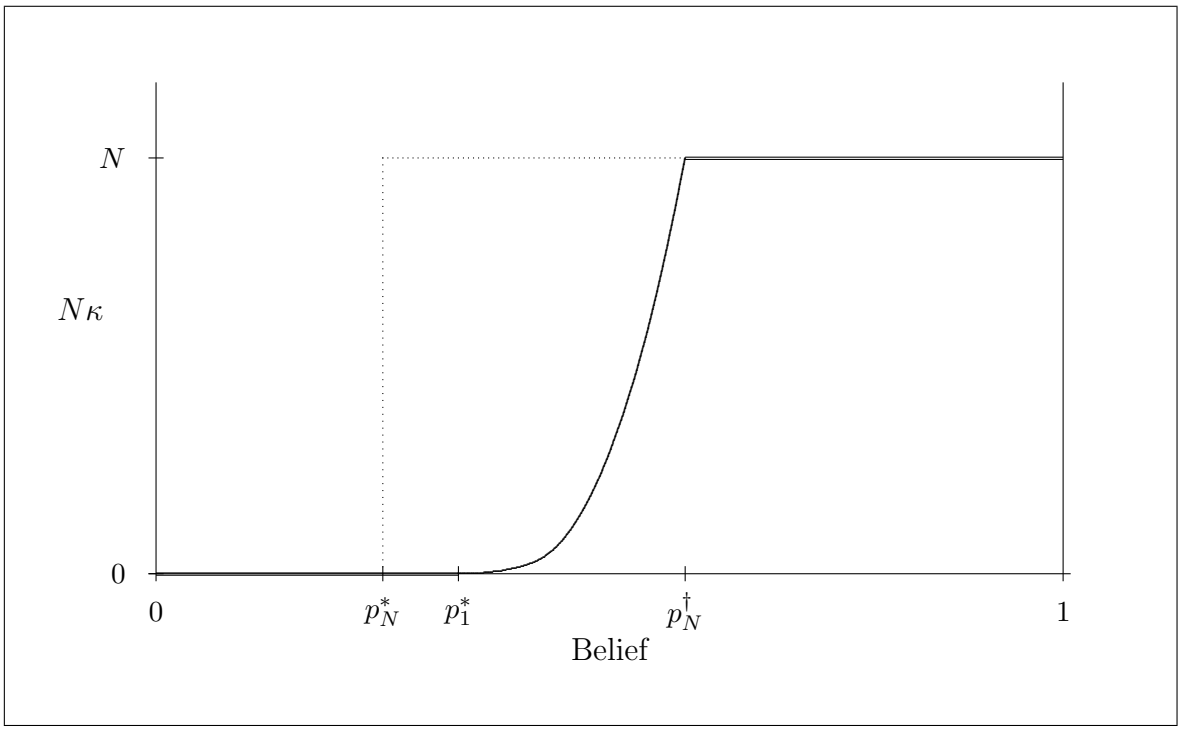

Figure 3: Total experimentation in the N-player symmetric equilibrium

6 Asymmetric Equilibria

We now turn to the behaviour that can arise in asymmetric pure-strategy Markov perfect equilibria. We will present two types of such equilibria. The first type of MPE consists of strategies where the action of each player switches at finitely many beliefs. As a consequence, there is a last point in time at which any player is willing to experiment. As in the symmetric MPE, the belief at which this happens (provided no breakthrough has occurred) will be the single-player cut-off p_1^*. So a similar inefficiency arises: both the amount and the intensity of experimentation are too low. Nevertheless, these equilibria differ in terms of the time taken to reach the belief where experimentation ceases, and also in terms of aggregate payoffs.

In the second type of MPE, each player's strategy has infinitely many switch-points, and although there is a finite time after which no player ever experiments again, no single player has a *last* time for experimentation. That is, somewhat prior to reaching a certain cut-off belief, the players switch roles after progressively smaller belief revisions, and infinitely often. We will see that we can take this cut-off belief arbitrarily closely to the efficient cut-off. Still, the equilibrium is inefficient: although an almost efficient amount of experimentation is performed, it is performed with an inefficient intensity.

6.1 Finitely many switches

For ease of exposition, we focus on the two-player case initially and then extend our results to more than two players. From Figure 2, we can see that a Markov perfect equilibrium with two players has three phases. When the players are optimistic, both play R; when they are pessimistic, both play S; in between, one of them free-rides by playing S while the other is playing R. We shall see that this mid-range of beliefs further splits into two regions: the roles of free-rider and pioneer are assigned for the whole of the upper region; in the lower region, players can swap roles.

The next proposition first describes the 'simplest' such equilibrium, in which one particular player experiments and the other free-rides throughout the lower region, and then characterizes all pure-strategy MPE where players' actions switch at finitely many beliefs.

Proposition 6.1 (Pure strategies, finite number of switches) *In the two-player strategic experimentation problem, there is a pure-strategy Markov perfect equilibrium where the players' actions depend as follows on the common posterior belief. There are two cut-offs, p_1^* and $\hat{p}_2$, and one switch-point, $\hat{p}_s$, with $p_1^* < \hat{p}_s < \hat{p}_2$ such that: on $]\hat{p}_2, 1]$, both players play R; on $]\hat{p}_s, \hat{p}_2]$, player 1 plays R and player 2 plays S; on $]p_1^*, \hat{p}_s]$, player 1 plays S and player 2 plays R; on $[0, p_1^*]$, they both play S. The low cut-off, p_1^*, is given in Proposition 2.1; the switch-point and other cut-off are given by the solution to*

$$\left(\frac{\Omega(\hat{p}_s)}{\Omega(p_1^*)}\right)^{\mu+1} + (\mu+1)\left[\frac{\Omega(\hat{p}_s)}{\Omega(p^m)} - 1\right] - 1 = 0$$

and the solution to

$$\left\{\frac{(\mu+1)(2\mu+1)}{\mu}\,\frac{\Omega(\hat{p}_s)}{\Omega(p^m)} - \frac{\mu^2+(\mu+1)(\mu+2)}{\mu}\right\}\left(\frac{\Omega(\hat{p}_2)}{\Omega(\hat{p}_s)}\right)^{\mu+1} + (\mu+1)\left[\frac{\Omega(\hat{p}_2)}{\Omega(p^m)} - 1\right] - 1 = 0.$$

Moreover, in any pure-strategy MPE with finitely many switch-points there are two cut-offs, p_1^* and $\bar{p}_2$, and one switch-point, $\bar{p}_s$, with $p_1^* < \bar{p}_s \leq \bar{p}_2$, and with $\hat{p}_s \leq \bar{p}_s$ and $\bar{p}_2 \leq \hat{p}_2$, such that: on $]\bar{p}_2, 1]$, both players play R; throughout $]\bar{p}_s, \bar{p}_2]$, one player plays R and the other plays S; on $]p_1^*, \bar{p}_s]$, the players share the burden of experimentation by taking turns; on $[0, p_1^*]$, they both play S.

Proof: Here we just sketch the proof; for details, see the Appendix.

We first note that there must be a last player to experiment since the level $u = s$ can only be reached via the part of the (p, u)-plane where R and S are mutual best responses. This player, say player 2, will necessarily stop experimenting at the single-agent cut-off belief p_1^*.

We can now work backwards (in time) from (p_1^*, s). On an interval to the right of p_1^*, player 2 plays R and his continuation value (as a function of the belief) is a slowly rising convex function. On this interval, player 1 free-rides by playing S and her continuation value is a steeply rising concave function. Thus, at some belief, player 1's

value meets $\mathcal{D}_1$ while player 2's value is still below it – this defines $\hat{p}_s$. On an interval to the right of $\hat{p}_s$, player 1 is content to be a pioneer and play R, while player 2 responds by free-riding with S. At some belief, player 2's value meets $\mathcal{D}_1$ while player 1's value is yet further above it – this defines $\hat{p}_2$. On the interval to the right of $\hat{p}_2$, both players optimally play R.

As to other equilibria of this sort, we again work backwards from (p_1^*, s). If the players swap roles (at least once) before the value of either of them has met $\mathcal{D}_1$, then the one with the higher value will be below that of player 1 in the 'simplest' equilibrium sketched above, and the one with the lower value will be above that of player 2. At some belief, the value of one of the players meets $\mathcal{D}_1$ while the other's value is still (weakly) below it – this defines $\bar{p}_s > \hat{p}_s$. The one with the higher value plays R to the right of $\bar{p}_s$, while the other one free-rides until the value meets $\mathcal{D}_1$ – this defines $\bar{p}_2 < \hat{p}_2$ – and then joins in by playing R. ■

The value functions of the two players in the 'simplest' equilibrium (with cut-offs p_1^* and $\hat{p}_2$, and switch-point $\hat{p}_s$) are illustrated in Figure 4, the faint straight line being $\mathcal{D}_1$. Observe that the higher payoff meets this line at $\hat{p}_s$ while the lower payoff meets it at $\hat{p}_2$. Note also that a player's payoff function is concave where the player is a free-rider, and convex otherwise.

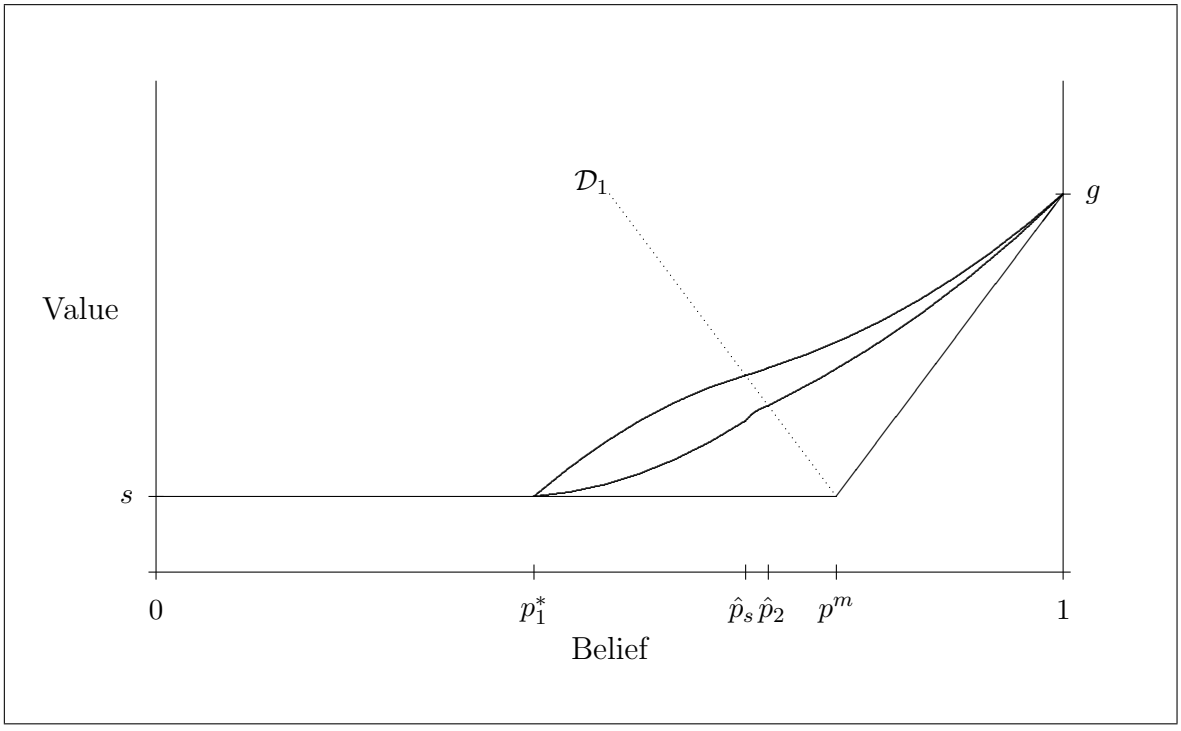

Figure 4: Payoffs in the simplest two-player asymmetric equilibrium

Before moving on to the N-player case, we note that it is not very difficult to construct equilibria in which the players take pure actions in some intervals of beliefs, and use both arms a positive fraction of the time in others. As these do not seem to lead to further insights, we do not pursue this issue further.

The N-player case

With N players, the equilibrium strategies will depend on where the players are in the (p,u)-plane, specifically where they are in terms of the diagonals $\mathcal{D}_0$, $\mathcal{D}_1$, ..., $\mathcal{D}_{N-1}$. If they are all weakly below $\mathcal{D}_0$ (i.e. $u = s$), then S is the dominant strategy, and if they are all above $\mathcal{D}_{N-1}$, then R is the dominant strategy; elsewhere we will look for mutual best responses involving some players taking the risky action and some playing safe.

If we define $N+1$ "bins" as the areas between $\mathcal{D}_{K-1}$ and $\mathcal{D}_K$ ($K = 1, \ldots, N-1$), with bin 0 being $\mathcal{D}_0$ and bin N being the area above $\mathcal{D}_{N-1}$, then mutual best responses will depend on how many players are in which bins. The size of this combinatorial task is reduced by first noting that we are seeking a relatively small number of cases: when it is the case that just one player is playing R, when two, and so on; then grouping the combinations of players-in-bins so that each grouping corresponds to a certain number playing R and the rest playing S; and finally allocating those actions to the various players.

Define areas of the plane as follows:

$$\begin{aligned} \mathcal{A}_K &:= \{(p,u) \in [0,1] \times I\!R_+ :\ u > s + K(s - gp)\} \\ \mathcal{B}_K &:= \{(p,u) \in [0,1] \times I\!R_+ :\ u \leq s + K(s - gp)\} \end{aligned}$$

so that $\mathcal{A}_K$ is the area *above* $\mathcal{D}_K$, and $\mathcal{B}_K$ is the area *below* it. The various cases are then given by:

(0) all players in $\mathcal{B}_0$;

(1) at least 1 player in $\mathcal{A}_0$, at least $N-1$ players in $\mathcal{B}_1$;

(2) at least 2 players in $\mathcal{A}_1$, at least $N-2$ players in $\mathcal{B}_2$;

$\vdots$

(K) at least K players in $\mathcal{A}_{K-1}$, at least $N-K$ players in $\mathcal{B}_K$;

$\vdots$

(N–1) at least $N-1$ players in $\mathcal{A}_{N-2}$, at least 1 player in $\mathcal{B}_{N-1}$;

(N) all players in $\mathcal{A}_{N-1}$.

To see that these cases exhaust all the possible combinations, observe that if we are not in case (0), then at least 1 player is in $\mathcal{A}_0$; if at least $N-1$ players are in $\mathcal{B}_1$, then we have case (1), else at least 2 players are in $\mathcal{A}_1$; if at least $N-2$ players are in $\mathcal{B}_2$, then we have case (2), else at least 3 players are in $\mathcal{A}_2$; and so on.

Lemma 6.1 *In case (K), the mutual best responses are such that K players in $\mathcal{A}_{K-1}$ play R, and N−K players in $\mathcal{B}_K$ play S. Moreover, when all players are playing mutual best responses any player in $\mathcal{A}_K$ plays R, any player in $\mathcal{B}_{K-1}$ plays S, and the players (if any) between $\mathcal{D}_{K-1}$ and $\mathcal{D}_K$ are assigned actions arbitrarily so that the experimental intensity is as prescribed.*

PROOF: Consider case (K).

Assume that $K-1$ players in $\mathcal{A}_{K-1}$ play R, and $N-K$ players in $\mathcal{B}_K$ play S. Then the best response of the remaining player in $\mathcal{A}_{K-1}$ is to play R, since he is above $\mathcal{D}_{K-1}$ (see Lemma 4.1).

Assume that K players in $\mathcal{A}_{K-1}$ play R, and $N-K-1$ players in $\mathcal{B}_K$ play S. Then the best response of the remaining player in $\mathcal{B}_K$ is to play S, since she is below $\mathcal{D}_K$ (see Lemma 4.1).

Since all the safe players are in $\mathcal{B}_K$, it follows that any player in $\mathcal{A}_K$ plays risky; similarly, since all the risky players are in $\mathcal{A}_{K-1}$, it follows that any player in $\mathcal{B}_{K-1}$ plays safe; any unassigned player must be between $\mathcal{D}_{K-1}$ and $\mathcal{D}_K$, and who plays which action is arbitrary as long as there are K risky players and $N-K$ safe players. ■

Generalizing the two-player situation, we work backwards (in time) from case (0), i.e., from (p_1^*, s). On an interval to the right of p_1^*, we are in case (1), where one player experiments while the rest free-ride. At some point, the value function of (at least) one player meets $\mathcal{D}_1$; that player then plays R while any player who is still below $\mathcal{D}_1$ indulges in a free-ride until two players have met $\mathcal{D}_1$. Now we are in case (2), where two players are above $\mathcal{D}_1$ and playing R while the rest are still below $\mathcal{D}_2$ and free-riding. We continue the construction in the obvious way, with more and more experimenters and fewer and fewer free-riders, until everyone is above $\mathcal{D}_{N-1}$ and playing R.

Again, there emerges a 'simplest' equilibrium. Number the players 1 through N. Whenever we enter case (K), look at each player in turn, starting with the lowest numbered, and assign him the risky action provided he is in the appropriate bin, else let him free-ride; continue until K players are assigned to R, and let the rest free-ride. (When any player crosses a diagonal, thereby moving into a different bin, we might have to reassign actions.) The beliefs at which we move from one case to an adjacent case are then as high as possible in equilibrium.

These actions are illustrated for the 'simplest' three-player equilibrium in Figure 5 (at the end of the paper), the faint dotted line showing the efficient outcome. To the right of p_1^*, in case (1), player 1 experiments while the other two free-ride until, at $\bar{p}_2$, we enter case (2), which has four regions. Now, player 1 free-rides on the experiments of the others as long as he is below $\mathcal{D}_1$; when he crosses into a different bin, actions are reassigned so that players 1 and 2 are the experimenters as long as the third player is below $\mathcal{D}_2$. When she crosses that diagonal, actions are again reassigned so that players 1 and 3 are the experimenters until player 2 also hits $\mathcal{D}_2$. Actions are reassigned for a third time, R now being dominant for both players 2 and 3, so player 1 has

another free-ride as long as he is below $\mathcal{D}_2$. At $\bar{p}_3$ we enter case (3) and all players are experimenting.

Welfare results

Note that with finitely many beliefs at which players change actions, the average payoff is determined by a decreasing sequence of cut-off beliefs $\bar{p}_N, \bar{p}_{N-1}, \ldots, \bar{p}_K, \ldots, \bar{p}_1$ at which the intensity of experimentation drops from N to $N-1$, from $N-1$ to $N-2$, and so on.[8] The cut-off belief at which all experimentation stops, $\bar{p}_1$, is again that of a single agent, namely p_1^*; in particular, it is the same for all equilibria of this type (and thus they all exhibit the same amount of experimentation – see Lemma 3.2), whereas the higher cut-off beliefs are determined endogenously by how the burden of experimentation is shared at beliefs to the right of p_1^* (and hence the intensity of experimentation will vary across these equilibria). Only for beliefs in a neighbourhood of p_1^* – specifically when fewer than two players are experimenting – is the average payoff the same across all these equilibria (see the note regarding equation (14)).

The 'simplest' equilibria described above are also the 'worst' from an efficiency perspective. This is because the cut-off beliefs at which the intensity of experimentation drops from K to K–1 are as high as they can be in equilibrium. The 'simplest' equilibria therefore exhibit the *slowest* experimentation – in equilibria where the cut-off beliefs are lower, greater intensities of experimentation are maintained for a wider range of beliefs, so the same overall amount of information is acquired faster. As the following proposition shows, such equilibria are more efficient.

Proposition 6.2 (Welfare ranking) *In terms of aggregate payoffs, the pure-strategy Markov perfect equilibria with finitely many switches can be partially ordered as follows: consider two equilibria characterized by cut-offs* $\{\bar{p}_K\}_{K=1}^N$ *and* $\{\bar{p}'_K\}_{K=1}^N$*, respectively; if* $\bar{p}_K \leq \bar{p}'_K$ *for all* K *with at least one inequality being strict, then the equilibrium with the lower cut-off(s) yields a higher aggregate payoff.*

Proof: Consider a pure-strategy MPE with cut-offs $\{\bar{p}_K\}_{K=1}^N$. For $K = 1, \ldots, N$ let $u_{N,K}$ denote the solution to the ODE (14) whenever we are in case (K), with $u_{N,1}(\bar{p}_1) = s$ and $u_{N,K}(\bar{p}_K) = u_{N,K-1}(\bar{p}_K)$ for $K = 2, \ldots, N$.

The players' average payoff function is $u_{N,K-1}$ just to the left of $\bar{p}_K$; to the right of $\bar{p}_K$ it is $u_{N,K}$. It is straightforward to verify that $u'_{N,K}(p) > u'_{N,J}(p)$ whenever $u_{N,K}(p) = u_{N,J}(p) > s$ and $K > J$, which in turn implies that if we increase $\bar{p}_K$ to $\bar{p}'_K < \bar{p}_{K+1}$, then the average payoff on $[\bar{p}_K, \bar{p}'_K]$ is now only $u_{N,K-1}$ whereas it was $u_{N,K}$, and to the right of $\bar{p}'_K$ it is still of the form $u_{N,K}$ but now lower since its value at $\bar{p}'_K$ has decreased. To the left of $\bar{p}_K$, if the cut-offs are unchanged then so is the average payoff. ■

[8] For convenience, we shall also write $\bar{p}_{N+1} = 1$.

The way to achieve a more efficient equilibrium is to move the payoff functions towards each other by sharing the burden of experimentation more equally, that is, by switching roles more often. The least upper bound on aggregate payoffs is then given by a situation of payoff symmetry where each player obtains exactly $1/N$ of the payoff of the cooperative strategy that has all N players experiment above $\mathcal{D}_{N-1}$ and $K < N$ players experiment between the diagonals $\mathcal{D}_K$ and $\mathcal{D}_{K-1}$.[9] This is the same payoff as if each player allocated exactly K/N of his time to the risky arm on the entire region between $\mathcal{D}_K$ and $\mathcal{D}_{K-1}$ (and so the players cross successive diagonals together), and hence clearly different from the payoff in the symmetric mixed-strategy equilibrium where the fraction of time allocated to the risky arm falls gradually from 1 to 0 over the region below $\mathcal{D}_{N-1}$.

In particular, there is a region of beliefs close to the single-agent cut-off where the intensity of experimentation in the symmetric equilibrium is lower than even in the 'worst' asymmetric one. By the logic of the last proposition, this ought to mean that welfare in the symmetric equilibrium should be lower at those beliefs than in any asymmetric equilibrium. The following proposition confirms this.

Proposition 6.3 (Welfare comparison with symmetric MPE) *For beliefs in the interval* $]p_1^*, p_N^\dagger]$ *the average payoff in any pure-strategy Markov perfect equilibrium with finitely many switches is strictly greater than the common payoff in the symmetric mixed-strategy equilibrium.*

Proof: See the Appendix. ■

The intuition for this result is that, at each belief in the stated range, players face a coordination problem. If this coordination problem is solved by mixing (i.e., using both arms a positive fraction of the time), there is a positive probability for players to mis-coordinate, and they do worse in aggregate.[10]

6.2 Infinitely many switches

Propositions 6.2 and 6.3 show that alternating between the roles of free-rider and pioneer as the belief changes is an effective (and incentive-compatible) way of increasing players' aggregate payoff. Players can do even better if we allow them to switch between actions at *infinitely* many beliefs. In that case, they can take turns experimenting in

[9] This least upper bound on the players' average payoff function is easy to calculate. Let $p_{N,1} = p_1^*$, and let $\bar{u}_{N,1}$ solve (14) with $\bar{u}_{N,1}(p_{N,1}) = s$; the cut-off $p_{N,2}$ is determined by the intersection of $\bar{u}_{N,1}$ with $\mathcal{D}_1$. Now let $\bar{u}_{N,2}$ solve (14) with $\bar{u}_{N,2}(p_{N,2}) = \bar{u}_{N,1}(p_{N,2})$, and so on, until we have determined $p_{N,N}$ by the intersection of $\bar{u}_{N,N-1}$ with $\mathcal{D}_{N-1}$. Finally, let $\bar{u}_{N,N}$ solve (14) with value-matching at $p_{N,N}$; the bound is the continuous function thus constructed.

[10] Note that if $p_N^\dagger < \bar{p}_N$ then the symmetric equilibrium would exhibit a higher intensity of experimentation than the asymmetric ones at beliefs close to the rightmost cut-off $\bar{p}_N$, and so the symmetric mixed-strategy equilibrium could be more efficient at beliefs above $p_N^\dagger$. Numerically, however, we find that the asymmetric equilibria are more efficient on the entire interval $]p_1^*, 1[$.

such a way that no player ever has a last time (or lowest belief) at which he is supposed to use the risky arm. Surprisingly, it is then possible to reach cut-off beliefs below p_1^* in equilibrium – in fact, it is possible to (almost) attain the efficient cut-off p_N^*, but it is still reached too slowly.

The intuition for these equilibria is that for all beliefs above the cooperative cut-off there is a Pareto gain from performing more experiments, so provided any player's immediate contributions are sufficiently small relative to the long-run Pareto gain, performing experiments in turn can be sustained as an equilibrium.

The description of mutual best responses in Lemma 6.1 implies that at beliefs close to the cooperative cut-off, any MPE must have exactly one player experimenting. The equilibria constructed below specify a sequence of intervals of beliefs where each player assumes the role of pioneer on every Nth interval. Moreover, the intervals are such that a player's expected payoff when embarking on a round of single-handed experimentation equals s. While pinning down payoffs this way simplifies the construction, other choices would work as well.

Proposition 6.4 (Pure strategies, infinite number of switches) *For each belief $p_\infty^\ddagger$ with $p_N^* < p_\infty^\ddagger < p_1^*$, the N-player strategic experimentation game admits a pure-strategy Markov perfect equilibrium with infinitely many switches where at least one player experiments as long as the current belief is above $p_\infty^\ddagger$. More precisely, there exists a strictly decreasing sequence of beliefs $\{p_i^\ddagger\}_{i=1}^\infty$ with $p_1^\ddagger \leq p_1^*$ and $\lim_{i\to\infty} p_i^\ddagger = p_\infty^\ddagger$ such that the equilibrium strategies at beliefs below $p_1^\ddagger$ can be specified as follows: player $n = 1, \ldots, N$ plays R at beliefs in the intervals $]p_{n+jN+1}^\ddagger, p_{n+jN}^\ddagger]$ $(j = 0, 1, \ldots)$ and S at all other beliefs below $p_1^\ddagger$.*

PROOF: See the Appendix. ■

The payoffs in a two-player equilibrium with an infinite number of switches are illustrated in Figure 6. (For clarity, we focus on beliefs between p_2^* and $p_1^\ddagger$.) Coming from the right, a player's value function resembles a decaying saw-tooth, with rapidly falling concave sections (for a free-rider) alternating with slowly rising convex sections (for a pioneer).

Note that if no breakthrough occurs, the limit belief $p_\infty^\ddagger$ is reached in finite time as the overall intensity of experimentation is bounded away from zero at beliefs above $p_\infty^\ddagger$. Note also that as we take $p_\infty^\ddagger$ closer to the efficient cut-off p_N^*, the amount of experimentation performed in equilibrium approaches the efficient amount; see Lemma 3.2. The intensity of experimentation, however, remains inefficient; it is 1 at beliefs in $]p_\infty^\ddagger, p_1^\ddagger]$, for instance, and therefore too low relative to the efficient benchmark.

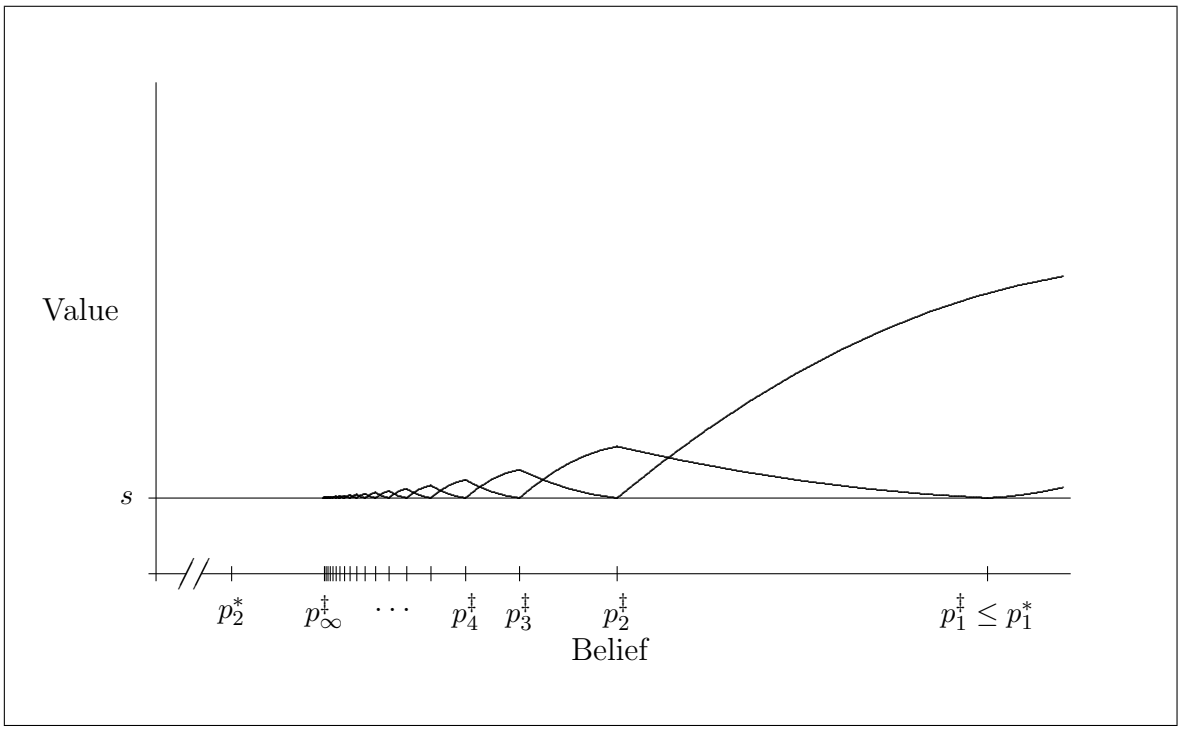

Figure 6: Payoffs in a two-player equilibrium with an infinite number of switches

7 Concluding Remarks

There are some generalizations of our results that follow with no or relatively little additional work. First, all our results apply to bandit problems where the known arm generates a stationary non-deterministic stream of payoffs – we can simply reinterpret s as the expected flow payoff. A second point is that the payoff g of an unknown arm that proves to be good has a more general interpretation as the flow-equivalent of the value of the continuation game that is reached after the first breakthrough. Further, the model can be extended to a *multi*-armed bandit with several risky arms – the best arm to use at any time is given by the familiar Gittins index rule. However, this requires more effort, as even the single-agent optimum involves a time-slicing strategy.

One extension that we are actively researching is where a single breakthrough is not fully revealing. In such a model, we would expect the encouragement effect identified by Bolton and Harris (1999) to reappear, because at least one of a group of players would have an incentive to continue experimenting at beliefs below the single-agent cut-off. Such a model could also be easily adapted to situations where an event is bad news: a 'breakdown' rather than a 'breakthrough'.

A second extension that we intend to pursue is the introduction of asymmetries between players, for example regarding the discount rate or the ability to generate

information from their experimentation effort. This may reduce the multiplicity of asymmetric equilibria that we have found for symmetric players. It may also allow us to investigate the question as to with whom a given agent would choose to play the strategic experimentation game.

More generally, we hope that exponential bandits will prove useful as building blocks for models with a richer structure. Interesting extensions in this direction could include rewards that depend on action profiles, unobservable outcomes, or costly communication.

Appendix

Proof of Lemma 4.1

First note that each player's value function is continuous as a function of p and takes the value g at $p = 1$ and s at $p = 0$; moreover it is differentiable *wherever he/she chooses optimally to change* (from playing R to playing S, or *vice versa*) and the other players do not change – if the right derivative is smaller, the player should change at a larger p; if the right derivative is larger, the player should change at a smaller p.

Assume that K players play R and $N-K-1$ play S for all beliefs in an interval $]p_\ell, p_r]$ with $p_\ell < p_r$. Let the remaining player, n, have a continuation payoff of $u_n \geq V_1^*(p_\ell)$ at the belief p_ℓ. Consider the response of player n on $]p_\ell, p_\ell + \varepsilon] \subseteq]p_\ell, p_r]$.

• $K > 0$: If his response is R then his value function on $]p_\ell, p_\ell + \varepsilon]$ is given by V_{K+1} from equation (11) with $V_{K+1}(p_\ell) = u_n$; if his response is S then his value function on $]p_\ell, p_\ell + \varepsilon]$ is given by F_K from equation (13) with $F_K(p_\ell) = u_n$. Now, if $V_{K+1}(p) = F_K(p) = u$, say, then $V'_{K+1}(p) > F'_K(p)$ if (p, u) lies above $\mathcal{D}_K$, and $V'_{K+1}(p) < F'_K(p)$ if (p, u) lies below $\mathcal{D}_K$. Thus, if (p_ℓ, u_n) lies on or above $\mathcal{D}_K$, then his best response is to "join in" by playing R; whereas if (p_ℓ, u_n) lies below $\mathcal{D}_K$, then his best response is to free-ride by playing S, and he can only change optimally at a belief $p_c \in]p_\ell, p_r]$ where $(p_c, F_K(p_c))$ is on $\mathcal{D}_K$.

• $K = 0$: If his response is S then his value function on $]p_\ell, p_\ell + \varepsilon]$ is simply s, so it must be the case that $u_n = s$. If his response is R then his value function on $]p_\ell, p_\ell + \varepsilon]$ is given by V_1 from equation (2) with $V_1(p_\ell) = u_n$; if it is also the case that $u_n = s$ and $p_\ell < p_1^*$ then $V_1'(p_\ell) < 0$ and his payoff on $]p_\ell, p_\ell + \varepsilon]$ would be less than s, but if $u_n > s$ or $p_\ell \geq p_1^*$ then (although maybe $V_1'(p_\ell) < 0$) his payoff on $]p_\ell, p_\ell + \varepsilon]$ is greater than s. Thus, if (p_ℓ, u_n) lies above $\mathcal{D}_0$ (the line $u = s$) or if $(p_\ell, u_n) = (p_1^*, s)$, then his best response is to "go it alone" by playing R; otherwise, his best response is to acquiesce by playing S, and he can only change optimally at a belief $p_c \in]p_\ell, p_r]$ where $p_c = p_1^*$. ■

Finitely many switches

Before proving Proposition 6.1 we have two preliminary lemmas.

Let v and f satisfy equation (2) (pioneer) and equation (13) (free-rider, with $K = 1$) respectively, with $v(p_1^*) = f(p_1^*) = s$. Let $\bar{p}_v$ be the belief where v meets $\mathcal{D}_1$, and let $\bar{p}_f$ be the belief where f meets $\mathcal{D}_1$. In Lemma A.1 we show first that there is a region of the (p, u)-plane that is bounded below by v between p_1^* and $\bar{p}_v$, bounded above by f between p_1^* and $\bar{p}_f$, and bounded on the right by $\mathcal{D}_1$. Next, we show that any solution to either equation (2) (pioneer) or equation (13) (free-rider, with $K=1$) that starts somewhere in that region can exit only through $\mathcal{D}_1$. Then in Lemma A.2 we show that any solution to equation (6) (both experimenting, $N=2$) that starts on $\mathcal{D}_1$ at a belief between $\bar{p}_f$ and $\bar{p}_v$ never hits $\mathcal{D}_1$ again. (Note that in each case, what is claimed and proved is a little stronger than necessary.)

Lemma A.1 *On $]p_1^*, p^m]$, $v < f$, v is strictly convex, f is strictly concave, and both are strictly increasing.*

Further, any function u_v which satisfies equation (2) (pioneer) on $[p_\ell, p_r] \subseteq [p_1^*, p^m]$ with $v(p_\ell) < u_v(p_\ell) \leq f(p_\ell)$ is strictly convex and strictly increasing, with $v(p) < u_v(p) < f(p)$ for $p_\ell < p \leq p_r$; and any function u_f which satisfies equation (13) (free-rider, with $K=1$) on $[p_\ell, p_r] \subseteq [p_1^*, p^m]$ with $v(p_\ell) \leq u_f(p_\ell) < f(p_\ell)$ is strictly concave and strictly increasing, with $v(p) < u_f(p) < f(p)$ for $p_\ell < p \leq p_r$.

PROOF: Whenever $v(p) = f(p)$, $v'(p) < f'(p)$ if $p < p^m$, and $v'(p) > f'(p)$ if $p > p^m$, so v and f can cross at most twice, once on either side of p^m, and, in between, $v < f$.

A calculation shows that the second derivative of the functions v and f has the same sign as the constant of integration. The boundary condition $v(p_1^*) = s$ implies a positive constant and therefore strict convexity of v, whereas the boundary condition $f(p_1^*) = s$ implies a negative constant and therefore strict concavity of f.

Evaluating, $v'(p_1^*) = 0$ and so v is strictly increasing on $[p_1^*, p^m]$.

To show that f is strictly increasing, we first define

$$Z_v(p) = \frac{(r+\lambda)gp}{r+\lambda p}, \quad Z_f(p) = \frac{rs+\lambda gp}{r+\lambda p}, \quad \text{and} \quad L(p) = s + \frac{g-s}{1-p_1^*}(p-p_1^*).$$

Z_v has the property that if u_v satisfies equation (2) (pioneer), then whenever $u_v(p) = Z_v(p)$ we have $u_v'(p) = 0$ (for $p \neq 1$), and whenever $u_v(p) < Z_v(p)$ we have $u_v'(p) > 0$. Similarly, Z_f has the property that if u_f satisfies equation (13) (free-rider, with $K=1$), then whenever $u_f(p) = Z_f(p)$ we have $u_f'(p) = 0$ (for $p \neq 1$), and whenever $u_f(p) < Z_f(p)$ we have $u_f'(p) > 0$. Finally, L is the linearization of f at (p_1^*, s), i.e. $L(p_1^*) = f(p_1^*) = s$, and $L'(p_1^*) = f'(p_1^*) = \frac{g-s}{1-p_1^*}$, so $L(p) > f(p)$ for $p \neq p_1^*$.

$Z_f(p) > Z_v(p)$ whenever $p < p^m$. Z_v is strictly concave, with $Z_v(p_1^*) = s$ and $Z_v(1) = g$, so Z_v and L coincide at (p_1^*, s) and $(1, g)$, and since L is linear we have $Z_v(p) > L(p)$ on $]p_1^*, 1[$. Thus we have the string of inequalities:

$$Z_f(p) > Z_v(p) > L(p) > f(p)$$

whenever $p_1^* < p < p^m$, showing that f is strictly increasing there.

Since u_v and v both satisfy equation (2) (pioneer) on $[p_\ell, p_r]$ with $v(p_\ell) < u_v(p_\ell)$, the constant of integration for v is strictly less than that for u_v; it follows from the strict convexity of v that both constants are positive and hence that u_v is strictly convex. Also, u_v and v cannot cross, so u_v remains above v. If $u_v(p_\ell) < f(p_\ell)$, then u_v remains below f on $[p_\ell, p_r]$ since it can only cross from below to the right of p^m; if $u_v(p_\ell) = f(p_\ell)$, then u_v falls below f immediately to the right of p_ℓ. Consequently, $u_v < f$ on $]p_\ell, p_r]$, and since $f < Z_v$ there, it follows that u_v is strictly increasing.

Similarly, since u_f and f both satisfy equation (13) (free-rider, with $K = 1$) on $[p_\ell, p_r]$ with $f(p_\ell) > u_f(p_\ell)$, the constant of integration for f is strictly greater than that for u_f; it follows from the strict concavity of f that both constants are negative and hence that u_f is strictly concave. Also, u_f and f cannot cross, so u_f remains below f. Moreover, since $f < Z_f$ on $]p_1^*, p^m[$, it follows that u_f is strictly increasing. If $u_f(p_\ell) > v(p_\ell)$, then u_f remains above v on $[p_\ell, p_r]$ since it can only cross from above to the right of p^m; if $u_f(p_\ell) = v(p_\ell)$, then u_f rises above v immediately to the right of p_ℓ. Consequently, $u_f > v$ on $]p_\ell, p_r]$. ■

Let u satisfy equation (6) (both experimenting, $N=2$) on $[\bar{p}, 1]$ for $p_1^* \leq \bar{p} \leq p^m$ and with $v(\bar{p}) \leq u(\bar{p}) \leq f(\bar{p})$.

Lemma A.2 *u is strictly convex and strictly increasing.*

PROOF: Let u_v satisfy equation (2) (pioneer) on $[\bar{p}, p^m]$ with $u_v(\bar{p}) = u(\bar{p})$. Inspection of equations (2) (pioneer) and (6) (both experimenting, $N = 2$) shows that the constant of integration for u has the same sign as that for u_v, namely it is positive (by the previous lemma). As above, the second derivative of the function u has the same sign as the constant of integration, hence u is strictly convex.

Since $(\bar{p}, u(\bar{p}))$ is above and to the left of the myopic payoff we have $u'(\bar{p}) > u'_v(\bar{p})$, and also $u'_v(\bar{p}) \geq 0$ (by the previous lemma), hence u is strictly increasing. ■

Thus, we have the following result.

Fix some $\bar{p}$ between $\bar{p}_f$ and $\bar{p}_v$. Let U be the continuous function on $[p_1^*, 1]$ with $U(p_1^*) = s$ that satisfies equation (13) (free-rider, with $K=1$) on some finite subpartition of $[p_1^*, \bar{p}]$ and equation (2) (pioneer) on its finite complement, such that U meets $\mathcal{D}_1$ at $\bar{p}$; further, U satisfies equation (6) (both experimenting, $N=2$) on $[\bar{p}, 1]$. Then U lies between v and f below and to the left of $\mathcal{D}_1$ and is strictly concave where it satisfies equation (13) (free-rider, with $K=1$), strictly convex elsewhere, and strictly increasing.

Proof of Proposition 6.1

Let $\bar{p}_2$ denote the smallest belief where each player's continuation payoff is (weakly) above $\mathcal{D}_1$, and let $\bar{p}_s$ denote the largest belief where each player's continuation payoff is (weakly) below $\mathcal{D}_1$; by Lemma A.1, $p_1^* < \bar{p}_s \leq \bar{p}_2 < p^m$.

For a belief in a neighbourhood of 1, specifically $p \in]\bar{p}_2, 1]$, R is the dominant strategy; and for a belief in a neighbourhood of 0, specifically $p \in [0, p_1^*]$, S is the dominant strategy. (We know that $u_n(0) = s$, and so S is a dominant response on any interval $[0, p_c] \subseteq [0, p_1^*]$). For beliefs $p \in]p_1^*, \bar{p}_s]$, the best response to S is to play R (act unilaterally), and the best response to R is to play S (free-ride). Now consider beliefs $p \in]\bar{p}_s, \bar{p}_2]$; let A be the player whose continuation payoff crosses $\mathcal{D}_1$ at $\bar{p}_s$ and let B be the player whose continuation payoff crosses $\mathcal{D}_1$ at $\bar{p}_2$. If B plays S, then A's best response is to play R ("go it alone"), and if B plays R, then A's best response is to play R ("join in"); thus R is the dominant response for A. So, given A plays R, B's best response is to play S (free-ride). To summarize:

Belief p	0	p_1^*	$\bar{p}_s$	$\bar{p}_2$	1
A's strategy	S	S/R	R	R	
B's strategy	S	R/S	S	R	
A's continuation payoff	s	$F_{1,A}/V_{1,A}$	$V_{1,A}$	$V_{2,A}$	
B's continuation payoff	s	$V_{1,B}/F_{1,B}$	$F_{1,B}$	$V_{2,B}$	

and the strategies on $]p_1^*, \bar{p}_s]$ determine $\bar{p}_s$ endogenously, which player plays R and which player plays S on $]\bar{p}_s, \bar{p}_2]$, and $\bar{p}_2$ endogenously. If the players have the above continuation payoffs, then the above strategies are best responses to each other; and if the players are using the above strategies, then the continuation payoffs are indeed those given above. Thus the above strategies constitute an equilibrium with the equilibrium value functions given by the continuation payoffs.

The 'simplest' equilibrium is where one player, say player 1, plays S on $]p_1^*, \hat{p}_s]$, and the other player, player 2, plays R on this interval.[11] Then player 1's value function F_1 satisfies equation (13) and player 2's value function V_1 satisfies equation (2), with $F_1(p_1^*) = V_1(p_1^*) = s$. Lemma A.1 shows that F_1 meets $\mathcal{D}_1$ at a *smaller* belief than does V_1, and that $F_1 > V_1$ on $]p_1^*, \hat{p}_s]$; that is, player 1 must be A and switch from playing R on $]\hat{p}_s, \hat{p}_2]$, and player 2 must

[11] $\hat{p}_s$ is the same belief as $\bar{p}_f$ used in Lemma A.1.

be B and switch from playing S on $]\hat{p}_s, \hat{p}_2]$. This equilibrium is thus given by:

Belief p	0	p_1^*	$\hat{p}_s$	$\hat{p}_2$	1
A's strategy	S	S	R	R	
B's strategy	S	R	S	R	
A's value function	s	$F_{1,A}$	$V_{1,A}$	$V_{2,A}$	
B's value function	s	$V_{1,B}$	$F_{1,B}$	$V_{2,B}$	

and the components of the value functions, plus the switch-point and cut-off, are determined as follows:

(1) C in $F_{1,A}$ from $F_{1,A}(p_1^*) = s$

(2) C in $V_{1,B}$ from $V_{1,B}(p_1^*) = s$

(3) $\hat{p}_s$ from $F_{1,A}(\hat{p}_s) = 2s - g\hat{p}_s$

(4) C in $V_{1,A}$ from $V_{1,A}(\hat{p}_s) = F_{1,A}(\hat{p}_s) = 2s - g\hat{p}_s$

(5) C in $F_{1,B}$ from $F_{1,B}(\hat{p}_s) = V_{1,B}(\hat{p}_s)$

(6) $\hat{p}_2$ from $F_{1,B}(\hat{p}_2) = 2s - g\hat{p}_2$

(7) C in $V_{2,A}$ from $V_{2,A}(\hat{p}_2) = V_{1,A}(\hat{p}_2)$

(8) C in $V_{2,B}$ from $V_{2,B}(\hat{p}_2) = F_{1,B}(\hat{p}_2) = 2s - g\hat{p}_2$

Note that the boundary condition at $p = 1$ is automatically satisfied because $V_{2,A}(1) = V_{2,B}(1) = g$ regardless of the constants of integration.

Noting that when $V_2(p) = V_1(p) = u$, say, $V_2'(p) > V_1'(p)$ iff $u > gp$ (the payoff from always playing R), we see that

- $0 < F'_{1,A}(p_1^*), \quad F'_{1,A}(\hat{p}_s) > V'_{1,A}(\hat{p}_s), \quad V'_{1,A}(\hat{p}_2) < V'_{2,A}(\hat{p}_2)$;
- $0 = V'_{1,B}(p_1^*), \quad V'_{1,B}(\hat{p}_s) < F'_{1,B}(\hat{p}_s), \quad F'_{1,B}(\hat{p}_2) = V'_{2,B}(\hat{p}_2)$.

Thus, as the common belief decays, B changes smoothly from R to S against R at $\hat{p}_2$ (where A has a kink), A and B switch actions at $\hat{p}_s$ (each with a kink), and B changes smoothly again from R to S against S at p_1^* (where A again has a kink).

Following steps (1) and (3) determines the equation for $\hat{p}_s$ given in the statement of the proposition; following steps (2), (5) and (6) determines the equation for $\hat{p}_2$ given in the statement of the proposition; the remaining steps are for completeness only.[12]

Other equilibria for the two-player strategic problem

Any finite partition of the interval to the right of p_1^* can be used to construct a pure-strategy equilibrium of the two-player strategic problem.

Take any finite (measurable) partition of $]p_1^*, p^m]$ and divide this into two subsets I_n, $n = 1, 2$. Build the continuous functions X_n on $[p_1^*, p^m]$ as follows: $X_n(p_1^*) = s$, X_n satisfies equation (13) on I_n (free-rider), X_n satisfies equation (2) on $I_{\neg n}$ (pioneer).

Define $\bar{p}_s = \min\{p \in [p_1^*, p^m] : X_1(p) \vee X_2(p) = 2s - gp\}$. If $X_n(\bar{p}_s) \geq X_{\neg n}(\bar{p}_s)$ then $A = n$, else $A = \neg n$; $B = \neg A$.

[12] Details are available from the authors on request.

Let u_f satisfy equation (13) (free-rider) with $u_f(\bar{p}_s) = X_B(\bar{p}_s)$, and define $\bar{p}_2$ by $u_f(\bar{p}_2) = 2s - g\bar{p}_2$, so $\bar{p}_2 \geq \bar{p}_s$.

Now take the partition $J_1 \cup J_2$ of $]p_1^*, \bar{p}_s]$, where $J_n = \{p \leq \bar{p}_s : p \in I_n\}$, i.e. J_n and I_n agree on $]p_1^*, \bar{p}_s]$.

Let A's strategy be as follows:
play S on $[0, p_1^*]$; play S on J_A and R on J_B; play R on $]\bar{p}_s, \bar{p}_2]$; play R on $]\bar{p}_2, 1]$.

Let B's strategy be as follows:
play S on $[0, p_1^*]$; play R on J_A and S on J_B; play S on $]\bar{p}_s, \bar{p}_2]$; play R on $]\bar{p}_2, 1]$.

Build the continuous functions Y_n on $[0, 1]$ as follows:

$Y_A(p) = s$ on $[0, p_1^*]$; Y_A satisfies equation (13) on J_A (free-rider) and satisfies equation (2) on J_B (pioneer); Y_A satisfies equation (2) on $]\bar{p}_s, \bar{p}_2]$ (pioneer); Y_A satisfies equation (6) on $]\bar{p}_2, 1]$ (both experimenting).

$Y_B(p) = s$ on $[0, p_1^*]$; Y_B satisfies equation (2) on J_A (pioneer) and satisfies equation (13) on J_B (free-rider); Y_B satisfies equation (13) on $]\bar{p}_s, \bar{p}_2]$ (free-rider); Y_B satisfies equation (6) on $]\bar{p}_2, 1]$ (both experimenting).

If the continuation payoffs are given by Y_n, then the above strategies are best responses to each other; and if the players are using the above strategies, then the continuation payoffs are indeed given by Y_n. Thus the above strategies constitute an equilibrium with the equilibrium value functions given by Y_n.

Lemmas A.1 and A.2 show that Y_A and Y_B lie between $F_{1,A}$ and $V_{1,B} \cup F_{1,B}$ below and to the left of $\mathcal{D}_1$. Thus $\hat{p}_s \leq \bar{p}_s \leq \bar{p}_2 \leq \hat{p}_2$ (at least one inequality being strict), and so the 'simplest' equilibrium exhibits the slowest experimentation. ■

Proof of Proposition 6.3

Consider a pure-strategy MPE with cut-offs $\{\bar{p}_K\}_{K=1}^N$. Let $u_{N,K}$ denote the solution to the ODE (14) whenever we are in case (K), with $u_{N,1}(\bar{p}_1) = s$ and $u_{N,K}(\bar{p}_K) = u_{N,K-1}(\bar{p}_K)$ for $K = 2, \ldots, N$. The continuous function thus constructed is the average payoff.

Define a related function $u_{N|1}$ which equals $u_{N,1}$ to the left of $\bar{p}_2$, but carries on as $u_{N,1}$ to the right, i.e. $u_{N|1}$ is continuous and solves (14) on $[\bar{p}_1, 1]$, not just on $[\bar{p}_1, \bar{p}_2]$. Since $u'_{N,K}(p) > u'_{N,J}(p)$ whenever $u_{N,K}(p) = u_{N,J}(p) > s$ and $K > J$, it is the case that $u_{N|1} < u_{N,K}$ to the right of $\bar{p}_2$. Thus if we can show that even $u_{N|1}$ does better than the common payoff in the symmetric mixed-strategy MPE, we are done.

To simplify notation we shall use a normalized odds ratio given by $R(p) = (\Omega(p)/\Omega(p_1^*))^\mu$, which is decreasing in p and less than 1 for $p > p_1^*$.

Using a combination of equations (11) and (13), $u_{N|1}$ satisfies

$$N\,\frac{u_{N|1}(p) - s}{s(1-p)} = \frac{\mu(\mu+N)}{(\mu+1)^2} R(p)^{-1/\mu} + \frac{1-(N-2)\mu}{(\mu+1)^2} R(p) - 1$$

where we have used the fact that $\Omega(p^m)/\Omega(p_1^*) = \mu/(\mu+1)$.

The common payoff in the symmetric mixed-strategy equilibrium is given by the function $W^\dagger$ on $[p_1^*, p_N^\dagger]$. Using equation (18), this function satisfies

$$\frac{W^\dagger(p) - s}{s(1-p)} = \mu R(p)^{-1/\mu} + \ln R(p) - \mu.$$

A simple calculation now gives

$$N\,\frac{u_{N|1}-W^\dagger}{s(1-p)}=-\frac{\mu^2(N\mu+2N-1)}{(\mu+1)^2}R^{-1/\mu}+\frac{1-(N-2)\mu}{(\mu+1)^2)}R-N\ln R+N\mu-1,$$

where we have suppressed the dependence of $u_{N|1}$, $W^\dagger$ and R on p. We want to show that the right-hand side is positive on the interval $]p_1^*,p^m]$. To this end, we consider the right-hand side as a function $h(R;N)$ on the interval $[R(p^m),1]$. As $h(1;N)=0$, $h'(1;N)=-(N-1)/(\mu+1)<0$ and $h''(R;N)=R^{-2}\{N-[(N\mu+2N-1)/(\mu+1)]R^{-1/\mu}\}<0$ on this interval, it suffices to show that $h(R(p^m);N)>0$. Now $R(p^m)=[\mu/(\mu+1)]^\mu$, so

$$h(R(p^m);N)=-\frac{N\mu+1}{\mu+1}+\frac{1-(N-2)\mu}{(\mu+1)^2}\left(\frac{\mu}{\mu+1}\right)^\mu-N\mu\ln\frac{\mu}{\mu+1}\,.$$

As a function of μ on the positive half-axis, this is quasi-concave with a limit of zero as μ tends to 0 or $+\infty$ for any $N>1$, hence positive throughout.

For $p\in]p_1^*,p_N^\dagger]$, therefore, the average payoff in the pure-strategy equilibrium lies strictly above the common payoff in the symmetric mixed-strategy equilibrium. ■

Proof of Proposition 6.4

Given $p_\infty^\ddagger$ with $p_N^*<p_\infty^\ddagger<p_1^*$, consider an arbitrary strictly decreasing sequence $\{p_i^\ddagger\}_{i=1}^\infty$ with $p_1^\ddagger\le p_1^*$ and $\lim_{i\to\infty}p_i^\ddagger=p_\infty^\ddagger$. Let player $n=1,\ldots,N$ play R at beliefs in the intervals $]p_{n+jN+1}^\ddagger,p_{n+jN}^\dagger]$ $(j=0,1,\ldots)$ and S at all other beliefs below $p_1^\ddagger$.

For arbitrary i, consider the player who who embarks on her round of single-handed experimentation at $p_i^\ddagger$, that is, who plays R on $]p_{i+1}^\ddagger,p_i^\ddagger]$ and S on $]p_{i+N}^\ddagger,p_{i+1}^\ddagger]$. Her payoff function u satisfies equation (2) (pioneer) on the former interval, and equation (13) (free-rider, with $K=1$) on the latter. Imposing the conditions $u(p_i^\ddagger)=u(p_{i+N}^\ddagger)=s$, we can solve for the respective constants of integration. This yields two equations for $u(p_{i+1}^\ddagger)$:

$$\begin{aligned}u(p_{i+1}^\ddagger)&=gp_{i+1}^\ddagger+(s-gp_i^\ddagger)\frac{1-p_{i+1}^\ddagger}{1-p_i^\ddagger}\left(\frac{\Omega(p_{i+1}^\ddagger)}{\Omega(p_i^\ddagger)}\right)^\mu,\\ u(p_{i+1}^\ddagger)&=s+\frac{g-s}{\mu+1}p_{i+1}^\ddagger-\frac{g-s}{\mu+1}p_{i+N}^\ddagger\frac{1-p_{i+1}^\ddagger}{1-p_{i+N}^\ddagger}\left(\frac{\Omega(p_{i+1}^\ddagger)}{\Omega(p_{i+N}^\ddagger)}\right)^\mu.\end{aligned}$$

After eliminating $u(p_{i+1}^\ddagger)$ from these equations, we change variables to

$$x_i=\frac{\Omega(p_i^\ddagger)}{\Omega(p^m)},$$

noting that $\Omega(p^m)=(g-s)/s$ and

$$\frac{s-gp_i^\ddagger}{(1-p_i^\ddagger)s}=1-\frac{1}{x_i}.$$

This leads to the N-th order difference equation

$$\frac{1}{\mu+1}\;x_{i+N}^{-\mu-1}=\left[1-\frac{\mu}{\mu+1}\frac{1}{x_{i+1}}\right]x_{i+1}^{-\mu}-\left[1-\frac{1}{x_i}\right]x_i^{-\mu}.$$

Introducing the variables

$$y_{i,n} = \frac{x_{i+n} - x_{i+n-1}}{x_{i+n-1}} \qquad (n = 1, \ldots, N-1),$$

we obtain the N-dimensional first-order system

$$\begin{aligned}
x_{i+1} &= x_i\,(1 + y_{i,1}),\\
y_{i+1,1} &= y_{i,2},\\
&\vdots\\
y_{i+1,N-2} &= y_{i,N-1},\\
y_{i+1,N-1} &= \left(\prod_{n=2}^{N-1}(1+y_{i,n})\right)^{-1}\Big[(\mu+1)x_i(1+y_{i,1}) - (\mu+1)(x_i-1)(1+y_{i,1})^{\mu+1} - \mu\Big]^{-\frac{1}{\mu+1}} - 1.
\end{aligned}$$

Writing

$$x_\infty = \frac{\Omega(p^\ddagger_\infty)}{\Omega(p^m)},$$

we clearly have a steady state of this system at $(x_\infty, 0, \ldots, 0)$.

The linearization of the system around this steady state is

$$\begin{pmatrix} x_{i+1} - x_\infty \\ y_{i+1,1} \\ y_{i+1,2} \\ y_{i+1,3} \\ \vdots \\ y_{i+1,N-3} \\ y_{i+1,N-2} \\ y_{i+1,N-1} \end{pmatrix} = \begin{pmatrix} 1 & x_\infty & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 1 \\ 0 & \xi & -1 & -1 & \ldots & -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} x_i - x_\infty \\ y_{i,1} \\ y_{i,2} \\ y_{i,3} \\ \vdots \\ y_{i,N-3} \\ y_{i,N-2} \\ y_{i,N-1} \end{pmatrix},$$

where

$$\xi = \mu x_\infty - \mu - 1.$$

Since the characteristic polynomial is $(-1)^{N-1}(1-\eta)h(\eta)$ with $h(\eta) = \eta^{N-1} + \eta^{N-2} + \ldots + \eta^2 + \eta - \xi$, the eigenvalues are 1 and the zeroes of h. As $p^*_N < p^\ddagger_\infty < p^*_1$, we have $(\mu+1)/\mu < x_\infty < (\mu+N)/\mu$ and so $0 < \xi < N-1$. Thus, $h(0) = -\xi < 0$ and $h(1) = N - 1 - \xi > 0$, implying the existence of an eigenvalue η_* strictly between 0 and 1. A corresponding eigenvector is

$$\begin{pmatrix} -x_\infty/(1-\eta_*) \\ 1 \\ \eta_* \\ \eta_*^2 \\ \vdots \\ \eta_*^{N-2} \end{pmatrix}.$$

This shows that under the linearized dynamics, $(x_\infty, 0, \ldots, 0)$ can be approached in such a way that the sequence $\{x_i\}$ is strictly increasing. By standard results from the theory of dynamical systems, the same is possible under the original nonlinear dynamics if we start from a suitable initial point in a neighbourhood of the steady state; see for example Wiggins (1990, Section 1.1C).

Starting appropriately close to the steady state, we can ensure in particular that the strategies we obtain for the corresponding sequence of beliefs $\{p_i^\ddagger\}_{i=1}^\infty$ are mutual best responses at all beliefs below $p_1^\ddagger$ (all we need for this is that $(p_1^\ddagger, u_N)$ be below $\mathcal{D}_1$, where u_N is the continuation payoff of player N when the common belief is $p_1^\ddagger$ – see Lemma 6.1). To complete the construction of the equilibrium, we now only have to move back from $p_1^\ddagger$ to higher beliefs and assign actions to the players in the way we did for the pure-strategy equilibria with a finite number of switches (see the outline after Lemma 6.1). ■

References

Admati, A.R. and M. Perry (1991): "Joint Projects without Commitment", *Review of Economic Studies*, **58**, 259–276.

Bergemann, D. and U. Hege (1998): "Venture Capital Financing, Moral Hazard and Learning", *Journal of Banking and Finance*, **22**, 703–735.

Bergemann, D. and U. Hege (2001): "The Financing of Innovation: Learning and Stopping" (CEPR Discussion Paper No. 2763).

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

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

Harris, C. (1993): "Generalized Solutions to Stochastic Differential Games in One Dimension" (mimeo, Nuffield College, Oxford).

Keller, G. and S. Rady (2003): "Price Dispersion and Learning in a Dynamic Differentiated-Goods Duopoly" (forthcoming in the *RAND Journal of Economics*).

Lockwood, B. and J.P. Thomas (2002): "Gradualism and Irreversibility", *Review of Economic Studies*, **69**, 339–356.

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

Marx, L. and S. Matthews (2000): "Dynamic Voluntary Contribution to a Public Project", *Review of Economic Studies*, **67**, 327–358.

Rothschild, M. (1974): "A Two-Armed Bandit Theory of Market Pricing", *Journal of Economic Theory*, **9**, 185–202.

Wiggins, S. (1990): *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. New York: Springer-Verlag.

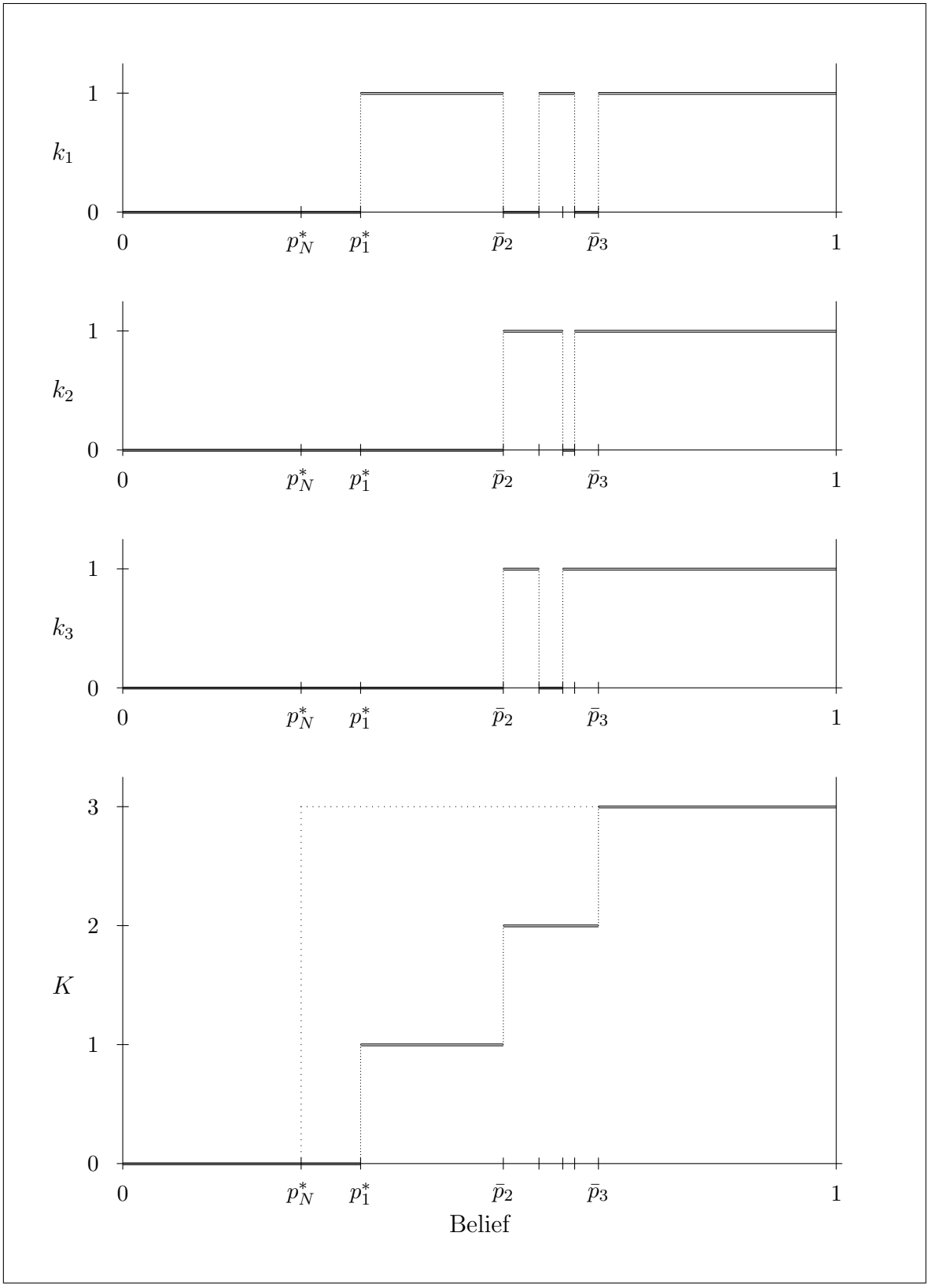

Figure 5: Action profiles in the simplest three-player asymmetric equilibrium