STRATEGIC EXPERIMENTATION WITH EXPONENTIAL BANDITS

By Godfrey Keller, Sven Rady, and Martin Cripps*

April 2004

Abstract

We analyze a game of strategic experimentation with two-armed bandits whose risky arm might yield payoffs after exponentially distributed random times. Free-riding causes an inefficiently low level of experimentation in any equilibrium where the players use stationary Markovian strategies with beliefs as the state variable. We construct the unique symmetric Markovian equilibrium of the game, followed by various asymmetric ones. There is no equilibrium where all players use simple cut-off strategies. Equilibria where players switch finitely often between experimenting and free-riding all yield a similar pattern of information acquisition, greater efficiency being achieved when the players share the burden of experimentation more equitably. When players switch roles infinitely often, they can acquire an approximately efficient amount of information, but still at an inefficient rate. In terms of aggregate payoffs, all these asymmetric equilibria dominate the symmetric one wherever the latter prescribes simultaneous use of both arms.

Keywords: Strategic Experimentation, Two-Armed Bandit, Exponential Distribution.

*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 London School of Economics, Stanford GSB, the universities of Munich, Oxford, Pennsylvania, Toulouse, Warwick, Wisconsin–Madison, the IGIER Workshop in Economic Theory at Erasmus University Rotterdam, 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 thank the Financial Markets Group at the London School of Economics, the Studienzentrum Gerzensee and the Institut d'Economie Industrielle at the University of Toulouse for their hospitality. This version has benefited from comments and suggestions from Glenn Ellison and anonymous referees.

1 Introduction

In this article we analyze a game of strategic experimentation based upon two-armed bandits with a safe arm that offers a known payoff and a risky arm of unknown type. If it is good, it generates positive payoffs after exponentially distributed random times; if it is bad, it never pays out anything. 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 is endowed with one unit of a perfectly divisible resource per unit of time, and continually has to decide what fraction of this resource to allocate to each arm. Players observe each others' actions and outcomes, so information about the type of the risky arm is a public good.

Such a game of strategic experimentation arises for example with drug trials by health authorities. Suppose there is an established drug with a known probability of curing a certain disease, while a new drug is either more highly effective or little better than a placebo. Health authorities in different jurisdictions face the allocation problem of deciding what fraction of patients to give what drug. Given a free flow of information, they will draw on each other's experience, and a lack of success with the new drug would make all parties more pessimistic. In the limiting case (assumed in our model) that a bad new drug is completely ineffective, the first success with the new drug reveals it to be good, leading all parties to adopt it.

As a second example, consider oil or gas companies any number of which may explore a virgin site. This new site could either be very rewarding, having more reserves than an old established one, or it could prove to be barren. Each company has to decide how much of its effort, time and money to allocate to the new site, and how much to the old one. The first breakthrough discovery by anyone at the new site reveals its superiority, and all companies would henceforth drill there.

In situations such as these, players gradually become less optimistic as long as no news arrives, and fully informed as soon as it does. In our model, this entails that the 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 Markov perfect equilibria with the players' posterior belief as state variable.

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, 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 were only one player. This means that 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 analyzed by Bolton and Harris (1999). Its absence in the exponential bandit framework is easy to explain, since the encouragement effect rests on two conditions: the additional experimentation by one player now must increase the likelihood that other players will experiment in the future, and this future experimentation must be valuable to the player who acts 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 – 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.

Having characterized the symmetric MPE, we go on to construct asymmetric equilibria in which players use *simple* strategies, i.e. strategies that at any given belief allocate the entire unit resource to one arm. We show that the players generate the same *amount* of information in all simple equilibria if they 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.

Quite surprisingly, there is no Markov perfect equilibrium where all players use the simple cut-off strategies that are familiar from single-agent bandit problems (use only the risky arm above some cut-off belief, and only the safe arm below it). In any two-player simple equilibrium, for example, there is an 'optimistic threshold' (where the number of experimenters drops from two to one) and a 'pessimistic threshold' (where the number of experimenters drops from one to zero), but it is not compatible with equilibrium that one player performs all the experimentation between these two thresholds while the other is free-riding. This is because the free-rider would have the higher continuation value at beliefs above the pessimistic threshold, and best responses are weakly increasing in one's continuation value. Consequently, there would be a range of beliefs around the optimistic threshold where he would find it optimal to experiment regardless of what his rival is doing. In turn, since best responses are weakly decreasing in the number of other players experimenting, the player who is supposed to run the final leg would find it optimal to free-ride just below the optimistic threshold. So there must be at least one intermediate belief where the players swap roles, and the simplest two-player MPE have precisely one such switch-point. This can be elaborated on by many switches between the role of free-rider and pioneer, and we give a complete characterization of when and how this can happen. Further, we indicate how these results generalize to the N-player case.

Although the amount of information acquired is fixed over all simple 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 more they are able to maintain higher rates of information acquisition, and the more efficient is the equilibrium. We also show that over the range of beliefs where the symmetric equilibrium prescribes simultaneous use of both arms, even the worst simple equilibria are more efficient than the symmetric one, and simple equilibria in which players frequently switch roles Pareto-dominate the symmetric one except for beliefs in an arbitrarily small interval just above the single-agent cut-off.

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 that is approximately efficient can then be induced; the rate at which this information is acquired is, however, 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), but has received little attention so far, Presman (1990) covering the single-agent case. Exponential bandits are embedded in the financial contracting models of Bergemann and Hege (1998, 2001) and the investment timing model of Décamps and Mariotti (2003). 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, which leads to the same structure of belief revisions as with exponential bandits. In these papers, however, the strategic interaction is entirely different from the situation that we consider.

The articles closest to us are Bolton and Harris (1999, 2000). 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 (1999) 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 asymmetric equilibria and show that they are more efficient than the symmetric one.

While the general case proves to be tractable in our framework, Bolton and Harris (2000) take their model to the undiscounted limit thereby shutting down the encouragement effect and facilitating the construction of asymmetric equilibria. They demonstrate an asymmetric equilibrium in pure strategies that depends primarily on the full-information payoff *and not* on the players' value functions; by contrast, in the equilibria that emerge from our model, the best responses and value functions are co-determined. In both models a graph of aggregate experimentation against beliefs is a step function, but in their model the cut-offs where there is a move from one step to the next are fixed, whereas in our model these cut-offs are determined endogenously – exchanging the roles of two players when aggregate experimentation is one, for example, will ripple all the way up, altering all the higher cut-offs and changing the equilibrium.

Our article is also related to a recent literature on the dynamic provision of public goods which has found that efficient provision is possible if the players make 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 because their net gain (from others' future contributions) outweighs their current cost of provision. Although our infinite switching equilibrium with its approximately efficient amount of information is quite different – time is continuous whereas actions are binary – 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 article is organized as follows. Section 2 introduces the exponential bandit model and Section 3 establishes the efficient benchmark where players cooperate in order to maximize joint expected payoffs. Section 4 sets up the strategic problem and provides several auxiliary results. Sections 5 and 6 present the unique symmetric Markov perfect equilibrium, asymmetric equilibria, and welfare comparisons. Section 7 contains some concluding remarks; some of the proofs are relegated to the Appendix.

2 Exponential Bandits

The purpose of this section is to introduce our model of strategic experimentation.

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

More precisely, if a player allocates the fraction $k_t \in [0, 1]$ of the resource to R over an interval of time $[t, t + dt)$, and consequently the fraction $1 - k_t$ to S, then she receives the payoff $(1 - k_t)s\, dt$ from S, where $s > 0$ is a constant known to all players. Moreover, the probability that she receives a lump-sum payoff from R at some point in the interval is $k_t \lambda \theta\, dt$ where $\lambda > 0$ is a constant known to all players, $\theta = 1$ if R is good, and $\theta = 0$ if R is bad. Lump-sums are independent draws from a time-invariant distribution on $\mathbb{R}_{++}$ with a known mean h. When the fraction k_t of the resource is allocated to R on $[t, t+dt)$, therefore, the overall expected payoff increment conditional on θ is $[(1 - k_t)s + k_t g\theta]\, dt$, where $g = \lambda h$.

We assume that $0 < s < g$, so each player strictly prefers R, if it is good, to S, and strictly prefers S to R, if it is bad. At $t = 0$, however, players do not know whether the risky arm is good or bad; they start with a common prior belief about θ. Thereafter, all players observe each other's actions and outcomes, so they hold common posterior beliefs throughout time. This means in particular that the arrival of the first lump-sum on any of the N risky arms reveals to all players that the risky arm is good.

If a player used a time-invariant allocation $k_t = k$, then by standard results the delay between the arrivals of successive lump-sums on a good risky arm would be exponentially distributed with parameter $k\lambda$, that is, with mean $1/(k\lambda)$; see Karlin and Taylor (1981, p.146), for instance. In particular, the time of the first-lump sum, and hence of the resolution of players' uncertainty about the type of the risky arm, would be exponentially distributed. This is the reason why we use the term *exponential bandits* for our set-up. Players are said to *experiment* when they use R while its type is still unknown.

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

units, is

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

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

Let p_t denote the subjective probability at time t that players assign to the risky arm being good, so that given a player's action k_t on $[t, t+dt)$, her expected payoff increment conditional on all available information is $[(1-k_t)s + k_t g p_t]\, dt$. By the Law of Iterated Expectations, we can rewrite the above total payoff as

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

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

To derive the law of motion of beliefs, suppose that over the interval of time $[t, t+dt)$ player $n = 1, \ldots, N$ allocates the fraction $k_{n,t}$ of the unit resource to her risky arm. Let $K_t = \sum_{n=1}^N k_{n,t}$. If the risky arms are good, the probability of none of the players achieving a breakthrough is $\prod_{n=1}^N (1 - k_{n,t}\lambda\, dt) = 1 - K_t\lambda\, dt$ (this is up to terms of the order $o(dt)$, which we can ignore here and in what follows); if the risky arms are bad, the probability of none of them achieving a breakthrough is 1. When the players start with the common belief p_t at time t and none of them achieves a breakthrough in $[t, t+dt)$, therefore, the updated belief at the end of that time period is

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

by Bayes' rule. Simplifying, we see that as long as there is no breakthrough, the belief changes by $dp_t = -K_t\lambda p_t(1-p_t)\, dt$. Once there is a breakthrough, of course, the posterior belief jumps to 1.

Were an agent to act myopically, she would simply maximize the expected short-run payoff $(1-k_t)s + k_t g p_t$. The belief that makes her indifferent between R and S is $p^m = s/g$. For $p_t > p^m$ it is myopically optimal to allocate the resource exclusively to R; for $p_t < p^m$ it is myopically optimal to allocate it exclusively to S. As we shall see next, it is efficient to play R exclusively for some beliefs below p^m.

3 The Cooperative Problem

Suppose that the N players work cooperatively, i.e. want to maximize the *average* expected payoff by jointly choosing the action profiles $\{(k_{1,t}, \ldots, k_{N,t})\}_{t\geq 0}$. This is a dynamic programming problem with the current belief p as the state variable.

If current actions are $(k_1, \ldots, k_N)$, the average expected payoff increment is given by $\left[(1-\frac{K}{N})s + \frac{K}{N}gp\right] dt$ with $K = \sum_{n=1}^{N} k_n$, and the subjective probability of a breakthrough on at least one risky arm is $pK\lambda\, dt$. As the evolution of beliefs also merely depends on K, the cooperative's problem reduces to choosing the optimal level of the overall allocation K given the current belief p.

By the Principle of Optimality, the value function of the cooperative, expressed as average payoff per agent, satisfies

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

where the first term is the average expected current payoff again and the second term is the discounted expected continuation payoff. As to the latter, with subjective probability $pK\lambda\, dt$ a breakthrough occurs and the value function jumps to $u(1) = g$; with probability $p(1-K\lambda\, dt)+(1-p) = 1-pK\lambda\, dt$ there is no breakthrough and the value function changes to $u(p) + u'(p)dp = u(p) - K\lambda p(1-p)u'(p)\, dt$.[1] Using these expectations, together with $1 - r\, dt$ as an approximation to $e^{-r\,dt}$, we replace the second term in the earlier equation, simplify and rearrange, to obtain the Bellman equation

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

which can be rewritten more succinctly as

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

where

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

and

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

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

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 linearity in K of the maximand in the Bellman equation immediately implies that it is always optimal to choose either $K=0$ (all agents use S exclusively), or $K=N$ (all agents use R exclusively), depending on whether or not the *shared* opportunity cost of playing R exceeds the full expected benefit. In the former case, $u(p)=s$; in the latter case, u satisfies the first-order ODE

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

This has the solution

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

where

$$\Omega(p)=\frac{1-p}{p}$$

denotes the odds ratio at the belief p, and $\mu=r/\lambda$. The first term, gp, is the expected payoff from committing to the risky arm, while the second term 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_N.

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

$$p_N^*=\frac{\mu s}{(\mu+N)(g-s)+\mu s} \tag{3}$$

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

$$V_N^*(p)=gp+(s-gp_N^*)\left(\frac{1-p}{1-p_N^*}\right)\left(\frac{\Omega(p)}{\Omega(p_N^*)}\right)^{\mu/N} \tag{4}$$

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

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

This solution exhibits all of the familiar properties, which were described in Roth-

schild (1974) for one agent and in Bolton and Harris (1999) for several agents: the optimal strategy has a threshold where the agents change irrevocably from R to S; there are occasions where the agents make a mistake by changing from R to S although the risky action is actually better (R is good); the probability of mistakes decreases as the reward from the safe action decreases, as agents become more patient, and as the number of agents increases. Moreover, each player's payoff $V_N^*(p)$ increases in N over the range of beliefs where playing the risky arm is optimal.

The above proposition determines the *efficient* strategies. More precisely, we can distinguish two aspects of efficiency here. Given an action 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 much of the N units of the resource is allocated to risky arms at a given time t – we will call this number the *intensity* of experimentation. On the other hand, the integral $\int_0^T K_t\, dt$ measures how much of the resource is allocated to risky arms overall up to time T – we will call this number the *amount* of experimentation that is performed.

In the absence of a breakthrough, 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 3.1 *Suppose that there is no breakthrough and all experimentation ceases when the common belief decays to $p_c < p_0$. Then the amount of experimentation performed is $(\ln \Omega(p_c) - \ln \Omega(p_0))/\lambda$.*

Proof: With an intensity of experimentation K_t, the change in the belief is given by $dp_t = -K_t \lambda p_t(1-p_t)\, dt$ when no breakthrough occurs. Thus

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

Q.E.D.

With this lemma, Proposition 3.1 implies that the efficient amount of experimentation is $(\ln \Omega(p_N^*) - \ln \Omega(p_0))/\lambda$. 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 Strategic Problem

From now on, we assume that there are $N \geq 2$ players acting non-cooperatively. We consider stationary Markovian strategies with the common belief as the state variable. We describe the best response correspondence, present explicit representations of players' payoff functions, establish bounds on equilibrium payoffs, and show that all Markov equilibria are inefficient.

Best responses

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

Proceeding in the same way as in the previous section, we find that player n's value function satisfies the Bellman equation

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

Note that the second term on the right-hand side measures the benefit to player n of the information generated by the other players.

Player n's best response, k_n^*, is determined by comparing the opportunity cost of playing R with the expected *private* benefit:

$$k_n^* \begin{cases} = 0 & \text{if } c(p) > b(p, u_n), \\ \in [0,1] & \text{if } c(p) = b(p, u_n), \\ = 1 & \text{if } c(p) < b(p, u_n). \end{cases} \tag{5}$$

Using the Bellman equation together with the relationship between opportunity cost and expected benefit that defines each of the three cases in (5), we obtain the following alternative representation of best responses:

$$k_n^* \begin{cases} = 0 & \text{if } u_n(p) < s + K_{\neg n}\, c(p), \\ \in [0,1] & \text{if } u_n(p) = s + K_{\neg n}\, c(p), \\ = 1 & \text{if } u_n(p) > s + K_{\neg n}\, c(p). \end{cases} \tag{6}$$

Thus, player n's best response depends on whether in the (p, u)-plane, the point $(p, u_n(p))$ lies below, on or above the line

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

For $K_{\neg n} > 0$ this is a downward sloping diagonal that cuts the safe payoff line $u = s$ at $p = p^m$, the myopic cut-off; for $K_{\neg n} = 0$, it coincides with the safe payoff line. Note that the higher is $K_{\neg n} > 0$, the larger is the range of continuation payoffs $u_n(p)$ for which it is optimal to play S exclusively – this illustrates the incentive to free-ride on other players' experimentation efforts.

Explicit solutions for payoff functions

On intervals of beliefs where $K_{\neg n}$ is constant and player n has a unique best response, we can solve for the value function u_n explicitly up to a constant of integration. If $k_n^* = 1$ then u_n satisfies the ODE

$$K\lambda p(1-p)u'(p) + (r + K\lambda p)u(p) = (r + K\lambda)gp \tag{7}$$

with $K = K_{\neg n} + 1$.[2] The solution to (7) is

$$V_K(p) = gp + C\,(1-p)\,\Omega(p)^{\mu/K}. \tag{8}$$

If $k_n^* = 0$ then u_n satisfies

$$K\lambda p(1-p)u'(p) + (r + K\lambda p)u(p) = rs + K\lambda gp \tag{9}$$

with $K = K_{\neg n}$. The solution to (9) is

$$F_K(p) = s + \frac{K(g-s)}{\mu + K}p + C\,(1-p)\,\Omega(p)^{\mu/K}. \tag{10}$$

If the graphs of $F_{K_{\neg n}}$ and $V_{K_{\neg n}+1}$ meet $\mathcal{D}_{K_{\neg n}}$ at the 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.

Using the indifference condition from (5), $c(p) = b(p, u_n)$, we can also provide an explicit representation for a player's value function on any interval of beliefs where that player is indifferent between the two arms. On such an interval, u_n satisfies

$$\lambda p(1-p)u'(p) + \lambda p u(p) = (r + \lambda)gp - rs, \tag{11}$$

[2]Note that equation (7) for the strategic problem is the same ODE as that for the cooperative problem with an overall endowment of K units of the resource; cf. equation (1). This is because the arrival probability of a breakthrough is the same in both situations.

which has the (strictly convex) solution

$$W(p) = s + (\mu + 1)(g - s) + \mu s(1 - p) \ln \Omega(p) + C\,(1 - p). \tag{12}$$

Bounds on equilibrium payoffs

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

LEMMA 4.1 *In any Markov perfect equilibrium, 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 $b(p, V_1^*) \geq 0$, 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 following from $V_1^*(p') > u_n(p')$:

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

which is a contradiction. *Q.E.D.*

Inefficiency of Markov perfect equilibria

We can use the above results to show the following.

PROPOSITION 4.1 (INEFFICIENCY) *All Markov perfect equilibria of the N-player experimentation game are inefficient.*

PROOF: Since the average payoff lies between V_1^* and V_N^*, 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. *Q.E.D.*

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.

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.

It is easy to see from our characterization of best responses that there can be no symmetric equilibrium where all players always allocate the entire unit resource to one arm – there must be an interval of beliefs where players allocate a positive fraction of the resource to either arm. More precisely, the diagonal $\mathcal{D}_{N-1}$ separates the region where all players exclusively use the risky arm from the region where they use both arms; along this diagonal, we have smooth pasting of the relevant solutions W and V_N. Moreover, as we show, the equilibrium value function reaches the level s smoothly.

Proposition 5.1 (Symmetric equilibrium) *The N-player experimentation game has a unique symmetric equilibrium in Markovian strategies with the common posterior belief as the state variable. In this equilibrium, the safe arm is used exclusively at beliefs below the single-player cut-off p_1^*; the risky arm is used exclusively 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 the resource is allocated to each arm at beliefs strictly between p_1^* and $p_N^\dagger$. The fraction of the resource that each player allocates to the risky arm at such a belief is

$$k_N^\dagger(p) = \frac{1}{N-1}\frac{W^\dagger(p)-s}{c(p)} \tag{13}$$

with

$$W^\dagger(p) = s + \mu s\left[\Omega(p_1^*)\left(1 - \frac{1-p}{1-p_1^*}\right) - (1-p)\ln\left(\frac{\Omega(p_1^*)}{\Omega(p)}\right)\right], \tag{14}$$

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 (2)

with $V_N(p_N^\dagger) = W^\dagger(p_N^\dagger)$.

PROOF: Consider a function W that solves (11) below and to the left of $\mathcal{D}_{N-1}$. Lemma 4.1 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 (11), this implies $p_c \geq p_1^*$, so the only possibility is that the solution to (11) has $p_c = p_1^*$, which implies $W(p_1^*) = s$ and $W'(p_1^*) = 0$.

Using $W(p_1^*) = s$ in equation (12) determines the constant of integration C, giving the expression (14) for the value function over the range where both arms are used. Given this function, the expression (13) for the fraction $k_N^\dagger$ of the unit resource allocated to R follows from the indifference condition in (6) and the fact that $K_{\neg n} = (N-1)k_N^\dagger$ by symmetry. As $W^\dagger$ is strictly convex, $k_N^\dagger$ is strictly increasing to $+\infty$ as $p \uparrow p^m$. Thus there is a unique cut-off $p_N^\dagger < p^m$ where $k_N^\dagger(p_N^\dagger) = 1$. Simplifying $W^\dagger(p_N^\dagger) - s = (N-1)\, c(p_N^\dagger)$ gives the equation satisfied by $p_N^\dagger$. *Q.E.D.*

A number of 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 3.1, 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 discussed in the introduction, this means that we do not have the encouragement effect of Bolton and Harris (1999).

Third, the expected equilibrium payoff that each player obtains at beliefs where both arms are used does not depend on the number of players either; see equation (14). The reason for this is that the relevant ODE, equation (11), 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. 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 (the indifference diagonal $\mathcal{D}_{N-1}$ rotates clockwise as N increases).

6 Asymmetric Equilibria

We now turn to the behavior that can arise in asymmetric Markov perfect equilibria. We will focus our attention on equilibria where at each belief, each player allocates the entire unit resource to one of the arms, that is, where the action profile $(k_1, \ldots, k_N)$ is always an element of $\{0,1\}^N$. We will call these equilibria *simple*.[3]

We will present two types of simple MPE. The first type 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 simple 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.

Figure 1 illustrates the best response correspondence for $N = 2$, the faint straight line being $\mathcal{D}_1$ and the solid kinked line being the myopic payoff. From this figure we can see that a simple 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 most inequitable equilibrium, in which one

[3] It is not difficult to construct asymmetric equilibria that are not simple. Apart from their use in constructing bounds for aggregate payoffs (see footnote 4), these equilibria do not lead to additional insights, so we do not pursue them further here.

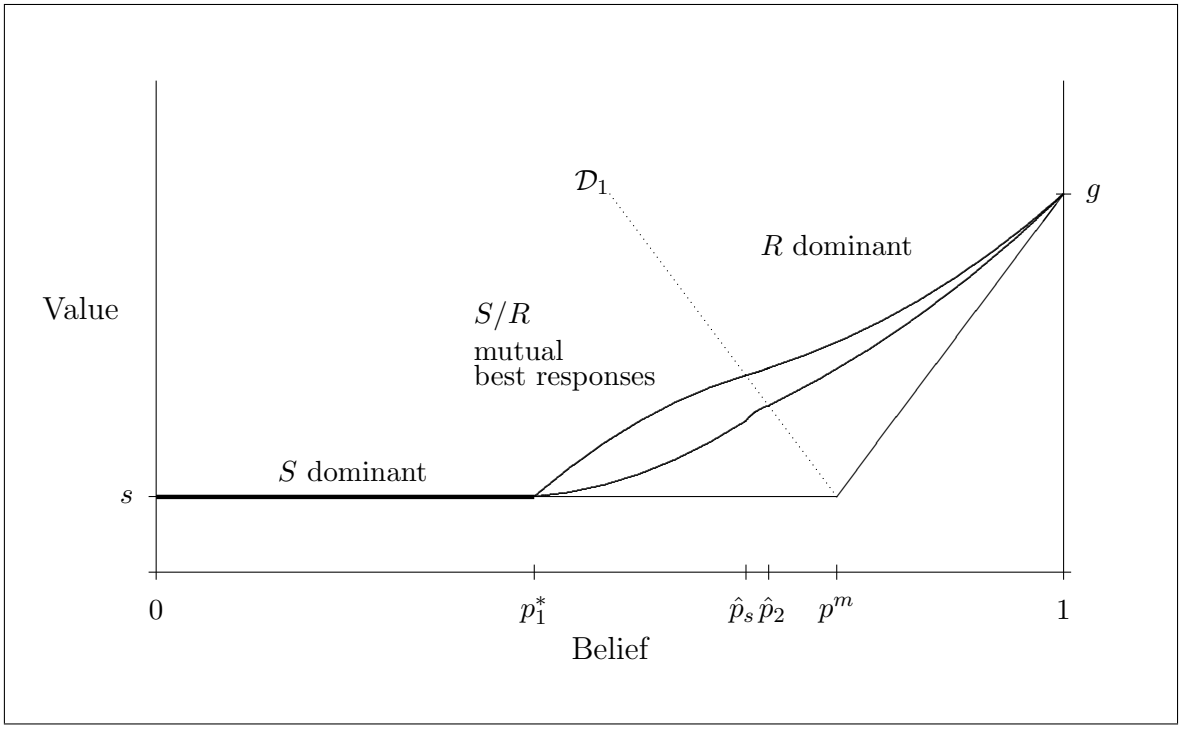

Figure 1: Best responses for $N = 2$, and payoffs in an asymmetric equilibrium

particular player experiments and the other free-rides throughout the lower region, and then characterizes all simple MPE where players' actions switch at finitely many beliefs.

Proposition 6.1 (Finite number of switches) *In the two-player experimentation game, there is a simple 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 S and player 2 plays R; on $(p_1^*, \hat{p}_s]$, player 1 plays R and player 2 plays S; on $[0, p_1^*]$, they both play S. The low cut-off, p_1^*, is given in Proposition 3.1; 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 simple MPE with finitely many switches there are two cut-offs, p_1^*

and $\bar{p}_2$, and a belief, $\bar{p}_s$, with $p_1^* < \hat{p}_s \leq \bar{p}_s \leq \bar{p}_2 \leq \hat{p}_2$, such that: on $(\bar{p}_2, 1]$, both players are above $\mathcal{D}_1$ and play R; throughout $(\bar{p}_s, \bar{p}_2]$, one player is above $\mathcal{D}_1$ and plays R, while the other is below $\mathcal{D}_1$ but above s and plays S; on $(p_1^*, \bar{p}_s]$, both players are below $\mathcal{D}_1$ but above s, exactly one plays R and one S, and they swap roles at least once; on $[0, p_1^*]$, they both play S.

PROOF: Here we just sketch the proof; for details, and a procedure how to construct all equilibria of this type, 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 1, 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 1 plays R and his continuation value (as a function of the belief) is a slowly rising convex function. On this interval, player 2 free-rides by playing S and her continuation value is a steeply rising concave function. Thus, at some belief, player 2's value meets $\mathcal{D}_1$ while player 1's value is still below it – this defines $\hat{p}_s$. On an interval to the right of $\hat{p}_s$, player 2 is content to be a pioneer and play R, while player 1 responds by free-riding with S. At some belief, player 1's value meets $\mathcal{D}_1$ while player 2's value is yet further above it – this defines $\hat{p}_2$. 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 lower value will be above that of player 1 in the inequitable equilibrium sketched above, and the one with the higher value will be below 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. *Q.E.D.*

The value functions of the two players in the equilibrium with cut-offs p_1^* and $\hat{p}_2$, and switch-point $\hat{p}_s$ are illustrated in Figure 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.

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:

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

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 $\mathcal{B}_0$;

(K) at least K players in $\mathcal{A}_{K-1}$, at least $N-K$ players in $\mathcal{B}_K$, for $0 < K < N$;

(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).

For $K \neq 0$, 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 $\mathcal{D}_{K-1}$.

For $K \neq N$, 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 $\mathcal{D}_K$.

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. *Q.E.D.*

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 most inequitable 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.

Figure 2 illustrates these actions for the most inequitable three-player equilibrium, 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. 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 3.1), 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 neighborhood of p_1^* – specifically when fewer than two players are experimenting – is the average payoff the same across all these equilibria.

The most inequitable equilibria described above are the worst from a welfare 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 most inequitable 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 simple 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: See the Appendix. *Q.E.D.*

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. The least upper bound on aggregate payoffs (in simple equilibria with finite switching) 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 $\mathcal{D}_{K-1}$. This is the same payoff as if each player allocated exactly K/N of his resource 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 equilibrium where the fraction of the resource that each player allocates to the risky arm falls gradually from 1 to 0 over the region below $\mathcal{D}_{N-1}$.[4]

[4]We can increase aggregate payoffs further if we consider non-simple equilibria, and the least upper bound on equilibrium payoffs is again given by a situation of payoff symmetry. It is the payoff that would obtain if the total intensity of experimentation were K on the entire region between $\mathcal{D}_{K-1}$ and $\mathcal{D}_{K-1+1/(K+1)}$, and increasing from K to $K+1$ on the region between $\mathcal{D}_{K-1+1/(K+1)}$ and $\mathcal{D}_K$, the incentive-compatible individual intensity being calculated in much the same way as we derived the expression (13) in the symmetric equilibrium. This is because just above $\mathcal{D}_{K-1}$ players using one arm exclusively produce more intense experimentation than is prescribed by symmetric equilibrium strategies, but less intense experimentation just below $\mathcal{D}_K$ – and exactly the same intensity on $\mathcal{D}_{K-1+1/(K+1)}$.

In particular, in an interval of beliefs above the single-agent cut-off where $W^\dagger$ is below $\mathcal{D}_{(N-1)/N}$, the intensity of experimentation in the symmetric equilibrium is lower than even in the most inequitable 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, and shows that there is a knock-on effect in a much wider interval of beliefs.

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

Proof: See the Appendix. *Q.E.D.*

The idea underpinning this result can be brought out by a closer comparison of the three-player symmetric equilibrium with a three-player simple equilibrium. If we were to superimpose a curve showing the intensity of experimentation in the symmetric equilibrium on the step function in Figure 2 for an asymmetric equilibrium, we would see that the curve connects $(p_1^*, 0)$ to $(p_3^\dagger, 3)$ (lying to the right of $(\bar{p}_3, 3)$), and almost goes through $(\bar{p}_2, 1)$ and $(\bar{p}_3, 2)$. Thus for a large majority of beliefs where some, but not all, of the aggregate resource is devoted to experimentation, the intensity of experimentation is higher in a simple equilibrium than in the symmetric equilibrium and the former is therefore more efficient.

Finally, note that for any $\epsilon > 0$ there is a simple MPE that Pareto-dominates the symmetric one at all beliefs in the interval $[p_1^* + \epsilon, p_N^\dagger]$. It is enough to take an equilibrium with players' payoff functions sufficiently close to the least upper bound identified above. Then, even the player who is the last one to experiment is better off at beliefs slightly above p_1^* as he benefits from some free-riding before running the last leg.

6.2 Infinitely many switches

Propositions 6.2 and 6.3 show that sharing the roles of free-rider and pioneer by alternating 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 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 above the cooperative cut-off and where we are in case (1), 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 (Infinite number of switches) *For each belief $p^{\ddagger}_{\infty}$ with $p^*_N < p^{\ddagger}_{\infty} < p^*_1$, the N-player experimentation game admits a simple Markov perfect equilibrium with infinitely many switches where at least one player experiments as long as the current belief is above $p^{\ddagger}_{\infty}$. More precisely, there exists a strictly decreasing sequence of beliefs $\{p^{\ddagger}_i\}_{i=1}^{\infty}$ with $p^{\ddagger}_1 \leq p^*_1$ and $\lim_{i\to\infty} p^{\ddagger}_i = p^{\ddagger}_{\infty}$ such that the equilibrium strategies at beliefs below $p^{\ddagger}_1$ can be specified as follows: player $n = 1, \ldots, N$ plays R at beliefs in the intervals $(p^{\ddagger}_{n+jN+1}, p^{\ddagger}_{n+jN}]$ $(j = 0, 1, \ldots)$ and S at all other beliefs below $p^{\ddagger}_1$.*

Proof: See the Appendix. *Q.E.D.*

The payoffs in a two-player equilibrium with an infinite number of switches are illustrated in Figure 3. (For clarity, we focus on beliefs between p^*_2 and $p^{\ddagger}_1$.) 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 $p^{\ddagger}_{\infty}$ is reached in finite time as the overall intensity of experimentation is bounded away from zero at beliefs above $p^{\ddagger}_{\infty}$. Note also that as we take $p^{\ddagger}_{\infty}$ closer to the efficient cut-off p^*_N, the amount of experimentation performed in equilibrium approaches the efficient amount; see Lemma 3.1. The intensity of experimentation, however, remains inefficient; it is 1 at beliefs in $(p^{\ddagger}_{\infty}, p^{\ddagger}_1]$, for instance, and therefore too low relative to the efficient benchmark.

7 Concluding Remarks

One extension of our model that we are actively researching is where a single breakthrough is not fully revealing, so that the encouragement effect identified by Bolton and

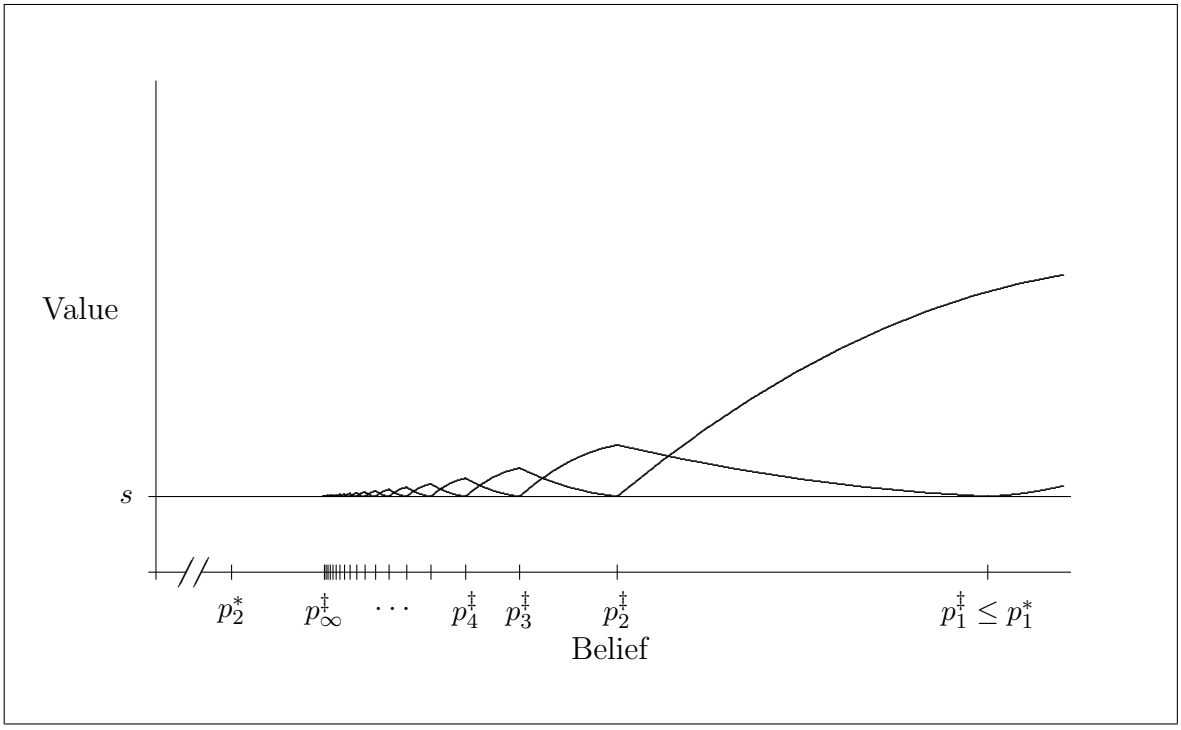

Figure 3: Payoffs in a two-player equilibrium with an infinite number of switches

Harris (1999) reappears. Such a model can 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.

Department of Economics, University of Oxford, Manor Road Building, Oxford, OX1 3UQ, U.K.; godfrey.keller@economics.ox.ac.uk,

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

and

Olin School of Business, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130, U.S.A.

Appendix

Before proving Proposition 6.1 we have two preliminary lemmas.

Let v and f satisfy equation (8) (pioneer, with $K=1$) and equation (10) (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 (8) (pioneer, with $K=1$) or equation (10) (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 (8) (both experimenting, $K=2$) 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 (8) *(pioneer, with $K=1$) 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* (10) *(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 (8) (pioneer, with $K=1$), 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 (10) (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 (8) (pioneer, with $K=1$) 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 (10) (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 $(p_\ell, p_r]$. *Q.E.D.*

Let u satisfy equation (8) (both experimenting, $K=2$) 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 (8) (pioneer, with $K = 1$) on $[\bar{p}, p^m]$ with $u_v(\bar{p}) = u(\bar{p})$. Inspection of equations (8) (pioneer, with $K=1$) and (8) (both experimenting, $K=2$) 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. *Q.E.D.*

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 (10) (free-rider, with $K = 1$) on some finite subpartition of $[p_1^*, \bar{p}]$ and equation (8) (pioneer, with $K=1$) on its finite complement, such that U meets $\mathcal{D}_1$ at $\bar{p}$; further, U satisfies equation (8) (both experimenting, $K = 2$) 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 (10) (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 neighborhood of 1, specifically $p \in (\bar{p}_2, 1]$, R is the dominant strategy; and for a belief in a neighborhood 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 most inequitable equilibrium is where one player, say player 1, plays R on $(p_1^*, \hat{p}_s]$, and the other player, player 2, plays S on this interval.[5] Then player 1's value function V_1 satisfies equation (8), and player 2's value function F_1 satisfies equation (10) with $V_1(p_1^*) = F_1(p_1^*) = s$. Lemma A.1 shows that V_1 meets $\mathcal{D}_1$ at a *larger* belief than does F_1, and that $V_1 < F_1$ on $(p_1^*, \hat{p}_s]$; that is, player 1 must be B and switch from playing S on $(\hat{p}_s, \hat{p}_2]$, and player 2 must

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

be A and switch from playing R 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 $F_{1,A}(\hat{p}_s) = s + c(\hat{p}_s)$; (4) C in $V_{1,A}$ from $V_{1,A}(\hat{p}_s) = F_{1,A}(\hat{p}_s) = s + c(\hat{p}_s)$; (5) C in $F_{1,B}$ from $F_{1,B}(\hat{p}_s) = V_{1,B}(\hat{p}_s)$; (6) $\hat{p}_2$ from $F_{1,B}(\hat{p}_2) = s + c(\hat{p}_2)$; (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 $V_{2,B}(\hat{p}_2) = F_{1,B}(\hat{p}_2) = s + c(\hat{p}_2)$. 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, $V_2'(p) > V_1'(p)$ if and only if $u > gp$ (the payoff from always playing R), we see that $0 < F_{1,A}'(p_1^*)$, $F_{1,A}'(\hat{p}_s) > V_{1,A}'(\hat{p}_s)$ and $V_{1,A}'(\hat{p}_2) < V_{2,A}'(\hat{p}_2)$, whereas $0 = V_{1,B}'(p_1^*)$, $V_{1,B}'(\hat{p}_s) < F_{1,B}'(\hat{p}_s)$ and $F_{1,B}'(\hat{p}_2) = V_{2,B}'(\hat{p}_2)$. 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).

Other equilibria for the two-player experimentation game

Take any finite 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 (10) on I_n (free-rider), X_n satisfies equation (8) on $I_{\neg n}$ (pioneer).

Define $\bar{p}_s = \min\{p \in [p_1^*, p^m] : X_1(p) \vee X_2(p) = s + c(p)\}$. If $X_1(\bar{p}_s) \geq X_2(\bar{p}_s)$ then $A = 1, B = 2$, else $A = 2, B = 1$. Let u_f satisfy equation (10) (free-rider) with $u_f(\bar{p}_s) = X_B(\bar{p}_s)$, and define $\bar{p}_2$ by $u_f(\bar{p}_2) = s + c(\bar{p}_2)$, 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 (10) on J_A (free-rider) and satisfies equation (8) on J_B (pioneer); Y_A satisfies equation (8) on $(\bar{p}_s, \bar{p}_2]$ (pioneer); Y_A satisfies equation (8) on $(\bar{p}_2, 1]$ (both experimenting). $Y_B(p) = s$ on $[0, p_1^*]$; Y_B satisfies equation (8) on J_A (pioneer) and satisfies equation (10) on J_B (free-rider); Y_B satisfies equation (10) on $(\bar{p}_s, \bar{p}_2]$ (free-rider); Y_B satisfies equation (8) 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 most inequitable equilibrium exhibits the slowest experimentation. *Q.E.D.*

Proof of Proposition 6.2

We first note that when everyone is playing either R or S exclusively, the average payoff satisfies

$$u(p) = s + Kb(p, u) - \tfrac{K}{N}c(p) \tag{A.1}$$

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 (A.1) follows directly from the developments in Section 3.) Consequently, u satisfies a convex combination of (7) and (9) with weightings K/N and $(N-K)/N$, and the solution is the corresponding combination of (8) and (10).

Consider a simple 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 (A.1) 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. *Q.E.D.*

Proof of Proposition 6.3

Consider a simple MPE with cut-offs $\{\bar{p}_K\}_{K=1}^N$. Let $u_{N,K}$ denote the solution to the ODE (A.1) 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 (A.1) 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 MPE, we are done.

To simplify notation define $R(p) = (\Omega(p)/\Omega(p_1^*))^\mu$, which is decreasing in p and less than 1

for $p > p_1^*$. Using a combination of equations (8) and (10), $u_{N|1}$ 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 equilibrium is given by the function $W^\dagger$ on $[p_1^*, p_N^\dagger]$. Using equation (14), 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. *Q.E.D.*

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 $(p_{n+jN+1}^\ddagger, p_{n+jN}^\ddagger]$ $(j = 0,1,\ldots)$ and S at all other beliefs below $p_1^\ddagger$.

For arbitrary i, consider the player 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 (8) (pioneer, with $K=1$) on the former interval, and equation (10) (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 $u(p^\ddagger_{i+1})$ from these equations, we change variables to $x_i = \Omega(p^\ddagger_i)/\Omega(p^m)$, noting that $\Omega(p^m) = (g-s)/s$ and

$$\frac{s - gp^\ddagger_i}{(1-p^\ddagger_i)s} = 1 - \frac{1}{x_i}.$$

This leads to the Nth 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

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

Writing $x_\infty = \Omega(p^\ddagger_\infty)/\Omega(p^m)$, we clearly have a steady state of this system at $(x_\infty, 0, \ldots, 0)$.

The Jacobian at this steady state is

$$\begin{pmatrix}
1 & x_\infty & 0 & 0 & \ldots & 0 \\
0 & 0 & 1 & 0 & \ldots & 0 \\
0 & 0 & 0 & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & 0 & \ldots & 1 \\
0 & \xi & -1 & -1 & \ldots & -1
\end{pmatrix}$$

with $\xi = \mu x_\infty - \mu - 1$. 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 $(-x_\infty/(1-\eta_*), 1, \eta_*, \eta_*^2, \ldots, \eta_*^{N-2})$. 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 neighborhood 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 strate-

gies 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 simple equilibria with a finite number of switches (see the outline after Lemma 6.1). *Q.E.D.*

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.

— and — (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.

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

Décamps, J.-P. and T. Mariotti (2003): "Investment Timing and Learning Externalities," forthcoming in *Journal of Economic Theory.*

Karlin, S. and H.M. Taylor (1981): *A Second Course in Stochastic Processes,* 2nd edition. New York: Academic Press.

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.

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

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.

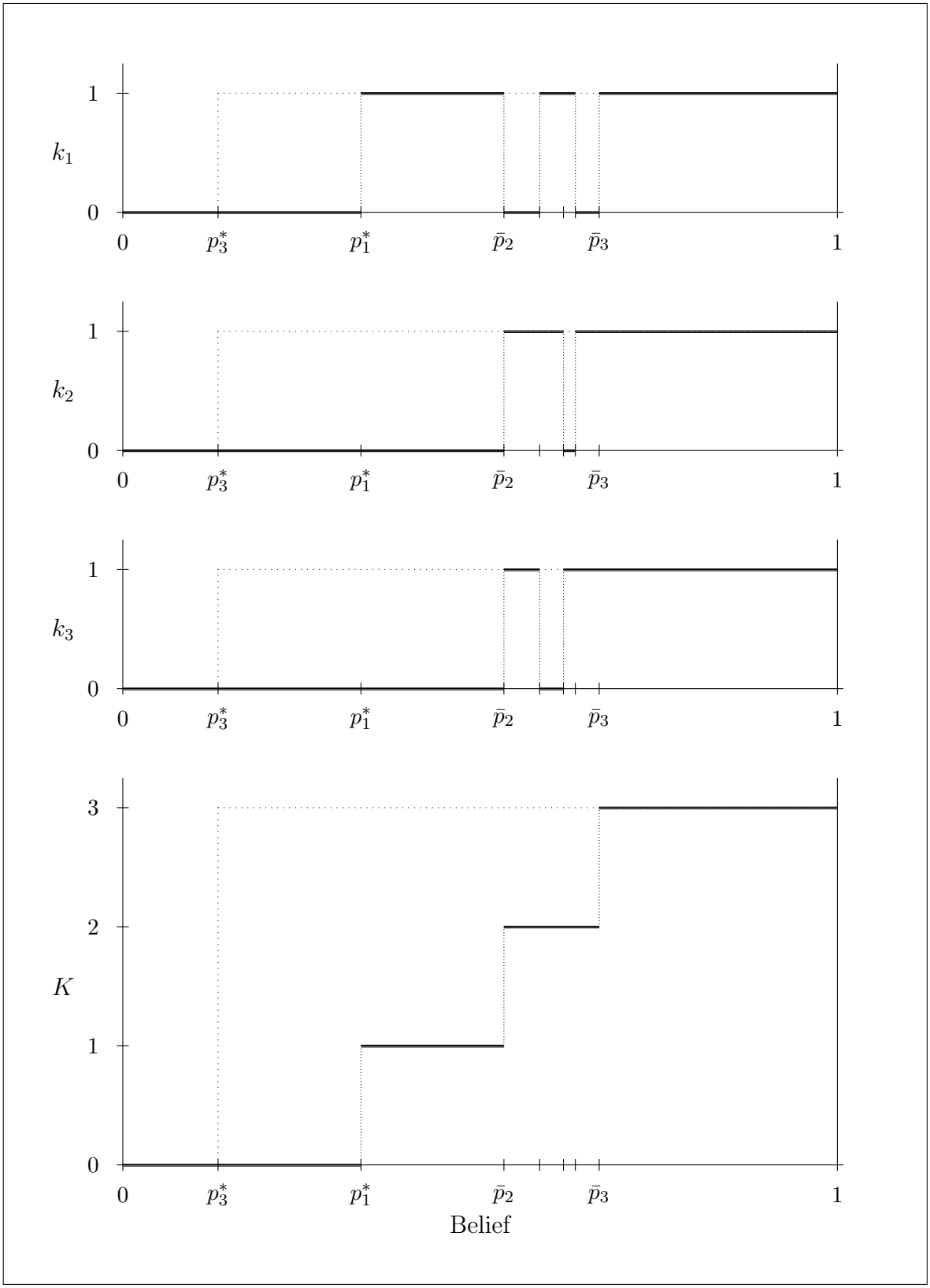

Figure 2: Action profiles in the most inequitable three-player asymmetric equilibrium