Econometrica, Vol. 67, No. 2 (March, 1999), 349–374

STRATEGIC EXPERIMENTATION

BY PATRICK BOLTON AND CHRISTOPHER HARRIS[1]

This paper extends the classic two-armed bandit problem to a many-agent setting in which N players each face the same experimentation problem. The main change from the single-agent problem is that an agent can now learn from the current experimentation of other agents. Information is therefore a public good, and a free-rider problem in experimentation naturally arises. More interestingly, the prospect of future experimentation by others encourages agents to increase current experimentation, in order to bring forward the time at which the extra information generated by such experimentation becomes available. The paper provides an analysis of the set of stationary Markov equilibria in terms of the free-rider effect and the encouragement effect.

KEYWORDS: Multi-agent two-armed bandit, informational public good, free-rider problem, encouragement effect.

1. INTRODUCTION

THIS PAPER ANALYSES A GAME of strategic experimentation in which individual players can learn from the experiments of others as well as their own. Given that experimentation typically entails an opportunity cost, and that information obtained from an experiment is valuable to all players, individual players attempt to free ride on the experiments of others. This informational externality drives a wedge between equilibrium experimentation and socially optimal experimentation. On the other hand, an individual player may be encouraged to experiment more if, by so doing, she can bring forward the time at which the information generated by the experimentation of others becomes available. This encouragement effect mitigates the free-rider effect. The objective of the paper is to analyze equilibrium experimentation strategies in terms of the free-rider and the encouragement effects.

The game of strategic experimentation we consider is a many-player common-value extension of the classic continuous-time two-armed bandit problem as presented from various points of view in Karatzas (1984), Berry and Fristedt (1985), and Mandelbaum (1987). In any given period of this game, each player must divide her time between the "safe" action and the "risky" action. The underlying payoff of the safe action is known and common to all players. The underlying payoff of the risky action is unknown but common to all players, and it can be either higher or lower than that of the safe action. The actual payoff obtained by a player from an action is the underlying payoff of that

[1]We would like to thank the Studienzentrum, Gerzensee, where a substantial part of the work reported in this paper was undertaken. We would also like to thank David Cox, Drew Fudenberg, Graham Hill, Meg Meyer, Khalid Sekkat, Neil Shephard, Hyun Shin, Richard Spady, Jean Tirole, and especially the co-editor and three anonymous referees for their help, and seminar participants at Berkeley, Cambridge, Chicago, ECARE, LSE, Minnesota, MIT, Northwestern, Oxford, Paris, Princeton, Stanford, Toulouse, UCLA, University College London, and Yale for their comments.

action plus noise. Once players have chosen how to allocate their time and the payoffs have been realized, all players observe all choices and all payoffs. They therefore obtain information about the underlying payoff of the risky action by observing the payoffs derived from the risky action.

In reality there are many situations in which a group of agents are involved in a game of strategic experimentation. Obvious examples are natural resource exploration, research and development, adoption of new technologies or products, identification of new investment opportunities, evaluation of the state of the economy, or consumer search. In fact, almost any situation of *social learning* can be represented as a game of strategic experimentation. Of course, most of these situations are more complex than our game. They may, for example, involve other forms of interaction among players besides the informational externality emphasized here. Also, other players' actions or payoffs may only be partially observable. Nevertheless, we believe that it is important to first develop a good understanding of the simplest form of interaction before proceeding to the analysis of more complex situations. Just as the single-agent two-armed bandit problem has become the backbone of single-agent active-learning theories, we believe that the multi-agent two-armed bandit problem is a natural benchmark for multi-agent active-learning theories.

The team-learning problem has a particularly simple solution in our setting: all players choose the risky action if and only if the common posterior probability p that the risky action has a higher underlying payoff than the safe action is above an endogenously determined cutoff c_*. Given that the opportunity cost of experimentation is divided by the number of players N, the team cutoff c_* is decreasing in N. When N tends to infinity, c_* converges to zero.

In analyzing equilibrium experimentation, we restrict attention to symmetric equilibria in stationary Markov strategies. We show that there exists a unique equilibrium which involves two cutoffs $\underline{c}_\dagger$ and $\bar{c}_\dagger$, where $0 < c_* < \underline{c}_\dagger < \bar{c}_\dagger < 1$. The proportion of time that players devote to experimentation is 0 when p falls below $\underline{c}_\dagger$, rises continuously from 0 to 1 as p rises from $\underline{c}_\dagger$ to $\bar{c}_\dagger$, and takes the value 1 when p exceeds $\bar{c}_\dagger$. Thus, because of free riding, equilibrium experimentation is less than socially efficient experimentation. As a consequence, the equilibrium payoff of a representative player is strictly less than the full-information payoff. This is even true in the limit as the number of players N tends to infinity.

We have chosen a continuous-time formulation of the two-armed bandit problem because of its tractability. The Bellman equation takes a particularly simple form in our formulation, and it can be used to establish extensive comparative-statics results for the unique symmetric equilibrium. For example, we are able to show that the equilibrium payoff is rising in the number of players. This result is an illustration of the encouragement effect. It is by no means obvious, since total experimentation itself is not monotonic in the number of players.

If one restricts attention to pure strategies, where players must devote any given period exclusively to one of the two actions available to them, then symmetric equilibria do not exist. (We do not analyze asymmetric equilibria here, but a characterization of the set of stationary Markov equilibria in an undiscounted version of our model can be found in Bolton and Harris (1996). See also the original version of the paper, namely Bolton and Harris (1993).) Thus, to obtain a symmetric equilibrium, one must allow players either to divide their time between the two actions in any given period, or to randomize in some way. We have chosen to focus on the divisible-time model. We do, however, discuss both private and public randomization towards the end of the paper. We argue there that the private-randomization model is isomorphic to the divisible-time model, and that any symmetric equilibrium where players exploit a public-randomization device to take turns in experimenting yields the same payoff as the symmetric private-randomization equilibrium.

To our knowledge only one paper, namely Smith (1991), considers a similar framework to ours. Smith focuses on limit beliefs and does not attempt to characterize socially optimal or equilibrium experimentation strategies. A number of recent papers deal with issues related to ours but in rather different settings. Some of these papers are discussed in Bolton and Harris (1996).

The paper is organized as follows. Section 2 describes the model. Section 3 derives the dynamics of p. Section 4 derives the Bellman equation for a player's best responses to a profile of strategies of the other players. Section 5 gives a partial characterization of symmetric equilibrium, and uses this characterization to obtain a detailed picture of a symmetric-equilibrium strategy. Section 6 exploits the encouragement effect to establish the existence and uniqueness of symmetric equilibrium. Section 7 sets out the comparative statics of the unique symmetric equilibrium. Section 8 considers the new indivisible-time game that is obtained from the original time-division game by requiring that players devote themselves exclusively to one or other action in any given period. It is argued there that the mixed extension of this indivisible-time game is isomorphic to the original game. All the results for the original game can therefore be reinterpreted as results for the mixed extension of the indivisible-time game. Section 9 considers the intuitively natural idea of taking turns to experiment. It is argued that this possibility is effectively covered by the analysis of mixed-strategy equilibrium.

The paper is deliberately written in a relatively informal way. The emphasis is on motivating the formulae that arise in the course of our analysis, and on explaining the intuition for our results. In particular, lemmas, theorems, and proofs serve simply to organize the discussion. Such an informal style may cause concern to some readers. We should therefore like to emphasize that the discussion of the present paper can be made fully rigorous by building on the material contained in the books by Jacod and Shiryayev (1988), Karatzas and Shreve (1988), and Krylov (1980).[2]

[2] The preceding draft of this paper, namely Bolton and Harris (1997), includes a mathematical appendix showing how the results of the present paper can be made rigorous. This draft is available on request from the authors.

2. THE TIME-DIVISION GAME

There are N identical infinitely lived risk-neutral players. At each time t, these players simultaneously and independently choose the proportion of the current period $[t, t+dt)$ to devote to each of the two actions available to them, namely 0 (the safe action) and 1 (the risky action). If player i chooses to devote a proportion α_i of the current period to the risky action, then she receives the total payoff

$$d\pi_i^0(t) = (1-\alpha_i)s\,dt + (1-\alpha_i)^{1/2}\sigma\,dZ_i^0(t)$$

from the safe action and the total payoff

$$d\pi_i^1(t) = \alpha_i\,\mu\,dt + \alpha_i^{1/2}\sigma\,dZ_i^1(t)$$

from the risky action. All players then observe all the proportions chosen and all the resulting payoffs. More explicitly, all players observe α_i, $d\pi_i^0$, and $d\pi_i^1$ for all $1 \le i \le N$. Here: s is fixed and known; $\mu \in \{l, h\}$ is unknown; $l < s < h$; and the $dZ_i^{a_i}(t)$ are independently and normally distributed with mean 0 and variance dt for $1 \le i \le N$, $a_i \in \{0,1\}$ and $t \in [0, \infty)$. Player i's objective is to maximize the expectation of the present discounted value of her payoff stream, namely $E[\int_0^\infty re^{-rt}(d\pi_i^0 + d\pi_i^1)(t)]$, where $r > 0$ is the discount rate.

Several features of this model are worthy of comment. First, $d\pi_i^0(t)$ is composed of the deterministic contribution $s\,dt$ and the stochastic shock $\sigma\,dZ_i^0(t)$. Since the contribution $s\,dt$ is known, it follows that $d\pi_i^0(t)$ conveys no information about μ. Similarly, $d\pi_i^1(t)$ is composed of the deterministic contribution $\mu\,dt$ and the stochastic shock $\sigma\,dZ_i^1(t)$. The first contribution ensures that $d\pi_i^1(t)$ conveys some information about μ. The second ensures that this information is noisy.

Secondly, if player i devotes a proportion α_i of the current period $[t, t+dt)$ to the risky action, then her total payoff $d\pi_i^0(t)$ from the safe action is distributed normally with mean $(1-\alpha_i)s\,dt$ and variance $(1-\alpha_i)\sigma^2 dt$, and her total payoff from the risky action is distributed normally with mean $\alpha_i\,\mu\,dt$ and variance $\alpha_i\sigma^2 dt$. These means and variances can be compared with the means and variances obtained when she devotes a proportion α_i of the periods $[t, t+dt)$ in the interval of time $[T, T+\Delta T)$ to the risky action, but devotes each period exclusively either to the safe action or to the risky action. More explicitly, suppose that the allocation of player i's time over the interval $[T, T+\Delta T)$ is determined by the function x_i: $[T, T+\Delta T) \to \{0, 1\}$, and put $\alpha_i = (1/\Delta T)\int_T^{T+\Delta T} x_i(t)\,dt$. Then the total payoff from the safe action over the interval $[T, T+\Delta T)$ is

$$\int_T^{T+\Delta T}(1-x_i(t))s\,dt + \int_T^{T+\Delta T}(1-x_i(t))\sigma\,dZ_i^0(t),$$

which is distributed normally with mean $(1-\alpha_i)s\Delta T$ and variance $(1-\alpha_i)\sigma^2\Delta T$, and the total payoff from the risky action over the interval $[T, T+\Delta T)$ is

$$\int_T^{T+\Delta T} x_i(t)\mu\,dt + \int_T^{T+\Delta T} x_i(t)\sigma\,dZ_i^1(t),$$

which is distributed normally with mean $\alpha_i\mu\Delta T$ and variance $\alpha_i\sigma^2\Delta T$. This explains the scaling used in the definition of $d\pi_i^0(t)$ and $d\pi_i^1(t)$.

Thirdly, by restricting players to choosing the actions 0 and 1, we obtain an indivisible-time version of our model. We shall argue informally in Section 8 below that the original time-division model is isomorphic to the mixed-action extension of this indivisible-time model.

3. THE FILTERING PROBLEM

We shall be concerned primarily with perfect equilibria in stationary Markov strategies. Such strategies depend only on the natural state variable for our problem, namely the players' common belief p that μ is high. In order to formulate the Bellman equations for equilibrium strategies, then, we need to determine how p evolves. This is the so-called filtering problem for our game.

Let $p(t)$ denote the prior belief that μ is high at time t, suppose that player i devotes a proportion α_i of the period $[t, t+dt]$ to the risky action, let $p(t+dt)$ denote the posterior belief that μ is high at time $t+dt$, and let $dp(t) = p(t+dt) - p(t)$ denote the change in beliefs over the period $[t, t+dt]$. Finally, let $\Phi(p) = (p(1-p)((h-l)/\sigma))^2$. Then we have the following lemma.

LEMMA 1: *Conditional on the information available to players at time t, the change in beliefs $dp(t)$ is distributed normally with mean* 0 *and variance* $(\sum_{i=1}^N \alpha_i)\Phi(p(t))\,dt$.

Note first that beliefs can be expected to follow a martingale. In other words, the expectation of $p(t+dt)$ conditional on current information should be $p(t)$. Or, equivalently, the expectation of $dp(t)$ conditional on current information should be 0. Lemma 1 confirms that this is indeed the case. Secondly, the better the information received about μ, the higher the variance of the posterior should be. In particular, the variance of the posterior should be higher the larger the total proportion $\sum_{i=1}^N \alpha_i$ of time devoted to the risky arm, and the higher the signal-to-noise ratio $(h-l)/\sigma$. Lemma 1 confirms these intuitions too. Finally, Lemma 1 makes clear that the posterior is unchanged from the prior when there is already certainty as to which state of the world obtains, i.e. whenever $p \in \{0, 1\}$.

PROOF: As we have pointed out above, players only derive information from the payoffs $d\pi_i^1(t)$. These payoffs are observationally equivalent to the signals $d\tilde{\pi}_i^1(t) = (\alpha_i)^{1/2}\tilde{\mu}dt + dZ_i^1(t)$, where $\tilde{\mu} = \mu/\sigma$.

Now $\tilde{\mu}$ takes the values $\tilde{l} = l/\sigma$ and $\tilde{h} = h/\sigma$ with probabilities $(1-p)$ and p, and the $dZ_i^1(t)$ are independently and normally distributed with mean 0 and variance dt. Hence, applying Bayes' Rule, we obtain

$$p(t+dt) = \frac{p(t)F(\tilde{h})}{p(t)F(\tilde{h}) + (1-p(t))F(\tilde{l})},$$

where $F(\tilde{\mu}) = (2\pi\, dt)^{-N/2} \exp(-(1/2\, dt)\Sigma_{i=1}^N (d\tilde{\pi}_i^1(t) - (\alpha_i)^{1/2}\tilde{\mu}\, dt)^2)$ is the probability of observing the payoff profile $d\tilde{\pi}^1(t) = \times_{i=1}^N d\tilde{\pi}_i^1(t)$ given $\tilde{\mu}$. Hence

$$dp = \frac{p(1-p)(\tilde{F}(\tilde{h}) - \tilde{F}(\tilde{l}))}{p\tilde{F}(\tilde{h}) + (1-p)\tilde{F}(\tilde{l})},$$

where $\tilde{F}(\tilde{\mu}) = \exp(\Sigma_{i=1}^N (\alpha_i)^{1/2}\tilde{\mu}\, d\tilde{\pi}_i^1 - 1/2\, \Sigma_{i=1}^N \alpha_i\, \tilde{\mu}^2\, dt)$, and where we have suppressed dependence on t. Moreover

$$\begin{aligned}
\tilde{F}(\tilde{\mu}) &= 1 + \left(\sum_{i=1}^N (\alpha_i)^{1/2} \tilde{\mu}\, d\tilde{\pi}_i^1 - \frac{1}{2} \sum_{i=1}^N \alpha_i\, \tilde{\mu}^2\, dt \right) \\
&\quad + \frac{1}{2} \left(\sum_{i=1}^N (\alpha_i)^{1/2} \tilde{\mu}\, d\tilde{\pi}_i^1 - \frac{1}{2} \sum_{i=1}^N \alpha_i\, \tilde{\mu}^2\, dt \right)^2 \\
&= 1 + \sum_{i=1}^N (\alpha_i)^{1/2} \tilde{\mu}\, d\tilde{\pi}_i^1,
\end{aligned}$$

where we have dropped terms of order $dt^{3/2}$ and higher, and where we have used the fact that $(d\tilde{\pi}_i^1)^2 = dt$ and $d\tilde{\pi}_i^1\, d\tilde{\pi}_j^1 = 0$ if $i \neq j$, respectively. Hence

$$\begin{aligned}
dp &= \frac{p(1-p)(\tilde{h} - \tilde{l})\Sigma_{i=1}^N (\alpha_i)^{1/2} d\tilde{\pi}_i^1}{1 + \Sigma_{i=1}^N (\alpha_i)^{1/2} \tilde{m}(p)\, d\tilde{\pi}_i^1} \\
&= p(1-p)(\tilde{h} - \tilde{l}) \left(\sum_{i=1}^N (\alpha_i)^{1/2}\, d\tilde{\pi}_i^1 \right) \left(1 - \sum_{i=1}^N (\alpha_i)^{1/2} \tilde{m}(p)\, d\tilde{\pi}_i^1 \right) \\
&= p(1-p)(\tilde{h} - \tilde{l}) \left(\sum_{i=1}^N (\alpha_i)^{1/2} d\tilde{\pi}_i^1 - \sum_{i=1}^N \alpha_i \tilde{m}(p)\, dt \right) \\
&= p(1-p)(\tilde{h} - \tilde{l}) \sum_{i=1}^N (\alpha_i)^{1/2} d\tilde{Z}_i^1,
\end{aligned}$$

where $\tilde{m}(p) = (1-p)\tilde{l} + p\tilde{h}$, where we have neglected terms of order $dt^{3/2}$ and higher, where we have noted once again that $(d\tilde{\pi}_i^1)^2 = dt$ and that $d\tilde{\pi}_i^1\, d\tilde{\pi}_j^1 = 0$ if $i \neq j$, and where $d\tilde{Z}_i^1 = d\tilde{\pi}_i^1 - \alpha_i \tilde{m}(p)\, dt$, respectively. Finally, the expectation of $d\tilde{Z}_i^1$ conditional on the information available to the players at time t is 0, and $d\tilde{Z}_i^1\, d\tilde{Z}_j^1 = dt$ if $i = j$ and $d\tilde{Z}_i^1\, d\tilde{Z}_j^1 = 0$ if $i \neq j$. That is, the profile $\tilde{Z}^1 = \times_{i=1}^N \tilde{Z}_i^1$ follows a standard N-dimensional Wiener process relative to the players' information. Hence dp has mean 0 and variance $(p(1-p)(\tilde{h}-\tilde{l}))^2(\sum_{i=1}^N \alpha_i)\,dt$. Recalling that $\tilde{l}=l/\sigma$ and $\tilde{h}=h/\sigma$, we obtain the required conclusion. *Q.E.D.*

4. BEST RESPONSES

In this section we analyze a player's best response to a profile of Markov strategies of the other players.

DEFINITION 2: A *Markov strategy* for player i is a mapping ξ_i: $[0,1]\to[0,1]$.[3]

In what follows we shall refer to Markov strategies simply as strategies wherever this will not lead to any misunderstanding.

Let $m(p)=(1-p)l+ph$ be the expectation of the flow payoff from the risky arm when μ is believed to be h with probability p. Then we have the following characterization of a player's value function.

LEMMA 3: *Suppose that the players $j\neq i$ employ the strategies ξ_j. Then player i's value function u_i: $[0,1]\to[l,h]$ is the unique solution of the Bellman equation*

$$u_i(p)=\max_{\alpha_i\in[0,1]}\left((1-\alpha_i)s+\alpha_i m(p)+\frac{1}{r}\left(\alpha_i+\sum_{j\neq i}\xi_j(p)\right)\Phi(p)\frac{u_i''(p)}{2}\right) \tag{1}$$

for all $p\in[0,1]$. In particular, $u_i(0)=s$ and $u_i(1)=h$.[4]

In a discrete-time setting, the Bellman equation states that the current payoff is equal to the maximum over the control variable of the expectation of the current flow payoff plus the expectation of the discounted value of the continuation payoff. In the present, continuous-time, setting: $u_i(p)$ is the current payoff; α_i is the control variable; $(1-\alpha_i)s+\alpha_i m(p)$ is the expectation of the current flow payoff; $1/r$ is the discount factor; and $(\alpha_i+\sum_{j\neq i}\xi_j(p))\Phi(p)(u_i''(p)/2)$ is the expectation of the rate of change of the continuation payoff. The Bellman equation therefore states that the current payoff is the maximum over the control variable of the expectation of the current flow payoff plus the discounted value of the *rate of change* of the continuation payoff.

[3] More precisely, a Markov strategy for player i is an equivalence class of Lebesgue measurable mappings ξ_i: $[0,1]\to[0,1]$.

[4] The solution of the Bellman equation is located in $\mathscr{C}([0,1])\cap\mathscr{W}^{2,1}([0,1])$. Here: $\mathscr{C}([0,1])$ is the space of continuous functions from $[0,1]$ to $\mathbb{R}$: $\mathscr{W}^{2,1}([0,1])$ is the Sobolev space of generalized functions from $[0,1]$ to $\mathbb{R}$ which, together with their first and second generalized derivatives, lie in $\mathscr{L}^1([0,1])$; and $\mathscr{L}^1([0,1])$ is the space of equivalence classes of Lebesgue measurable functions from $[0,1]$ to $\mathbb{R}$. Equation (1) is required to hold almost everywhere.

PROOF: Suppose that the current belief is $p \in [0,1]$, that player i devotes a proportion $\alpha_i \in [0,1]$ of the current period to the risky action, and that her continuation payoffs are given by c_i: $[0,1] \to [l,h]$. Then her current flow payoff will be $r(d\pi_i^0 + d\pi_i^1)$, and her continuation payoff will be $e^{-r\,dt}c_i(p+dp)$. Now $E[\mu] = m(p)$. Hence the expectation of player i's current flow payoff is

(2) $$r((1-\alpha_i)s + \alpha_i m(p))\,dt.$$

Moreover $E[dp] = 0$ and $E[(dp)^2] = (\alpha_i + \sum_{j \neq i} \xi_j(p))\Phi(p)\,dt$. Hence, neglecting terms of order $dt^{3/2}$ and higher, we have $e^{-r\,dt} = 1 - r\,dt$ and $c_i(p+dp) = c_i(p) + c_i'(p)dp + \frac{1}{2}c_i''(p)(dp)^2$. Hence the expectation of player i's continuation payoff is

(3) $$(1 - r\,dt)\left(c_i(p) + \frac{1}{2}c_i''(p)\left(\alpha_i + \sum_{j \neq i} \xi_j(p)\right)\Phi(p)\,dt\right).$$

Adding (2) and (3), and dropping terms of order $dt^{3/2}$ and higher, we obtain the expectation of her current payoff, namely

$$H_i(\alpha_i, p, c_i, \xi_{-i}) = c_i(p) + r\,dt\Bigg((1-\alpha_i)s + \alpha_i m(p) + \frac{1}{r}\left(\alpha_i + \sum_{j \neq i} \xi_j(p)\right)\Phi(p)\frac{c_i''(p)}{2} - c_i(p)\Bigg).$$

Finally, her value function u_i is, as usual, the unique bounded solution of the Bellman equation

$$u_i(p) = \max_{\alpha_i \in [0,1]} H_i(\alpha_i, p, c_i, \xi_{-i}) \quad \text{for all} \quad p \in [0,1].$$

It is easy to see that this equation reduces to (1). In particular, $u_i(0) = s$ and $u_i(1) = h$ since $\Phi(0) = \Phi(1) = 0$. *Q.E.D.*

The Bellman equation also tells us how to choose the optimal policy ξ_i.

LEMMA 4: *The strategy ξ_i is a best response to the strategy profile $\xi_{-i} = \times_{j \neq i} \xi_j$ iff*

(4) $$\xi_i(p) \in \operatorname*{argmax}_{\alpha_i \in [0,1]} \Bigg((1-\alpha_i)s + \alpha_i m(p) + \frac{1}{r}\left(\alpha_i + \sum_{j \neq i} \xi_j(p)\right)\Phi(p)\frac{u_i''(p)}{2}\Bigg)$$

for all $p \in [0,1]$. In particular, $\xi_i(0) = 0$ *and* $\xi_i(1) = 1$.[5]

[5] Relation (4) is required to hold almost everywhere.

In other words, ξ_i is a best response iff

$$\xi_i(p)\left\{\begin{array}{ll} =0 & \text{if } \frac{1}{r}\Phi(p)\frac{u_i''(p)}{2} < s-m(p) \\ \in[0,1] & \text{if } \frac{1}{r}\Phi(p)\frac{u_i''(p)}{2} = s-m(p) \\ =1 & \text{if } \frac{1}{r}\Phi(p)\frac{u_i''(p)}{2} > s-m(p) \end{array}\right\}.$$

Notice that: $1/r$ is the discount factor; $\Phi(p)$ is the amount of information revealed by any given experiment; and $u_i''(p)/2$ is the shadow value of information. We may therefore interpret $(1/r)\Phi(p)(u_i''(p)/2)$ as the shadow value of experimentation. Also, $s-m(p)$ is the opportunity cost of experimentation. The Bellman equation therefore tells us that we should maximize experimentation if the shadow value of experimentation exceeds its opportunity cost, and minimize experimentation otherwise.

PROOF: The strategy ξ_i is a best response for player i iff it maximizes the expectation of her overall payoff when her continuation payoffs are given by her value function u_i, i.e. iff $\xi_i(p)\in \operatorname{argmax}_{\alpha_i\in[0,1]} H_i(\alpha_i, p, u_i, \xi_{-i})$ for all $p\in[0,1]$. In particular, $\xi_i(0)=0$ and $\xi_i(1)=1$ since $\Phi(0)=\Phi(1)=0$. *Q.E.D.*

5. SYMMETRIC EQUILIBRIA: CHARACTERIZATION

In this section we obtain a partial characterization of symmetric equilibrium. This characterization will be used in the present section to derive properties of equilibrium strategies from properties of the associated equilibrium value functions. It will be helpful to begin with an analysis of the team problem.

5.1. *The Team Problem*

In the team problem, the social planner chooses the profile $\alpha = \times_{i=1}^{I}\alpha_i \in [0,1]^N$ at each moment in time in such a way as to maximize the average payoff of the players. Arguing as in Section 4, we arrive at the following formulation for the Bellman equation for the team problem.

LEMMA 5: *The value function $u_*:[0,1]\to[l,h]$ for the N-player team problem is the unique solution of the Bellman equation*

$$(5) \qquad u_* = \max_{\alpha\in[0,1]^N}\left(\frac{1}{N}\left(\left(N-\sum_{i=1}^{N}\alpha_i\right)s+\left(\sum_{i=1}^{N}\alpha_i\right)m\right)+\frac{1}{r}\left(\sum_{i=1}^{N}\alpha_i\right)\Phi\frac{u_*''}{2}\right),$$

where we have suppressed the dependence of u_*, m, and Φ on p. Moreover the strategy profile $\xi = \times_{i=1}^{N} \xi_i$ is an optimal policy for the team problem iff

$$(6) \qquad \xi \in \operatorname*{argmax}_{\alpha \in [0,1]^N} \left(\frac{1}{N} \left(\left(N - \sum_{i=1}^{N} \alpha_i \right) s + \left(\sum_{i=1}^{N} \alpha_i \right) m \right) + \frac{1}{r} \left(\sum_{i=1}^{N} \alpha_i \right) \Phi \frac{u''_*}{2} \right),$$

where we have suppressed the dependence of ξ, m, u_* and Φ on p.[6]

Equation (5) can be solved explicitly for u_*. Let $\zeta = \sqrt{1 + \left(8r\sigma^2/N(h-l)^2\right)}$, let $z(p) = p^{-(\zeta-1)/2}(1-p)^{(\zeta+1)/2}$, let $c_* = (\zeta-1)(s-l)/((\zeta-1)(s-l)+(\zeta+1)(h-s))$, let

$$u_*(p) = \begin{cases} s & \text{for } p \in [0, c_*] \\ m(p) + \dfrac{s - m(c_*)}{z(c_*)} z(p) & \text{for } p \in (c_*, 1] \end{cases},$$

and let

$$\xi_*(p) = \begin{cases} \times_{i=1}^{N} 0 & \text{for } p \in [0, c_*] \\ \times_{i=1}^{N} 1 & \text{for } p \in (c_*, 1] \end{cases}.$$

Then obviously $u_* = s$ on $[0, c_*]$, and it is easy to check by explicit calculation that $u_* = m + (1/r)N\Phi(u''_*/2)$ on $(c_*, 1]$. That is, u_* is the payoff function associated with the policy ξ_*. Moreover $(1/r)\Phi(u''_*/2) = 0 < (s-m)/N$ on $[0, c_*]$ and $(1/r)\Phi(u''_*/2) = (u_* - m)/N > (s-m)/N$ on $(c_*, 1]$. That is, ξ_* is optimal when continuation payoffs are given by u_*. Overall, then, u_* solves (5) and ξ_* satisfies (6).

We see from this explicit solution that optimal experimentation in the team problem takes a particularly simple form: all players choose the safe action when $p \in [0, c_*)$; and all players choose the risky action when $p \in (c_*, 1]$.

5.2. *Best Responses*

First we establish some properties of player i's value function when she plays a best response. Let $\underline{u}(p) = \max\{s, m(p)\}$ denote the myopic payoff, let $\bar{u}(p) = (1-p)s + ph$ denote the full-information payoff, and let b denote the break-even point at which $m = s$.

LEMMA 6: *Suppose that player i plays a best response to the strategy profile $\xi_{-i} = \times_{j \neq i} \xi_j$. Let u_i denote her value function. Then $\underline{u} \leq u_i \leq \bar{u}$ and $u''_i \geq 0$ on $[0,1]$.*

[6] For a more precise interpretation of equation (5) and relation (6), see footnotes 4 and 5 above.

PROOF: Note first that one feasible strategy for player i is to choose the safe action when $p \in [0, b]$ and to choose the risky action when $p \in (b, 1]$. This strategy yields the flow payoff $\underline{u}$. Denote the overall payoff to which it gives rise by $\hat{u}$. Then clearly $u_i \geq \hat{u}$. Moreover $\hat{u} \geq \underline{u}$ by Jensen's inequality. Hence $u_i \geq \underline{u}$ as required.

Secondly, any strategy that can be played when there is incomplete information can also be played when there is complete information. Hence u_i must be less than or equal to the full-information payoff $\bar{u}$.

Thirdly, the Bellman equation (14) holds iff

$$u_i \geq m + \frac{1}{r}\left(1 + \sum_{j \neq i} \xi_j\right)\Phi \frac{u_i''}{2} \quad \text{and} \quad u_i \geq s + \frac{1}{r}\left(\sum_{j \neq i} \xi_j\right)\Phi \frac{u_i''}{2}$$

with at least one equality. If the first inequality holds as an equality, then we have

$$\frac{1}{r}\Phi \frac{u_i''}{2} = \frac{u_i - m}{1 + \sum_{j \neq i} \xi_j} \geq 0.$$

Similarly, if the second inequality holds as an equality and $\sum_{j \neq i} \xi_j > 0$, then

$$\frac{1}{r}\Phi \frac{u_i''}{2} = \frac{u_i - s}{\sum_{j \neq i} \xi_j} \geq 0.$$

Finally, if the second inequality holds as an equality and $\sum_{j \neq i} \xi_j = 0$, then $u_i = s$. Since $u_i \geq s$ on $[0, 1]$, u_i attains a minimum in this case. Hence $u_i'' \geq 0$. Overall, then, $u_i'' \geq 0$ on $[0, 1]$. *Q.E.D.*

Next, using the fact that $u_i'' \geq 0$—i.e. the shadow value of information is nonnegative—we can show that player i's payoff is nondecreasing in the total experimentation of the other players.

LEMMA 7: *Let $\xi_{-i} = \times_{j \neq i} \xi_j$ and $\hat{\xi}_{-i} = \times_{j \neq i} \hat{\xi}_j$ be strategy profiles of the other players, and let u_i and $\hat{u}_i$ be the value functions of player i when she plays best responses to ξ_{-i} and $\hat{\xi}_{-i}$ respectively, let $\Xi_i = \sum_{j \neq i} \xi_j$ and $\hat{\Xi}_i = \sum_{j \neq i} \hat{\xi}_j$, and suppose that $\Xi_i \geq \hat{\Xi}_i$. Then $u_i \geq \hat{u}_i$.*

PROOF: The value function u_i solves the Bellman equation

$$u_i = \max_{\alpha_i \in [0,1]} \left((1 - \alpha_i)s + \alpha_i m + \frac{1}{r}(\alpha_i + \Xi_i)\Phi \frac{u_i''}{2}\right).$$

Hence

$$u_i \geq \max_{\alpha_i \in [0,1]} \left((1 - \alpha_i)s + \alpha_i m + \frac{1}{r}\left(\alpha_i + \hat{\Xi}_i\right)\Phi \frac{u_i''}{2}\right).$$

Comparing this inequality with the Bellman equation,

$$\hat{u}_i = \max_{\alpha_i \in [0,1]} \left((1-\alpha_i)s + \alpha_i m + \frac{1}{r}\left(\alpha_i + \hat{\Xi}_i\right)\Phi\frac{\hat{u}_i''}{2} \right)$$

for $\hat{u}_i$, we see that $u_i \geq \hat{u}_i$. *Q.E.D.*

Finally, it turns out that there is a particularly simple way of choosing a best response for player i. We present this result here since it provides significant insight into the nature of the strategic interaction among the players. We do not prove it since its proof is similar to that of Theorem 9, and since it will not be used below.

THEOREM 8: *Suppose that a strategy profile $\xi_{-i} = \times_{j \neq i} \xi_j$ is given. Let u_i denote the value function of player i when she plays a best response to ξ_{-i}, let $\Xi_i = \Sigma_{j \neq i} \xi_j$, and let*

$$\beta_i = \frac{u_i - s}{s - m} - \Xi_i$$

be the private incentive to experiment. Then the strategy

$$\xi_i = \begin{Bmatrix} 0 & \text{if } \beta_i \leq 0 \text{ and } p < b \\ 1 & \text{if } \beta_i > 0 \text{ or } p \geq b \end{Bmatrix}$$

is a best response of ξ_{-i}. *Q.E.D.*

Combining Lemma 7 and Theorem 8, we see that increasing Ξ_i has two effects: it leads directly to a reduction in β_i via the dependence of β_i on Ξ_i; and it leads to an increase in u_i, and thereby indirectly to an increase in β_i via the dependence of β_i on u_i. The first effect is the free-rider effect. The second effect is the encouragement effect.

The free-rider effect is easy to explain: extra current experimentation by the other players provides player i with information at no cost, and this information is used as a substitute for information that she would otherwise have had to supply for herself at the opportunity cost $s-m$. As for the encouragement effect: extra future experimentation by the other players encourages player i to increase her current experimentation in order to bring forward the time at which the extra information generated by the other players becomes available.

An illustrative example may help to get a better grasp of the encouragement effect. Suppose that $N = 2$, that player 2 initially undertakes no experimentation, and that player 1 optimizes. Let c_{*1} denote the experimentation cutoff for the 1-player team problem. Then player 1 will experiment if $p > c_{*1}$, will not experiment if $p < c_{*1}$, and is indifferent between experimenting and not experimenting when $p = c_{*1}$. Suppose now that player 2 decides to experiment whenever $p > b$. Then player 1 has a strict incentive to experiment when $p = c_{*1}$. Indeed, if she does not experiment, then the belief p will not move, and she will never benefit from the information that experimentation by player 2 could provide. By the same token, if she does experiment now, and if the outcome of her experiment is favorable, then player 2 may be induced to join in the experimentation.

In summary, then, current experimentation by one player is a strategic substitute for current experimentation by another, but future experimentation by one player is a strategic complement for current experimentation by another. In particular, our game is not a supermodular game. Indeed, even in a two-player game, increasing the level of experimentation of one player in all states may lead the other player to increase experimentation in some states, and to decrease experimentation in other states.

5.3. *A Characterization of Symmetric Equilibrium*

We turn now to the characterization of equilibrium experimentation. Since the case $N = 1$ has already been treated in Section 5.1, we confine ourselves to the case $N \geq 2$.

THEOREM 9: *Let $\xi_\dagger$ be a strategy, let $u_\dagger$ be a player's value function of player when she plays a best response to the strategy profile $\times_{j \neq i} \xi_\dagger$, and let*

$$\beta_\dagger = \frac{u_\dagger - s}{s - m}$$

be the collective incentive to experiment. Then $\xi_\dagger$ is a symmetric equilibrium iff

$$\xi_\dagger = \begin{cases} \dfrac{\beta_\dagger}{N-1} & \text{if } \beta_\dagger \leq N-1 \text{ and } p < b \\ 1 & \text{if } \beta_\dagger > N-1 \text{ or } p \geq b \end{cases}. \tag{7}$$

PROOF: In view of Lemmas 3 and 4, $\xi_\dagger$ is a symmetric equilibrium iff

$$\xi_\dagger \in \operatorname*{argmax}_{\alpha_i \in [0,1]} \left((1 - \alpha_i)s + \alpha_i m + \frac{1}{r}\big(\alpha_i + (N-1)\xi_\dagger\big)\Phi \frac{u''_\dagger}{2} \right), \tag{8}$$

where $u_\dagger$ is the unique solution of the equation

$$u_\dagger = \max_{\alpha_i \in [0,1]} \left((1 - \alpha_i)s + \alpha_i m + \frac{1}{r}\big(\alpha_i + (N-1)\xi_\dagger\big)\Phi \frac{u''_\dagger}{2} \right). \tag{9}$$

Suppose that $\xi_\dagger$ is a symmetric equilibrium. If $p < b$, then there are three possibilities. First, if $(1/r)\Phi(u''_\dagger/2) < s - m$ then $\xi_\dagger = 0$ from (8). Hence $u_\dagger = s$ from (9), and $\beta_\dagger = (u_\dagger - s)/(s - m) = 0$. Secondly, if $(1/r)\Phi(u''_\dagger/2) = s - m$, then

$$\begin{aligned} u_\dagger &= (1 - \xi_\dagger)s + \xi_\dagger m + \big(\xi_\dagger + (N-1)\xi_\dagger\big)(s - m) \\ &= s + (N-1)\xi_\dagger(s - m) \end{aligned}$$

from (9). Hence

$$\xi_\dagger = \frac{1}{N-1}\left(\frac{u_\dagger - s}{s-m}\right) = \frac{\beta_\dagger}{N-1}.$$

Thirdly, if $(1/r)\Phi(u''_\dagger/2) > s - m$, then $\xi_\dagger = 1$ from (8) and $u_\dagger = m + (1/r)N\Phi(u''_\dagger/2)$. Hence

$$\beta_\dagger = \frac{u_\dagger - s}{s-m} = \frac{\frac{1}{r}N\Phi\frac{u''_\dagger}{2} - (s-m)}{s-m} > N-1.$$

On the other hand, if $p \geq b$, then $s - m \leq 0$ and $u''_\dagger \geq 0$ with at least one strict inequality. Indeed: $s - m \leq 0$ for all $p \geq b$; $s - m < 0$ for all $p > b$; $u''_\dagger \geq 0$ for all p by Lemma 6; and $u''_\dagger(b) > 0$ since if $u''_\dagger(b) = 0$ then we would also have $u_\dagger(b) = \underline{u}(b)$ by (9), and Lemma 6 would then imply that $u_\dagger$ had an upward kink at b. Hence $\xi_\dagger = 1$ from (8). In all cases, then, $\xi_\dagger$ satisfies (7).

Suppose now that $\xi_\dagger$ satisfies (7), where $u_\dagger$ is the unique solution of (9). If $p < b$ then there are three possibilities. First, if $(1/r)\Phi(u''_\dagger/2) < s - m$, then $u_\dagger = s + (1/r)(N-1)\xi_\dagger\Phi(u''_\dagger/2)$ from (9). Hence

$$\begin{aligned}\xi_\dagger &= \min\left\{\frac{\beta_\dagger}{N-1}, 1\right\} = \min\left\{\frac{1}{N-1}\left(\frac{u_\dagger - s}{s-m}\right), 1\right\} \\ &= \min\left\{\frac{\frac{1}{r}\xi_\dagger\Phi\frac{u''_\dagger}{2}}{s-m}, 1\right\} = \left(\frac{\frac{1}{r}\Phi\frac{u''_\dagger}{2}}{s-m}\right)\xi_\dagger.\end{aligned}$$

Hence $\xi_\dagger = 0$. Secondly, if $(1/r)\Phi(u''_\dagger/2) = s - m$, then $\xi_\dagger$ automatically satisfies (8). Thirdly, if $(1/r)\Phi(u''_\dagger/2) > s - m$, then $u_\dagger = m + (1/r)(1 + (N-1)\xi_\dagger)\Phi(u''_\dagger/2)$ from (9). Hence

$$\beta_\dagger = \frac{u_\dagger - s}{s-m} = \frac{\frac{1}{r}\big(1 + (N-1)\xi_\dagger\big)\Phi\frac{u''_\dagger}{2} - (s-m)}{s-m} > (N-1)\xi_\dagger.$$

But $\xi_\dagger = \min\{\beta_\dagger/(N-1), 1\}$. It follows that $\xi_\dagger = 1$. On the other hand, if $p \geq b$, then $s - m \leq 0$, and Lemma 6 implies that $u''_\dagger \geq 0$ with at least one strict inequality. In all cases, then, $\xi_\dagger$ satisfies (8). *Q.E.D.*

5.4. *Properties of Equilibrium*

By combining Lemma 6 and Theorem 9, we can build up a detailed picture of symmetric equilibrium. Once again, we confine ourselves to the case $N \geq 2$.

We begin with a lemma. Recall that u_* denotes the value function of the N-player team problem, and let u_{*1} denote the value function of the 1-player team problem.

LEMMA 10: *Let $u_{\dagger}$ be the value function of a symmetric equilibrium. Then* $u_{*1} \leq u_{\dagger} \leq u_{*}$ *on* $[0,1]$.

PROOF: The inequality $u_{\dagger} \geq u_{*1}$ follows at once from Lemma 7. As for the inequality $u_{\dagger} \leq u_{*}$, note that the average payoff in equilibrium is $u_{\dagger}$, and that the maximum achievable average payoff is u_{*}. *Q.E.D.*

Let c_{*1} denote the experimentation cutoff for the one-player team problem as before. Then, combining Lemmas 6 and 10, we see that there is a unique cutoff $\underline{c}_{\dagger} \in [c_{*}, c_{*1}] \subset (0,b)$ such that $\beta_{\dagger} = 0$ on $[0, \underline{c}_{\dagger}]$; that $\beta_{\dagger}$ increases strictly from 0 to $+\infty$ as p increases from $\underline{c}_{\dagger}$ to b; and that $\beta_{\dagger}$ is convex on $[0,b)$. In particular, there is a unique cutoff $\bar{c}_{\dagger} \in (\underline{c}_{\dagger}, b)$ such that $\beta_{\dagger} = N-1$ at $\bar{c}_{\dagger}$. Theorem 9 then implies that equilibrium experimentation $\xi_{\dagger}$ is 0 for $p \in [0, \underline{c}_{\dagger}]$, increases strictly from 0 to 1 as p increases from $\underline{c}_{\dagger}$ to $\bar{c}_{\dagger}$, and is 1 for $p \geq [\bar{c}_{\dagger}, 1]$. Moreover $\xi_{\dagger}$ is convex on $[0, \bar{c}_{\dagger}]$.

It follows at once from this characterization of equilibrium that the indivisible-time version of our model possesses no symmetric equilibrium: all symmetric equilibria involve a nontrivial division of time between the two actions for all $p \in (\underline{c}_{\dagger}, \bar{c}_{\dagger})$.

It is also interesting to compare equilibrium experimentation with experimentation in the N-player team problem. We have already noted that $c_{*} \leq \underline{c}_{\dagger} < \bar{c}_{\dagger}$. It follows that total equilibrium experimentation $N\beta_{\dagger}/(N-1)$ falls short of the socially optimal level N for $p \in [c_{*}, \bar{c}_{\dagger})$. This shortfall is attributable to the free-rider effect.

Finally, it can be shown that if p lies initially in the interval $(\underline{c}_{\dagger}, 1)$, then strategic experimentation will go on indefinitely. Indeed, $\xi_{\dagger}$ converges quadratically to zero as p approaches $\underline{c}_{\dagger}$, and this implies that the rate of acquisition of information is too slow for p actually to reach $\underline{c}_{\dagger}$. Similarly, Φ converges quadratically to zero as p approaches 1, so the rate of acquisition of information is too slow for p actually to reach 1 either. On the other hand, since p follows a bounded martingale, it must converge; and since the rate of information acquisition is bounded away from zero on compact subsets of $(\underline{c}_{\dagger}, 1)$, it can only converge to $\underline{c}_{\dagger}$ or 1.

6. SYMMETRIC EQUILIBRIA: EXISTENCE AND UNIQUENESS

In this section we exploit the encouragement effect to establish the existence of symmetric equilibrium. More precisely, we know from Lemma 7 that:

- a player's payoff from her best response is increasing in the total experimentation of the other players.

Moreover Theorem 9 tells us that:

- in a symmetric equilibrium, experimentation is increasing in the payoff.

We may therefore construct an increasing mapping in the space of value functions. Applying Tarski's fixed-point theorem to this mapping, we obtain the existence of a minimal and a maximal fixed point. A simple comparison of these fixed points then shows that they actually coincide. In other words, equilibrium is unique.

THEOREM 11: *There exists a unique symmetric equilibrium.*

PROOF: Let $\mathscr{U}$ denote the set of Lipschitz continuous functions $u: [0,1] \to [l,h]$ such that $0 \le u' \le h-l$ almost everywhere on $[0,1]$, and let $\mathscr{X}$ denote the set of Borel measurable functions $\Xi: [0,1] \to [0, N-1]$. Here $\mathscr{U}$ might be thought of as the space of value functions, and $\mathscr{X}$ might be thought of as the space of total-experimentation schedules. For all $u \in \mathscr{U}$, let $\psi_1(u) \in \mathscr{X}$ be defined by the formula

$$(\psi_1(u))(p) = \begin{Bmatrix} \min\left\{ \dfrac{u(p)-s}{s-m(p)}, N-1 \right\} & \text{if } p < b \\ N-1 & \text{if } p \ge b \end{Bmatrix};$$

and for all $\Xi \in \mathscr{X}$, let $\psi_2(\Xi) \in \mathscr{U}$ be the value function of a player who plays a best response when the total experimentation of the other players is Ξ. Finally, let $\psi = \psi_2 \circ \psi_1$.

Theorem 9 implies that $u_\dagger$ is the value function of a symmetric equilibrium iff $u_\dagger$ is a fixed point of ψ. In order to establish existence and uniqueness, then, we need only show that ψ possesses a unique fixed point. It is obvious that ψ_1 is nondecreasing, and it follows from Lemma 7 that ψ_2 is nondecreasing. It therefore follows from Tarski's fixed-point theorem that ψ possesses minimal and maximal fixed points $u_£$ and $u_\$$.

Theorem 9 also implies that if $u_\dagger$ is the value function of a symmetric equilibrium, then there is a unique cutoff $\underline{c}_\dagger \in (0,b)$ such that

$$u_\dagger = \begin{Bmatrix} s & \text{on } [0, \underline{c}_\dagger] \\ (1-\xi_\dagger)s + \xi_\dagger m + \dfrac{1}{r} N \xi_\dagger \Phi \dfrac{u_\dagger''}{2} & \text{on } (\underline{c}_\dagger, 1] \end{Bmatrix},$$

where

$$\xi_\dagger = \begin{Bmatrix} \dfrac{\beta_\dagger}{N-1} & \text{if } \beta_\dagger \le N-1 \text{ and } p < b \\ 1 & \text{if } \beta_\dagger > N-1 \text{ or } p \ge b \end{Bmatrix}$$

and $\beta_\dagger = (u_\dagger - s)/(s-m)$. It follows that

$$(10) \qquad \frac{1}{r} \Phi \frac{u_\dagger''}{2} = \begin{Bmatrix} 0 & \text{on } [0, \underline{c}_\dagger] \\ \max\left\{ s-m, \dfrac{u-m}{N} \right\} & \text{on } (\underline{c}_\dagger, 1] \end{Bmatrix}.$$

Now let $\Delta u = u_\$ - u_£$, where $u_£$ and $u_\$$ are the minimal and maximal fixed points of ψ. Then (10) implies that

$$\frac{1}{r}\Phi\frac{(\Delta u)''}{2}$$
$$= \left\{\begin{array}{ll} 0 & \text{on } [0, \underline{c}_\$] \\ \max\left\{s-m, \dfrac{u_\$ - m}{N}\right\} & \text{on } (\underline{c}_\$, \underline{c}_£] \\ \max\left\{s-m, \dfrac{u_\$ - m}{N}\right\} - \max\left\{s-m, \dfrac{u_£ - m}{N}\right\} & \text{on } (\underline{c}_£, 1] \end{array}\right\}.$$

It follows that $(\Delta u)'' \geq 0$ on $[0,1]$. On the other hand, $\Delta u \geq 0$ and $\Delta u(0+) = \Delta u(1-) = 0$. We conclude that $\Delta u = 0$ on $[0,1]$. *Q.E.D.*

One possible intuition for the uniqueness result is as follows. As in the proof, let $u_£$ and $u_\$$ be the value functions of the minimal and maximal symmetric equilibria respectively. Then $u_\$ \geq u_£$. The encouragement effect therefore implies that experimentation will be higher in the maximal equilibrium. On the other hand, the opportunity cost of experimentation in any given state is the same in both equilibria. Hence, if experimentation is to be higher in the maximal equilibrium then the shadow value of information must also be higher. That is, we must have $(u''_\$/2) \geq (u''_£/2)$. Combining these two inequalities we see that $u_\$ - u_£$ is a nonnegative convex function on $[0,1]$ that vanishes at both 0 and 1. That is, $u_\$ - u_£ = 0$.

7. SYMMETRIC EQUILIBRIA: COMPARATIVE STATICS

In this section we examine the comparative statics of the unique symmetric equilibrium with respect to the discount factor $\rho = 1/r$ and the number of players N.

7.1. *Comparative Statics of Equilibrium Payoffs*

Let $u_\dagger$ denote the value function of the unique symmetric equilibrium, and recall that $\underline{u}$ and $\bar{u}$ denote the myopic payoff and the full-information payoff respectively. Then:

THEOREM 12: *$u_\dagger$ is nondecreasing in both ρ and N. Moreover: (i) $u_\dagger \to \underline{u}$ as $\rho \to 0$; (ii) $u_\dagger \to \bar{u}$ as $\rho \to \infty$; (iii) $u_\dagger(p)$ is bounded away from $\bar{u}(p)$ as $N \to \infty$ for all $p \in (0,1)$.*

Notice that the complete-information payoff is not obtained even in the limit as $N \to \infty$. This underlines the strength of the free-rider effect in our model. It should also be contrasted with the fact that the experimentation cutoff $\underline{c}_\dagger$ for the symmetric equilibrium converges to 0 as $N \to \infty$. (This is established in the proof of the theorem.) In other words, although complete learning is approached in the limit as $N \to \infty$, this learning is still too slow in social terms for the complete-information payoff to be obtained.

PROOF: The monotonicity of $u_\dagger$ in ρ and N is an easy corollary of the existence construction. In the case of ρ, ψ_1 is independent of ρ. Moreover, if $u = \psi_2(\Xi; \rho)$, $\hat{u} = \psi_2(\Xi; \hat{\rho})$, and $\rho \geq \hat{\rho}$, then

$$\begin{aligned} u &= \max_{\alpha_i \in [0,1]} \left((1-\alpha_i)s + \alpha_i m + \rho(\alpha_i + \Xi)\Phi \frac{u''}{2} \right) \\ &\geq \max_{\alpha_i \in [0,1]} \left((1-\alpha_i)s + \alpha_i m + \hat{\rho}(\alpha_i + \Xi)\Phi \frac{u''}{2} \right). \end{aligned}$$

Comparing this with the Bellman equation

$$\hat{u} = \max_{\alpha_i \in [0,1]} \left((1-\alpha_i)s + \alpha_i m + \hat{\rho}(\alpha_i + \Xi)\Phi \frac{\hat{u}''}{2} \right)$$

for $\hat{u}$, we see that $u \geq \hat{u}$. Hence ψ_2 is nondecreasing in ρ. Similarly, in the case of N, $\psi_1(\Xi; N)$ is nondecreasing in N, and ψ_2 is independent of N.

The limit results involving ρ can be deduced from the corresponding results for the team problem. Let $u_*(\cdot; \rho, N)$ denote the value function for the N-player team problem with discount factor ρ. Then $u_*(\cdot; \rho, 1) \leq u_\dagger \leq u_*(\cdot; \rho, N)$. Moreover $u_*(\cdot; \rho, 1) \geq \underline{u}$ and $u_*(\cdot; \rho, N) \leq \bar{u}$. Since $u_*(\cdot; \rho, N)$ converges uniformly to $\underline{u}$ as $\rho \to 0$, and $u_*(\cdot; \rho, 1)$ converges uniformly to $\bar{u}$ as $\rho \to \infty$, $u_\dagger$ too converges uniformly to $\underline{u}$ as $\rho \to 0$ and to $\bar{u}$ as $\rho \to \infty$.

As for the limit result involving N, let

$$\xi_\dagger = \begin{Bmatrix} \min\left\{ \dfrac{1}{N-1} \dfrac{u_\dagger - s}{s - m}, 1 \right\} & \text{if } p < b \\ 1 & \text{if } p \geq b \end{Bmatrix}.$$

Then $u_\dagger$ satisfies

$$u_\dagger = (1-\xi_\dagger)s + \xi_\dagger m + \rho N \xi_\dagger \Phi \frac{u_\dagger''}{2}. \tag{11}$$

Moreover $u_\dagger$ is nondecreasing in N and $\underline{u} \leq u_\dagger \leq \bar{u}$. Now $\rho\Phi(u_\dagger''/2) = 0$ on $[0, \underline{c}_\dagger]$. Moreover (11) implies that

$$\rho\Phi \frac{u_\dagger''}{2} = \frac{(1-\xi_\dagger)(u_\dagger - s) + \xi_\dagger(u_\dagger - m)}{N\xi_\dagger} = \max\left\{ s - m, \frac{u_\dagger - m}{N} \right\}$$

on $(c_\dagger, 1]$. In particular, $\rho\Phi(u_\dagger''/2)$ is uniformly bounded as $N \to \infty$. Hence there exists a function $u_\#$ such that: $u_\dagger$ converges uniformly to $u_\#$ as $N \to \infty$; $\underline{u} \leq u_\# \leq \bar{u}$; $u_\#$ is twice differentiable; and $\rho\Phi(u_\#''/2)$ is bounded. Also, if

$\underline{c}_\dagger = \max\{p | u_\dagger(p) = s\}$ and $\underline{c}_\# = \max\{p | u_\#(p) = s\}$, then $\underline{c}_\dagger$ is nonincreasing in N and $\underline{c}_\dagger$ converges to $\underline{c}_\#$ as $N \to \infty$. It follows that $u_\#$ and $\underline{c}_\#$ together solve the equation

$$\left\{\begin{array}{ll} u_\# = s & \text{on } [0, \underline{c}_\#] \\ \rho\Phi \dfrac{u''_\#}{2} = \max\{s - m, 0\} & \text{on } (\underline{c}_\#, 1) \\ u_\# = h & \text{at } 1 \end{array}\right\}.$$

Finally, it can be verified that this equation has the unique solution

$$\left\{\begin{array}{l} u_\#(p) = \displaystyle\int_0^1 G(p,q) \frac{2}{\rho\Phi(q)} \max\{s - m(q), 0\}\, dq \\ \underline{c}_\# = 0 \end{array}\right\},$$

where $G(p,q) = p(1-q)$ if $p \le q$ and $G(p,q) = (1-p)q$ if $p \ge q$. Since $u_\dagger \le u_\#$ on $[0,1]$ and $u_\# < \bar{u}$ on $(0,1)$, we have $u_\dagger < \bar{u}$ on $(0,1)$ as required. *Q.E.D.*

7.2. *Interpretation of the Comparative Statics of Payoffs*

Further insight into the source of the monotonicity results can be obtained by considering the equation of variations for the value function of the symmetric equilibrium. Indeed, the total information available in the symmetric equilibrium to a player due to experimentation by the other players is

$$\Xi_\dagger = \left\{\begin{array}{ll} \beta_\dagger & \text{if } \beta_\dagger \in [0, N-1] \text{ and } p < b \\ N-1 & \text{if } \beta_\dagger > N-1 \text{ or } p \ge b \end{array}\right\}$$

by Theorem 9, where $\beta_\dagger = (u_\dagger - s)/(s - m)$. Hence the value function of this equilibrium must satisfy the Bellman equation

$$u_\dagger = \max_{\alpha_i \in [0,1]} \left((1-\alpha_i)s + \alpha_i m + \rho(\alpha_i + \Xi_\dagger)\Phi \frac{u''_\dagger}{2} \right).$$

Hence, applying the envelope theorem, we obtain

$$\frac{\partial u_\dagger}{\partial \rho} = f + \rho(\xi_\dagger + \Xi_\dagger) \frac{\Phi}{2} \left(\frac{\partial u_\dagger}{\partial \rho} \right)'', \tag{12}$$

where

$$f = (\xi_\dagger + \Xi_\dagger)\Phi \frac{u''_\dagger}{2} + \rho \frac{\partial \Xi_\dagger}{\partial \rho} \Phi \frac{u''_\dagger}{2} \tag{13}$$

and

$$\frac{\partial \Xi_\dagger}{\partial \rho} = \left\{ \begin{array}{ll} \dfrac{1}{s-m}\dfrac{\partial u_\dagger}{\partial \rho} & \text{if } \beta_\dagger \in [0, N-1] \text{ and } p < b \\ 0 & \text{if } \beta_\dagger > N-1 \text{ or } p \geq b \end{array} \right\}.$$

Here f is the state shadow price of an extra unit of patience, and equation (12) states that the shadow value of an extra unit of patience is the expected present discounted value of the state shadow price.

Examining (13), we see that there are two components to the state shadow price of patience. These two components correspond to the two effects on a player's payoff of an increase in ρ:

(i) The direct effect on her objective of the increase in ρ.

(ii) The indirect effect on her objective of the change induced in the behavior of the other players by the change in ρ.

The first effect is positive: she now values the future more relative to the present, and so she values the opportunity to experiment more. The second effect is a strategic effect: the rise in ρ leads to an increase in experimentation by the other players when $\beta_\dagger \in [0, N-1]$ and $p < b$, and this further raises her payoff. The effect on her objective of the change induced in her own behavior by the change in ρ is of course zero by the envelope theorem. Notice the positive feedback implicit in the second effect. This is yet another manifestation of the encouragement effect.

A similar analysis applies to the case of N. Increasing N has no direct effect on a player's objective. However, it does raise experimentation by the other players through an encouragement effect when $\beta_\dagger \in [0, N-1]$ and $p < b$, and through an increase in the total number of players experimenting when $\beta_\dagger > N-1$ or $p \geq b$. It also lowers per capita experimentation by the other players through a free-rider effect when $\beta_\dagger \in [0, N-1]$ and $p < b$, but this decrease is exactly offset by the increase in the number of players.

7.3. *Comparative Statics of Experimentation*

The comparative statics in ρ of the equilibrium payoffs translate directly into comparative statics in ρ of the equilibrium mixed strategy $\xi_\dagger$ via the formula

$$\xi_\dagger = \left\{ \begin{array}{ll} \dfrac{1}{N-1}\beta_\dagger & \text{if } \beta_\dagger \leq N-1 \text{ and } p < b \\ 1 & \text{if } \beta_\dagger > N-1 \text{ or } p \geq b \end{array} \right\}.$$

The same is not true of the comparative statics in N. Indeed, increases in N involve both the free-rider effect—in that $1/(N-1)$ is strictly decreasing in N—and the encouragement effect—in that $u_\dagger$ is nondecreasing in N.

Perhaps the easiest way to see that the influence of the free-rider effect on $\xi_\dagger$ can outweigh that of the encouragement effect is to calculate the limit $\xi_\#$ of $\xi_\dagger$

as $\rho \to 1$. We know from Theorem 12 that $u_{\dagger} \to \bar{u}$ as $\rho \to 1$. Hence

$$\xi_{\#} = \begin{cases} \dfrac{1}{N-1}\bar{\beta} & \text{if } \bar{\beta} \le N-1 \text{ and } p < b \\ 1 & \text{if } \bar{\beta} > N-1 \text{ or } p \ge b \end{cases},$$

where $\bar{\beta} = (\bar{u} - s)/(s - m)$. Hence $\xi_{\#}$ is strictly decreasing in N for $p \in (0, \bar{c}_{\#})$, where

$$\bar{c}_{\#} = \frac{(N-1)(s-l)}{(h-s) + (N-1)(h-l)}.$$

Indeed, even total experimentation $N\xi_{\#}$ is strictly decreasing in N for p in this range.

In other words, neither individual nor total experimentation is in general monotonic in N. This underlines the subtlety of the comparative statics in N of the equilibrium payoff $u_{\dagger}$.

8. MIXED STRATEGIES

As we have already pointed out in Section 2 above, an indivisible-time model can be obtained from our model by restricting players to the actions 0 and 1. In this section we argue that the best-response correspondences for the mixed extension of the indivisible-time model are isomorphic to the best-response correspondences for the original time-division model. In particular, the analysis of equilibrium given in Sections 5, 6, and 7 above applies equally well to the mixed extension. One need only reinterpret the proportion of time that player i devotes to the risky action as the probability with which player i chooses that action.

Since our game is played in continuous time, consideration of mixed strategies raises a variant of the old question of how to formulate a continuum of independent and identically distributed random variables in a tractable way.[7] Our variant of this question is more complicated than the usual one since it is necessary to take account of the interplay between the dynamics of our game on the one hand and the choice of mixed actions by the players on the other. More

[7] It is straightforward to construct a probability space on which a continuum of independent and identically distributed random variables is defined. For example, for all $t \in [0,1]$, let Ω_t be the unit interval $[0,1]$ endowed with the usual Euclidean topology; let $\Omega = \times_{t \in [0,1]} \Omega_t$; let $\mathscr{F}$ be the Borel σ-algebra generated by the product topology on Ω; and let μ be the unique probability measure on Ω, the finite dimensional distributions of which are products of uniform distributions. Then the projections proj_t: $\Omega \to \Omega_t$ are all random variables on the probability space $(\Omega, \mathscr{F}, \mu)$, and they are independently and uniformly distributed. The difficulty consists in deriving results about such random variables. For example, one would like to be able to show that the strong law of large numbers holds in the sense that $\int \mathrm{proj}_t(\omega)\, dt = \frac{1}{2}$ for a set of ω having μ-measure 1. It is not clear, however, how to set about establishing this, since the probability that the function $t \mapsto \mathrm{proj}_t(\omega)$ is Lebesgue measurable is zero. See Anderson (1991) for one approach to this problem.

explicitly, the choice of mixed actions by the players depends on the past evolution of the state, and the evolution of the state depends on the choice of actions by the players.

The ideal response to this question would be to formulate the required continuum of independent and identically distributed random variables appropriately, and then to derive the generalization of the Bellman equation of player i from this formulation. Such an approach is beyond the scope of this paper. We therefore confine ourselves to a heuristic derivation of the generalization of the Bellman equation of player i.

8.1. *The Bellman Equation of Player i*

We begin with a definition. Let $\Delta(\{0,1\})$ denote the set of probability distributions on $\{0,1\}$.

DEFINITION 13: A *mixed Markov strategy* for player i is a Borel measurable mapping ξ_i: $[0,1] \to \Delta(\{0,1\})$.

In what follows we shall refer to mixed Markov strategies simply as mixed strategies wherever this will not lead to any misunderstanding.

Next, we emphasize that, in deriving the Bellman equation of player i, we shall take it that the order of events in any given period is as follows:

(i) at the outset of the period, players share a common belief p;

(ii) the players choose actions $a_i \in \{0,1\}$, the choices being made simultaneously and independently;

(iii) the payoffs $d\pi_i$ are realized, where $d\pi_i = s\,dt + \sigma\,dZ_i^0$ if $a_i = 0$ and $d\pi_i = \mu\,dt + \sigma\,dZ_i^1$ if $a_i = 1$;

(iv) the players observe the action profile $a = \times_{i=1}^N a_i$ and the payoff profile $d\pi = \times_{i=1}^N d\pi_i$;

(v) a new belief $p + dp$ is generated.

This order of events corresponds closely to the order of events which obtains in any given period of a discrete-time stochastic game. Indeed, the difference between our continuous-time stochastic differential game and a discrete-time stochastic game consists not in the way a period of the game is played, but rather in the relationship among periods. In a discrete-time stochastic game, each period has a well defined subsequent period, and it is therefore possible to determine the evolution of such a game beginning with the initial period and moving inductively from period to period. By contrast, in our continuous-time stochastic differential game, no period has a well defined subsequent period, and it is therefore necessary to exploit the differential structure of the game to determine the evolution of the game from a knowledge of the way a period of the game is played.

Suppose accordingly that the players $j \neq i$ employ the mixed strategies ξ_j, that the current belief is $p \in [0,1]$, that player i chooses mixed action $\alpha_i \in \Delta(\{0,1\})$, and that her continuation payoffs are given by c_i: $[0,1] \to \mathbb{R}$. Then, arguing as in the proof of Lemma 3, we see that the expectation of her current payoff conditional on the realization $a = \times_{j=1}^{N} a_j$ of the current action profile would be expected to be

$$c_i(p) + r\,dt\left((1-a_i)s + a_i m(p) + \frac{1}{r}\left(a_i + \sum_{j \neq i} a_j\right)\Phi(p)\frac{c_i''(p)}{2} - c_i(p)\right).$$

The unconditional expectation of her current payoff would then be

$$H_i(\alpha_i, \xi_{-i}, p, c_i) = c_i(p) + r\,dt\left((1-\alpha_i(\{1\}))s + \alpha_i(\{1\})m(p) + \frac{1}{r}\left(\alpha_i(\{1\}) + \sum_{j \neq i} \xi_j(\{1\}|p)\right)\Phi(p)\frac{c_i''(p)}{2} - c_i(p)\right),$$

where $\xi_{-i} = \times_{j \neq i} \xi_j$; and her value function u_i when the other players use the mixed-strategy profile ξ_{-i} would be the unique solution of the Bellman equation

$$u_i(p) = \max_{\alpha_i \in \Delta(\{0,1\})} H_i(\alpha_i, \xi_{-i}(p), p, c_i) \quad \text{for all } p \in [0,1]. \tag{14}$$

Finally, arguing as in the proof of Lemma 4, the mixed strategy ξ_i would be a best response to the mixed-strategy profile $\xi_{-i} = \times_{j \neq i} \xi_j$ iff

$$\xi_i(p) \in \operatorname*{argmax}_{\alpha_i \in \Delta(\{0,1\})} \left((1-\alpha_i(\{1\}))s + \alpha_i(\{1\})m(p) + \frac{1}{r}\left(\alpha_i(\{1\}) + \sum_{j \neq i} \xi_j(\{1\}|p)\right)\Phi(p)\frac{u_i''(p)}{2}\right) \tag{15}$$

for all $p \in [0,1]$.

It is easy to see that equations (14) and (15) are isomorphic to equations (1) and (4). One need only identify the probabilities $\alpha_i(\{1\})$ and $\xi_j(\{1\}|p)$ with which players choose the risky action in (14) and (15) with the proportions of time α_i and $\xi_j(p)$ which they devote to the risky action in (1) and (4).

9. PUBLIC RANDOMIZATION

In this section we analyze the public-randomization extension of the indivisible-time version of the model. This extension is motivated by the idea that one way of capturing the intuitively natural experimentation pattern of taking turns is to introduce a public-randomization device, and to have players coordinate their experimentation on the basis of the realization of this device in such a way that exactly one player experiments at any given time and each player experiments with equal probability. For example, two players might toss a coin, with player 1 experimenting if the coin comes up heads and player 2 experimenting if the coin comes up tails. Or again, six players might roll a die, with player i experimenting if and only if the die shows i.

In this extension, the order of events in any given period is as follows:

(i) at the outset of the period, players share a common belief p;

(ii) a referee chooses a pure-action profile $g \in \{0,1\}^N$ according to a distribution $\gamma \in \Delta(\{0,1\}^N)$;

(iii) the players observe g;

(iv) the players choose actions $a_i \in \{0,1\}$, the choices being made simultaneously and independently;

(v) the payoffs $d\pi_i$ are realized, where $d\pi_i = s\,dt + \sigma\,dZ_i^0$ if $a_i = 0$ and $d\pi_i = \mu\,dt + \sigma\,dZ_i^1$ if $a_i = 1$;

(vi) the players observe the action profile $a = \times_{i=1}^N a_i$ and the payoff profile $d\pi = \times_{i=1}^N d\pi_i$;

(vii) a new belief $p + dp$ is generated.

A mixed joint strategy Γ: $[0,1] \to \Delta(\{0,1\}^N)$ is a public-randomization equilibrium of our game if and only if it is incentive compatible for the players to carry out the actions g_i recommended by the referee. It is a symmetric public-randomization equilibrium if and only if it is unchanged by any permutation of the players.

By exploiting the additive separability of the players' payoffs in the stage game in the action profile a, it can be shown that Γ is a public-randomization equilibrium if and only if the marginals ξ_i of Γ over the action sets of the individual players constitute a mixed-strategy equilibrium of the original game. Moreover the value functions of the players in the public-randomization equilibrium Γ coincide with their value functions in the mixed-strategy equilibrium $\xi = \times_{i=1}^N \xi_i$. In particular, our characterization of symmetric mixed-strategy equilibria leads directly to a characterization of symmetric public-randomization equilibria. Cf. Harris (1993).

For example, there is a symmetric public-randomization equilibrium in which: with probability $1 - \xi_\dagger$ players all play safe; and with probability $\xi_\dagger$ players all experiment. The difference between the two equilibria is that in the mixed-strategy equilibrium the experiments are independent, whereas in the public-randomization equilibrium the experiments are perfectly correlated.

Or again, there is a symmetric public-randomization equilibrium in which: with probability $\max\{1 - 2\xi_\dagger, 0\}$ both players play safe; with probability

$$\xi_\dagger - \frac{1}{2}\max\{2\xi_\dagger - 1, 0\}$$

player 1 alone experiments; with probability

$$\xi_{\dagger} - \frac{1}{2}\max\{2\xi_{\dagger} - 1, 0\}$$

player 2 alone experiments; and with probability $\max\{2\xi_{\dagger} - 1, 0\}$ both players experiment. In particular, there is a cutoff $c_{\$} \in (\underline{c}_{\dagger}, \overline{c}_{\dagger})$ such that: for $p \in (\underline{c}_{\dagger}, c_{\$})$, at most one player experiments at any given time; and for $p \in (c_{\$}, \overline{c}_{\dagger})$, at least one player experiments at any given time. This time the difference between the two equilibria is that in the mixed-strategy equilibrium the experiments are independent, whereas in the public-randomization equilibrium the experiments are as negatively correlated as the incentive-compatibility constraints allow.

The negative correlation between the actions of the two players in the second example captures something of the flavor of taking turns. Taking turns as such is not, however, incentive compatible.

10. CONCLUSION

In this paper we have analyzed team and equilibrium experimentation in a many-player common-value two-armed bandit problem in terms of the free-rider effect and the encouragement effect. The model which we used for this purpose has at least three limitations. First, there are only two possible states of the world: the underlying payoff from the risky action is either high or low. Secondly, there is no distinction made between players' signals and payoffs. Thirdly, players choose between only two actions. Fourthly, one of the actions is safe. The first, second, and fourth of these limitations can be removed completely: there can be any (finite) number of states of the world; the choice of the risky action can generate a whole vector of signals, which may or may not include the payoff; and players can be allowed to choose between two risky actions, provided that the two actions differ in the quantity but not in the pattern of information that they produce. See Bolton and Harris (1996).

Allowing for many arms, all yielding the same pattern of information, raises one new issue: what is the best incentive-compatible pattern of experimentation? Continuing to restrict attention to two arms, but allowing the pattern of information generated by one arm to differ from the pattern of information generated by the other, raises another: what is the best way of exploiting the different patterns of information generated by the two arms? We hope to address these issues in future work.

Dept. of Economics, 001 Fisher Hall, Princeton University, Princeton, NJ 08544, U.S.A.

and

King's College, Cambridge University, Cambridge CB2 1ST, U.K.

Manuscript received February, 1993; final revision received April, 1998.

REFERENCES

ANDERSON, ROBERT M. (1991): "Non-Standard Analysis with Applications to Economics," in *Handbook of Mathematical Economics, Volume IV*, ed. by W. Hildenbrand and H. Sonnenschein. Amsterdam: North-Holland, Ch. 39, pp. 2145–2208.

BERRY, D. A., AND B. FRISTEDT (1985): *Bandit Problems*. New York: Chapman and Hall, 1985.

BOLTON, PATRICK, AND CHRISTOPHER HARRIS (1993): "Strategic Experimentation. Theoretical Economics," Discussion Paper TE/93/261, STICERD, London School of Economics.

——— (1996): "Strategic Experimentation: A Revision," Discussion Paper 9627, CentER, Tilburg University.

——— (1997): "Strategic Experimentation," Mimeo, Princeton University.

HARRIS, CHRISTOPHER (1993): "Generalized Solutions of Stochastic Differential Games in One Dimension," Theoretical Economics Discussion Paper 44, Boston University.

JACOD, J., AND A. N. SHIRYAYEV (1988): *Limit Theorems for Stochastic Processes*. Berlin: Springer-Verlag.

KARATZAS, I. (1984): "Gittins Indices in the Dynamic Allocation Problem for Diffusion Processes," *Annals of Probability*, 12, 173–192.

KARATZAS, I., AND S. E. SHREVE (1988): *Brownian Motion and Stochastic Calculus*. Berlin: Springer-Verlag.

KRYLOV, N. V. (1980): *Controlled Diffusion Processes*. Berlin: Springer-Verlag.

MANDELBAUM, A. (1987): "Continuous Multi-Armed Bandits and Multi-Parameter Processes," *Annals of Probability*, 15, 1527–1556.

SMITH, L. (1991): "Error Persistence, and Experiential versus Observational Learning," Foerder Discussion Paper, Tel-Aviv University and MIT.