ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 75, No. 6 (November, 2007), 1591–1611

SOCIAL LEARNING IN ONE-ARM BANDIT PROBLEMS

DINAH ROSENBERG
Institut Galilée, Université Paris Nord, 93430 Villetaneuse, France, and Laboratoire d'Econométrie de l'Ecole Polytechnique, Paris, France

EILON SOLAN
The School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel

NICOLAS VIEILLE
HEC, 78 351 Jouy-en-Josas, France

Econometrica, Vol. 75, No. 6 (November, 2007), 1591–1611

SOCIAL LEARNING IN ONE-ARM BANDIT PROBLEMS

BY DINAH ROSENBERG, EILON SOLAN, AND NICOLAS VIEILLE[1]

We study a two-player one-arm bandit problem in discrete time, in which the risky arm can have two possible types, high and low, the decision to stop experimenting is irreversible, and players observe each other's actions but not each other's payoffs. We prove that all equilibria are in cutoff strategies and provide several qualitative results on the sequence of cutoffs.

KEYWORDS: Social learning, one-arm bandit, equilibrium, cutoff strategies.

INTRODUCTION

RECENT MODELS of strategic experimentation (see Bolton and Harris (1999) and Keller, Rady, and Cripps (2005)) feature players who face identical bandit problems with a risky arm and a safe arm. It is assumed that players are free to switch from one arm to the other and that all information—both the actions and the actual rewards of the players—is publicly disclosed. The latter assumption appears to be restrictive in many economic setups, and the necessity to drop it has been recognized; see, for example, Bolton and Harris (1999).[2]

Here we address this task partially and study the following model. Each of two players faces a bandit machine with a risky arm and a safe arm. The risky arm is either of the high type, which yields independent and identically distributed (i.i.d.) payoffs with positive expectation, or of the low type, which yields payoffs with negative expectation. The machines of the two players have the same type. The decision to switch to the safe arm is irreversible. Along the play, each player observes her opponent's choices, but not her opponent's payoffs.

Dropping the assumption that payoffs are publicly observed raises new issues. Player i would like to make inferences about player j's observations on the basis of player j's actions, but cannot do so without knowing how player j's decisions relate to player j's observations, that is, j's strategy. As a consequence, there is no commonly observed state variable, such as a common posterior belief, on which to condition one's actions.

[1]We are indebted to Itzhak Gilboa for his help. We thank Yisrael Aumann, Pierpaolo Battigalli, Thomas Mariotti, Peter Sozou, and Shmuel Zamir for their suggestions, and seminar audiences at Caltech, Hebrew University of Jerusalem, LSE, Northwestern, Paris, Stanford, Tel Aviv University, Toulouse 1 University, the Workshop on Stochastic Methods in Game Theory in Erice, and the Second World Congress of the Game Theory Society. The research of the second author was supported by the Israel Science Foundation, grant 69/01.

[2]Other models of strategic experimentation include Bergemann and Välimäki (1997, 2000) and Décamps and Mariotti (2004). Bergemann and Välimäki (1997, 2000) studied a model of sellers and buyers who learn the value of new products by experimentation. Décamps and Mariotti (2004) studied a specific duopoly model where each player learns about the quality of a common value project by observing some public information plus the experience of her rival.

Our main result states that, nevertheless, all equilibrium strategies process information in a simple way: at each stage a player (i) computes the conditional probability that the type is high, using *only* her private observations (i.e., her own payoffs), (ii) determines a time-dependent cutoff, which depends on the public information (the decisions of the other player), and (iii) switches to the safe arm if the conditional probability does not exceed the cutoff.

The intuition for this result runs as follows. Once player j's strategy is given, player i faces an optimal stopping problem. It turns out that the pair formed by player j's status (active or not) and player i's private belief follows a Markov process. This observation crucially relies on the property that the payoffs to both players are conditionally independent given the type of the machines, and it allows one to recast player i's optimal continuation payoff as a function of this pair. We next prove that, everything else being equal, player i's continuation payoff increases with respect to (w.r.t.) her private belief and is, therefore, positive (resp. negative) above (resp. below) some cutoff.

We also prove that the equilibrium cutoffs are nonincreasing as long as the other player is active and are constant afterward: seeing the other player active induces a player to be more optimistic and, therefore, to stay active with lower beliefs associated to private payoffs. Finally, we argue that, as the number of players increases, there is eventually a unique equilibrium that can be explicitly derived.

Our model is equivalent to a multiplayer version of the standard real-options problem (see Dixit and Pindyck (1994, p. 136)) in which an investor has to choose when to invest in some project with uncertain prospects. This equivalence is discussed in Section 2.2.

The model also relates to the literature on social learning with endogenous timing. In Chamley and Gale (1994) (see also Chamley (2004)), players are endowed with private information on the state of nature and must decide when to "invest." Like here, externalities are purely informational, decisions are irreversible, and information is private. The main difference is that private information is received only once, in contrast to our setup where it keeps flowing in.

Finally, some studies in biology address similar issues. In some contexts, animals can learn some relevant information by observing the behavior of other animals of the same species. Such behavior has been studied by, for example, Valone and Templeton (2002) and Giraldeau, Valone, and Templeton (2002).

The paper is organized as follows. In Section 1, we present the model and the main results. Comments and extensions appear in Section 2. All proofs are relegated to the Appendix.

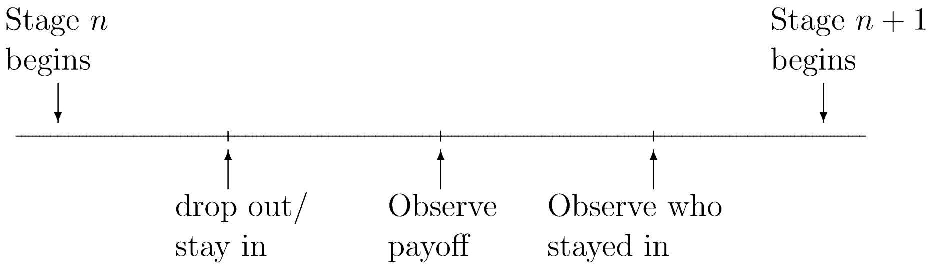

FIGURE 1.—Evolution of the game.

1. MODEL AND RESULTS

1.1. *The Game*

Each of two players operates a one-arm bandit machine in discrete time and must decide when to stop operating the machine. The decision to stop is irreversible and yields an outside payoff normalized to zero.

At each stage $n \geq 0$, the following sequence of events unfolds (see Figure 1). First, each (active) player i decides whether to drop out (that is, stop operating the machine) or not. If she chooses the latter, she receives a random payoff X_n^i and observes who decided to stay in the game. Thus, payoffs are private information, while the exit decisions are publicly observed.

Player i's machine is one of two types: *high* or *low*. The two machines have the same type Θ, which is chosen by nature at the outset of the game according to a known prior. We assume that, conditional on Θ, the payoffs (X_n^i, X_n^j), $n \geq 0$, are i.i.d.

Denote by $\overline{\theta}$ (resp. $\underline{\theta}$) the expected stage payoff of a machine of type high (resp. low), which we identify with the type of the machine. To eliminate trivial cases, we assume that $\underline{\theta} < 0 < \overline{\theta}$. The players discount payoffs at a common rate $\delta \in (0, 1)$.

Note that the payoff to player i is not affected by player j's decisions. However, insofar as player j's decisions are affected by her payoffs, and since the payoffs to both players are correlated, player j's decisions may be used by player i to infer information about Θ.

1.1.1. *Strategies*

Let $(\Omega, \mathbf{P})$ be the probability space over which all random variables are defined. $\mathbf{P}_\theta$ stands for the conditional probability given $\Theta = \theta$. Expectations w.r.t. $\mathbf{P}$ and $\mathbf{P}_\theta$ are denoted $\mathbf{E}$ and $\mathbf{E}_\theta$, respectively. The prior probability that the machines' type is high is denoted by $p_0 := \mathbf{P}(\Theta = \overline{\theta})$.

A strategy of player i specifies when to drop out from the game. At stage n, player i's private information consists of her past payoffs $(X_0^i, \ldots, X_{n-1}^i)$. We denote by $\mathcal{F}_n^i = \sigma(X_0^i, \ldots, X_{n-1}^i)$ the σ-algebra over Ω defined by player i's private information at stage n.

In addition, player i knows if (and when) the other player, player j, dropped out. We denote by $\alpha \in \mathbf{N} \cup \{\blacktriangle\}$ the status variable of player j: $\alpha = \blacktriangle$ if player j is still active and $\alpha = k$ if player j dropped out at stage k. Accordingly, we have the following definition of a pure strategy.

DEFINITION 1: A *pure strategy* of player i is a family $\phi^i = (\tau^i(\alpha), \alpha \in \mathbf{N} \cup \{\blacktriangle\})$ of stopping times for the filtration $(\mathcal{F}_n^i)_{n\in\mathbf{N}}$, with the property that $\tau^i(k) > k$, $\mathbf{P}$ almost surely ($\mathbf{P}$-a.s.) for each $k \in \mathbf{N}$.

Player i's behavior until player j drops out is described by $\tau^i(\blacktriangle)$, while $\tau^i(k)$ ($k \in \mathbf{N}$) describes her behavior after stage k, in the event player j drops out at stage k.

Cutoff strategies process information in the simplest way. At any stage, the decision whether to drop out or to continue is made by computing the belief assigned to the state high, given only *one's own private information*, and comparing it to a *time-dependent cutoff*.

Formally, for every stage $n \in \mathbf{N}$, we define $p_n^i := \mathbf{P}(\Theta = \overline{\theta} \mid \mathcal{F}_n^i)$. This is the posterior belief over the machine's type, taking into account player i's private information. We call it the *private belief* of player i.

DEFINITION 2: A strategy ϕ^i is a *cutoff strategy* if there exist $\pi_n^i(\alpha) \in [0, 1]$, ($n \in \mathbf{N}$ and $\alpha \in \{\blacktriangle, 0, 1, \ldots, n-1\}$), such that $\tau^i(\blacktriangle) = \inf\{n \geq 0 : p_n^i \leq \pi_n^i(\blacktriangle)\}$ and $\tau^i(k) = \inf\{n > k : p_n^i \leq \pi_n^i(k)\}$ for each $k \in \mathbf{N}$.[3]

Figure 2 shows typical cutoffs of a cutoff strategy (here we depict only $\pi_n^1(\blacktriangle)$; they are decreasing, and are denoted by large circles), with a typical evolution of the private belief of the player (assuming player 2 stays in throughout). In this figure, the player stops at stage 7, once her private belief falls below her cutoff.

Mixed strategies are probability distributions over pure strategies. Following Aumann (1964), this is formalized by supplying each player i with an external randomization device uniformly distributed over $[0, 1]$, which is privately observed at the outset of the game.

Given a pair of (pure or mixed) strategies $\phi = (\phi^1, \phi^2)$, we denote $t^i(\phi) \in \mathbf{N} \cup \{+\infty\}$ as the stage in which player i drops out. That is, $t^1(\phi) = n$ if either (i) both $\tau^2(\blacktriangle) \geq n$ and $\tau^1(\blacktriangle) = n$ or if (ii) both $\tau^2(\blacktriangle) = k$, $\tau^1(\blacktriangle) > k$, and $\tau^1(k) = n$ for some $k < n$. In the former case, player 1 drops out before or with player 2, whereas in the latter, player 2 drops out first at stage k.

Player i's overall payoff is the discounted sum $r^i(\phi) := \sum_{n=0}^{t^i(\phi)-1} \delta^n X_n^i$ of payoffs received prior to dropping out. Her expected payoff is $\gamma^i(\phi) := \mathbf{E}[r^i(\phi)]$.

[3]These cutoffs need not be uniquely defined. For example, if π and π' are such that $\mathbf{P}(\pi' \leq p_i^1 \leq \pi) = 0$, then any choice of $\pi_n^1(\alpha)$ in the interval $[\pi', \pi]$ gives rise to the same strategy.

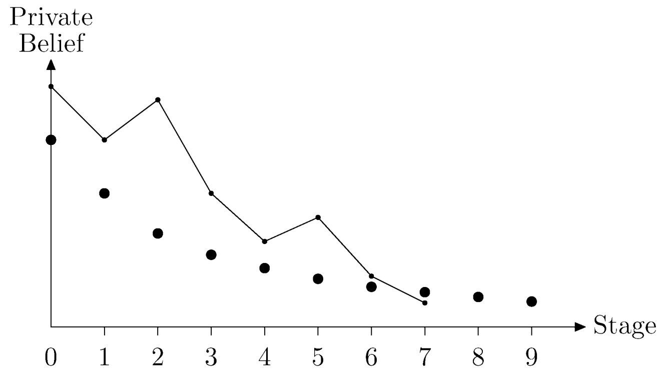

FIGURE 2.—Typical evolution of the private belief.

1.2. *Main Results*

All of our results are obtained under Assumption A below.

ASSUMPTION A: The law of the private belief p_1^i held by player i at stage 1 has a density (w.r.t. Lebesgue measure).

The assumption means that the distribution of X_n^i given θ has a density f_θ whose support does not depend on θ, and, moreover, the likelihood ratio $(f_{\overline{\theta}}(X_n^i))/(f_{\underline{\theta}}(X_n^i))$ is a random variable (r.v.) that has a density. In particular, the probability that $p_1^i = p_2^i$ is 0; with probability 1, the players hold different private beliefs. By Bayes' rule, the likelihood ratio at stage 1 is then given by

$$\frac{p_1^i}{1-p_1^i} = \frac{p_0}{1-p_0} \times \frac{f_{\overline{\theta}}(X_0^i)}{f_{\underline{\theta}}(X_0^i)}.$$

Thus, the existence of a density for p_1^i is equivalent to the existence of a density for the r.v. $(f_{\overline{\theta}}(X_0^i))/(f_{\underline{\theta}}(X_0^i))$. Under Assumption A, the private belief p_n^i has a density for each $n \geq 0$.

Our first result is standard and claims that a symmetric equilibrium exists.

THEOREM 3—Existence: *The game has a symmetric equilibrium*.

The uniqueness issue is addressed in Section 2.3. The next two results characterize the equilibrium strategies.

THEOREM 4—Structure: *All equilibria are in cutoff strategies*.

We actually prove that any best reply is a cutoff strategy and, therefore, so is any rationalizable strategy.

According to Theorem 4, *all* equilibria are pure and process information in a simple way. The interaction is incorporated in the way cutoff values depend on time and public information: if player j drops out at stage n, player i takes it into account by changing the cutoff and using $\pi^i_\cdot(n)$ rather than $\pi^i_\cdot(\blacktriangle)$ from then on; if player j does not drop out by stage n, player i takes this fact into account by using at stage $n+1$ the cutoff $\pi^i_{n+1}(\blacktriangle)$ rather than $\pi^i_n(\blacktriangle)$ that she used at stage n.

At a given stage, equilibrium behavior is monotonic w.r.t. the private belief. However, private belief need not be monotonic w.r.t. payoffs, unless the likelihood ratio $f_{\overline{\theta}}/f_{\underline{\theta}}$ is monotonic; high payoffs need not be good news; see Milgrom (1981).

If player j drops out at stage k, player i remains alone and gets no more public information from player j. The continuation game starting at stage $k+1$ is then analogous to a one-player game, where the initial prior takes into account the private belief p^i_{k+1} of player i and the fact that player j dropped out at stage k. This one-player version is equivalent to the usual one-arm bandit problem, in which exit decisions are reversible; see Chow and Robbins (1963) or Ferguson (2004). The optimal policy is to drop out as soon as the belief assigned to $\overline{\theta}$ (given public and private information) drops below a *time-independent* cutoff value, which we denote π_*.

Accordingly, after player j drops out, player i faces an auxiliary decision problem, in which she drops out whenever her belief is lower than π_*. This is equivalent to using a cutoff strategy, that is, dropping out whenever the *private* belief is lower than some cutoff. The cutoff is calculated by Bayes' rule from π_* and the fact that player j dropped out at stage k. We denote it by $\pi^i(k)$, so that $\pi^i_n(k) = \pi^i(k)$ for every $n > k$.

THEOREM 5: *Let an equilibrium with cutoffs* (π^i_n) *be given*.
P1. *The cutoff sequences* $(\pi^i_n(\blacktriangle))_{n\in\mathbf{N}}$ *are nonincreasing for* $i = 1, 2$.[4]
P2. $\lim_{n\to\infty} \pi^i_n(\blacktriangle) = 0$ *for at least one player* i.
P3. *For each player* i, $\pi^i_n(\blacktriangle) < \pi_*$, *and* $\pi^i_n(\blacktriangle) < \pi^i(k)$ *whenever* $k < n$.

Statements P1–P3 hold for any pair of rationalizable strategies, and not only for equilibrium strategies.[5] We comment briefly on these statements. Two possibly conflicting effects combine in P1. Assume that player i reaches stages n and $n+1$ with the same private belief p, while player j is still active. State $\overline{\theta}$ is then more likely at stage $n+1$ than at stage n: indeed, the longer player j is

[4]Since the sequence $(\pi^i_n(\blacktriangle))_{n\in\mathbf{N}}$ need not be uniquely defined, P1 should be interpreted as to mean that given an equilibrium, the corresponding cutoffs may be chosen in such a way that the sequences $(\pi^i_n(\blacktriangle))_{n\in\mathbf{N}}$ are nonincreasing.

[5]Indeed, by our proof of Theorem 4, every best reply is a cutoff strategy. Moreover, by properly adapting our proof of Theorem 5, one obtains that every best reply to a mixture of cutoff strategies satisfies P1–P3. It follows that P1–P3 hold for any pair of rationalizable strategies.

active, the better news it is on Θ. On the other hand, the optimal continuation value for player i also depends on whether and when she is likely to infer information about θ through player j's behavior in the future. That is, it involves both player i's current belief over the private belief currently held by player j, and player j's future cutoffs. This second effect is ambiguous. According to P1, the combined effect is such that the optimal continuation payoff is higher at stage $n+1$: a lower value for the private belief is necessary to trigger exit.

The intuition behind P2 is the following. If player j never drops out, p_n^j converges to 1 if $\Theta = \overline{\theta}$ and to 0 if $\Theta = \underline{\theta}$. If player j's cutoffs are bounded away from 0, the fact that she stays in longer provides very good news on Θ. Player i will therefore remain active, unless she gets strong private information in the opposite direction: player i's cutoffs will converge to zero.

We turn to P3 that incorporates two effects. On the one hand, having an active opponent is good news on Θ and is better news than the opponent dropping out in some earlier stage k. This is reflected in a higher posterior belief. On the other hand, having an active opponent creates an informational externality that does not exist in the one-player case and no longer exists if the opponent has already dropped out. For a given posterior, this is reflected in a higher option value.

Note that the one-player cutoff π_* decreases to zero when the discount rate goes to 1: the cost of experimentation drops to zero. According to P3, all equilibrium cutoffs $(\pi_n^i(\blacktriangle))$ then converge to zero.

We conclude by mentioning that our results still hold whenever the players receive private signals at every stage that are conditionally independent given Θ, in addition to being told their own payoff. They also hold if the players hold different (thereby inconsistent) prior beliefs on Θ, provided we use the notion of subjective Nash equilibrium.

1.2.1. *Large games*

When the number of players exceeds two, a strategy keeps track of the status of every other player. The definition of cutoff strategies is similar, except that cutoffs now depend on who dropped out and when. Theorems 3, 4, and 5(P2, P3) hold for any finite number of players.

When the number of players gets large, equilibrium behavior can be fully characterized.[6] To simplify the characterization, we will assume that (i) p_1^i has full support and (ii) $p_0 > \pi_*$, so that dropping out at stage 0 is a strictly dominated strategy.

Let ϕ^N be an arbitrary equilibrium of the N-player game, with cutoffs $(\pi_n^{i,N})$. At stage 1, all cutoffs $\pi_1^{i,N}$ are bounded away from zero. All players are active at stage zero, and more players drop out at stage 1 if $\Theta = \underline{\theta}$ than if $\Theta = \overline{\theta}$. Hence

[6] We refer to Rosenberg, Solan, and Vieille (2004) for a detailed proof.

the proportion of players who drop out at stage 1 reveals Θ with high probability to all remaining players. Therefore, the continuation payoff at stage 1 may (asymptotically) be computed under the assumption that θ will be revealed at the beginning of stage 2: all players use cutoffs that are close to π_c, the unique solution of $\pi\overline{\theta} + (1-\pi)\underline{\theta} + \pi\delta\overline{\theta}/(1-\delta) = 0$.

Denote by ρ^N the fraction of players who drop out at stage 1. By a large deviations argument, there is $\rho_c \in (0,1)$, such that the public likelihood that $\Theta = \overline{\theta}$ is close to $+\infty$ (resp. close to zero) if $\rho^N < \rho_c$ (resp. if $\rho^N > \rho_c$). Hence, except in the unlikely event where $\rho^N = \rho_c$, herding takes place at stage 2.

To summarize, as the number N of players increases to $+\infty$:

Stage 1: $\sup_{i=1,2,\ldots,N} |\pi_1^{i,N}(\blacktriangle) - \pi_c|$ converges to zero.

Stage 2: Let a list $\vec{\alpha}^N$ be given that specifies the status of the other players at stage 2, and such that the fraction $\rho^N(\vec{\alpha}^N)$ of drop outs converges to $\rho \in [0,1]$:

- If $\rho > \rho_c$, the maximal cutoff $\max_i \pi_2^{i,N}(\vec{\alpha}^N)$ converges to 0.
- If $\rho < \rho_c$, the minimal cutoff $\min_i \pi_2^{i,N}(\vec{\alpha}^N)$ converges to 1.

In a sense, this result provides an asymptotic version of the results of Caplin and Leahy (1994), which deal with the continuum-of-players case.

2. COMMENTS AND EXTENSIONS

2.1. *Reversible Decisions*

In the one-player bandit problem, assuming reversible decisions rather than irreversible ones does not change the equilibrium behavior. This property does not extend to the two-player case, since in the latter, free-riding effects appear.

The true multiplayer generalization of one-arm bandit problems allows the players to alternate between the two arms, as in Bolton and Harris (1999). The definition of cutoff strategies readily adapts to that case. However, public information is then more cumbersome, since it coincides with the list of the arms selected by each of the players in the past.

In this case, given an equilibrium, let $\vec{t}_n$ denote the public information available at stage n. The sequence $(p_n^i, \vec{t}_n)$ is a Markov chain for the filtration $(\mathcal{G}_n^i)$, so that the optimal continuation payoff W_n^i of player i can be written as a function of $(p_n^i, \vec{t}_n)$. Therefore, in equilibrium the players play as a function of their private beliefs and public information. We do not know whether all equilibria in this case are in cutoff strategies.

2.2. *Timing Games and Real Options*

In our model, a player keeps receiving payoffs until she drops out. In most real option problems, each player i chooses a time n at which she starts to receive a stream of payoffs $\tilde{X}_n^i, \tilde{X}_{n+1}^i, \ldots$. Our results apply to this situation as well. Indeed, observe that, given any strategy profile ϕ, the sum of the payoffs received by player i in the timing game we study and in the real options game is $\sum_{n=0}^{+\infty} \tilde{X}_n^i$, and is, therefore, *independent* of the strategy profile. As a result, the equilibria of the real option game coincide with the equilibria of the timing game with stage payoffs $X_n^i := -\tilde{X}_n^i$.

2.3. *Equilibrium Uniqueness*

In general, the equilibrium need not be unique. To see this, assume that the initial prior p_0 coincides with the optimal one-player cutoff π_*. One equilibrium is that both players drop out immediately. It can be shown that there is *another* symmetric equilibrium in which both players are active at stage 0: the fact that player j is active at stage 0 creates a positive informational externality that induces player i to enter as well. The uniqueness issue remains open in the following special cases: (a) when one restricts attention to equilibria in which *both* players are active at stage 0 and (b) when $p_0 > \pi_*$.

2.4. *Efficiency*

First-order efficiency would imply that both players drop out at stage 0 if $\Theta = \underline{\theta}$ and that both players stay in forever if $\Theta = \overline{\theta}$. Plainly, this cannot be achieved. Lemma 18 below proves that if the machines are high, then there is a positive probability that both players stay in forever. It is not difficult to prove that if both machines are low, then with probability 1 both players drop out in finite time. As discussed in Section 1.2.1, sharper results are obtained when the number of players gets large.

Laboratoire d'Analyse Géométrie et Applications, Institut Galilée, Université Paris Nord, avenue Jean-Baptiste Clément, 93430 Villetaneuse, France, and Laboratoire d'Econométrie de l'Ecole Polytechnique, Paris, France; dinah@math.univ-paris13.fr,

The School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel; eilons@post.tau.ac.il,

and

Dépt. Finance et Economie, HEC, 1 rue de la Libération, 78 351 Jouy-en-Josas, France; vieille@hec.fr.

Manuscript received January, 2006; final revision received April, 2007.

APPENDIX

For most of the appendix, we let a mixed strategy, ϕ^j, of player j be given. The payoff function to player i, $\gamma^i(\cdot, \phi^j)$, does not depend on player j's strategic decisions $\tau^j(k)$, once player i has dropped out. In the analysis of player i's best replies to ϕ^j, we therefore may, and will, assume that player j's behavior is independent of player i's actions, that is, $\tau^j(k) = \tau^j(\blacktriangle)$ for each $k \in \mathbf{N}$.

Accordingly, we denote by τ^j the unique stopping time that governs player j's decisions.

Section A contains preliminary material on beliefs. Each of the following three sections is devoted to the proof of one theorem.

A: BELIEFS

We first state the basic stochastic dominance properties of the private beliefs. Then we examine the Markov property of the sequence of beliefs w.r.t. various sequences of σ-algebras.

A.1. *Stochastic Dominance*

Recall that $\mathcal{F}_n^i := \sigma(X_0^i, \dots, X_{n-1}^i)$ is the private information of player i prior to some stage n and that the private belief of player i at that time is defined as $p_n^i := \mathbf{P}(\Theta = \overline{\theta} \mid \mathcal{F}_n^i)$.

By Bayes' rule, a version of p_n^i is given by

$$\frac{p_n^i}{1 - p_n^i} = \frac{p_0}{1 - p_0} \times \prod_{k=0}^{n-1} \frac{f_{\overline{\theta}}(X_k^i)}{f_{\underline{\theta}}(X_k^i)}. \tag{1}$$

It is well known that the sequence (p_n^i) is a martingale (resp. a submartingale, a supermartingale) under $\mathbf{P}$ (resp. under $\mathbf{P}_{\overline{\theta}}, \mathbf{P}_{\underline{\theta}}$). In addition, the law of p_n^i under $\mathbf{P}_{\overline{\theta}}$ stochastically dominates (in the first-order sense) the law of p_n^i under $\mathbf{P}_{\underline{\theta}}$: the private belief tends to be higher when the state is $\overline{\theta}$ than when it is not. A slightly stronger statement holds here. We omit the proof.[7]

LEMMA 6: *One has* $\mathbf{P}_{\overline{\theta}}(p_n^i \leq p) < \mathbf{P}_{\underline{\theta}}(p_n^i \leq p)$ *as soon as* $\mathbf{P}_{\underline{\theta}}(p_n^i \leq p) > 0$ *and* $\mathbf{P}_{\overline{\theta}}(p_n^i \leq p) < 1$.

This dominance property extends to vectors of beliefs. Again, we omit the proof.

LEMMA 7: *For each stage* $n \in \mathbf{N}$ *and* $x_1, \dots, x_n \in [0, 1]$, *one has*

$$\mathbf{P}_{\overline{\theta}}\big((p_1^i, \dots, p_n^i) \leq (x_1, \dots, x_n)\big) \leq \mathbf{P}_{\underline{\theta}}\big((p_1^i, \dots, p_n^i) \leq (x_1, \dots, x_n)\big).$$

Moreover, $\mathbf{P}_{\underline{\theta}}((p_1^i, \dots, p_n^i) > (x_1, \dots, x_n)) \leq \mathbf{P}_{\overline{\theta}}((p_1^i, \dots, p_n^i) > (x_1, \dots, x_n))$.

[7]All omitted proofs can be found in Rosenberg, Solan, and Vieille (2004).

A.2. *Markov Properties*

We denote by t_n^j the *status* of player j at stage n: $t_n^j := \blacktriangle$ if $\tau^j \geq n$ and $t_n^j = k$ if $\tau^j = k < n$. Since player j's strategy is given, her past decisions are informative and t_n^j is a well defined random variable. Prior to stage $n \geq 0$, player i's information over Ω is given by the σ-algebra $\mathcal{G}_n^i := \sigma(\mathcal{F}_n^i, t_n^j)$, and we define her *posterior belief* as $q_n^i := \mathbf{P}(\Theta = \overline{\theta} \mid \mathcal{G}_n^i)$. Our goal is to establish Proposition 9 below.

Conditional on the arm type Θ, the status t_n^j is independent of the payoffs $(X_k^i)_{k\geq 0}$, hence also of $\mathcal{F}_n^i$. Thus, a version of the posterior belief is given by

$$\frac{q_n^i}{1-q_n^i} = \frac{p_n^i}{1-p_n^i} \times \frac{\mathbf{P}_{\overline{\theta}}(t_n^j = \alpha)}{\mathbf{P}_{\underline{\theta}}(t_n^j = \alpha)} \quad \text{whenever} \quad t_n^j = \alpha. \tag{2}$$

In particular, q_n^i is measurable w.r.t. the pair (p_n^i, t_n^j). For later use, note also that the version $Q_n^i(p_n^i, t_n^j)$ of q_n^i given by (2) is continuous and increasing in p_n^i.

We now prove that the pair (p_n^i, t_n^j) is a Markov chain. This will later allow us to prove that (p_n^i, t_n^j) contains all relevant information for player i's best-reply problem.

We use the following definition.

DEFINITION 8—Shiryaev (1996, p. 564): Let $(\mathcal{G}_n)_{n\in\mathbf{N}}$ be a filtration over a probability space $(\Omega, \mathbf{P})$ and let $(A_n)_{n\in\mathbf{N}}$ be a sequence of random variables, adapted to $(\mathcal{G}_n)_{n\in\mathbf{N}}$ and with values in $\mathbf{R}^d$. The sequence (A_n) is a *Markov chain* for the filtration $(\mathcal{G}_n)_{n\in\mathbf{N}}$ if, for each Borel set $B \subseteq \mathbf{R}^d$ and $n \in \mathbf{N}$, one has $\mathbf{P}(A_{n+1} \in B \mid \mathcal{G}_n) = \mathbf{P}(A_{n+1} \in B \mid A_n)$ $\mathbf{P}$-a.s.

PROPOSITION 9: *Under* $\mathbf{P}$, *the sequence* $(p_n^i, t_n^j)_{n\in\mathbf{N}}$ *is a Markov chain for* $(\mathcal{G}_n^i)$.

We will use the following technical observation, stated without proof.

LEMMA 10: *Let* $\mathcal{H}^1$ *and* $\mathcal{H}^2$ *be two independent* σ-*algebras on a probability space* $(\Omega, \mathbf{P})$ *and let* $\mathcal{A}^i$ *be a sub-*σ*-field of* $\mathcal{H}^i$, $i = 1, 2$. *For each* $C^1 \in \mathcal{H}^1$, $C^2 \in \mathcal{H}^2$, *one has*

$$\mathbf{P}(C^1 \cap C^2 \mid \sigma(\mathcal{A}^1, \mathcal{A}^2)) = \mathbf{P}(C^1 \mid \mathcal{A}^1) \times \mathbf{P}(C^2 \mid \mathcal{A}^2).$$

PROOF OF PROPOSITION 9: Observe first that the sequence $(p_n^i)_{n\geq 0}$ is a Markov chain for $(\mathcal{F}_n^i)_{n\geq 0}$ under $\mathbf{P}_\theta$. Let a stage $n \geq 0$ be given, let $B \subseteq [0, 1]$ be a Borel set, and fix $\alpha \in \{\blacktriangle\} \cup \mathbf{N}$.

Under $\mathbf{P}_\theta(\theta \in \{\underline{\theta}, \overline{\theta}\})$, the σ-algebra $\mathcal{F}_{n+1}^i$ and the r.v. t_{n+1}^j are independent. By Lemma 10, this implies

$$\mathbf{P}_\theta(p_{n+1}^i \in B, t_{n+1}^j = \alpha \mid \mathcal{G}_n^i) = \mathbf{P}_\theta(p_{n+1}^i \in B \mid \mathcal{F}_n^i) \times \mathbf{P}_\theta(t_{n+1}^j = \alpha \mid t_n^j).$$

Since (p_n^i) is a Markov chain under $\mathbf{P}_\theta$, $\mathbf{P}_\theta(p_{n+1}^i \in B \mid \mathcal{F}_n^i) = \mathbf{P}_\theta(p_{n+1}^i \in B \mid p_n^i)$. Therefore, $\mathbf{P}_\theta(p_{n+1}^i \in B, t_{n+1}^j = \alpha \mid \mathcal{G}_n^i)$ is measurable w.r.t. the pair (p_n^i, t_n^j).

Since $q_n^i = \mathbf{P}(\Theta = \overline{\theta} \mid \mathcal{G}_n^i)$ is measurable w.r.t. (p_n^i, t_n^j), the conditional probability

$$\begin{aligned}
&\mathbf{P}(p_{n+1}^i \in B, t_{n+1}^j = \alpha \mid \mathcal{G}_n^i) \\
&\quad = \sum_\theta \mathbf{P}(\Theta = \theta \mid \mathcal{G}_n^i) \times \mathbf{P}_\theta(p_{n+1}^i \in B, t_{n+1}^j = \alpha \mid \mathcal{G}_n^i)
\end{aligned}$$

is also measurable w.r.t. (p_n^i, t_n^j), hence

$$\mathbf{P}(p_{n+1}^i \in B, t_{n+1}^j = \alpha \mid \mathcal{G}_n^i) = \mathbf{P}(p_{n+1}^i \in B, t_{n+1}^j = \alpha \mid p_n^i, t_n^j)$$

as desired. *Q.E.D.*

B: PROOF OF THEOREM 4

When facing ϕ^j, player i must choose when to stop; that is, a stopping time for the filtration $(\mathcal{G}_n^i)_{n\geq 0}$. If she stops at stage n, her overall realized payoff is $Y_n^i := \sum_{k=0}^{n-1} \delta^k X_k^i$. Hence, player i's best-reply problem is equivalent to the optimal stopping problem,

$$\text{problem } \mathcal{P}\colon \text{ find a solution to } \sup_\sigma \mathbf{E}[Y_\sigma^i],$$

where the supremum is taken over all stopping times σ for $(\mathcal{G}_n^i)$. That is, any best reply to ϕ^j yields an optimal solution to $(\mathcal{P})$ and vice versa. We will prove that $(\mathcal{P})$ admits a unique optimal stopping time, which moreover corresponds to a cutoff strategy.

We first recall standard material on optimal stopping problems.

STEP 0: *Optimal stopping problems*. Given $n \geq 0$, we let Λ_n be the set of stopping times $\sigma \geq n$ ($\mathbf{P}$-a.s.). The *Snell envelope* of the sequence (Y_n^i) is defined to be the sequence (V_n^i), where

$$V_n^i := \operatorname*{ess\,sup}_{\sigma\in\Lambda_n} \mathbf{E}[Y_\sigma^i \mid \mathcal{G}_n^i].$$

It is the optimal payoff to player i when she is constrained not to drop out before stage n. The following lemma is well known; see, for example, Chow and Robbins (1963), Ferguson (2004, Chap. 3), or Neveu (1972).

LEMMA 11: *The stopping time* $\sigma^* := \inf\{n \geq 0 : V_n^i = Y_n^i\}$ (*with* $\inf \emptyset = +\infty$) *is a solution to* $\mathcal{P}$. *Moreover*, $\sigma \geq \sigma^*$ *for every optimal stopping time* σ.

The Snell envelope (V_n^i) can be obtained as the (**P**-a.s.) limit of the optimal payoff in the finite horizon versions of $(\mathcal{P})$. To be specific, define $V_{n,k}^i$ for $n, k \geq 0$ by $V_{n,0}^i = Y_n^i$, and

$$V_{n,k+1}^i = \max\{Y_n^i, \mathbf{E}[V_{n+1,k}^i \mid \mathcal{G}_n^i]\} \quad \text{for every } n \geq 0 \text{ and } k \geq 1. \tag{3}$$

In (3), $V_{n,k+1}^i$ is the optimal payoff to player i when she is constrained not to drop out before stage n, but must drop out at stage $n+k+1$ at the latest. Thus, (3) is a dynamic programming principle: the optimal payoff is obtained as the maximum of the two choices that are available at stage n, dropping out or staying active. Then $V_n^i = \lim_{k\to\infty} V_{n,k}^i$.

The quantity $V_n^i - Y_n^i$ is the optimal payoff from stage n onward. We first prove that this quantity is measurable w.r.t. (p_n^i, t_n^j). We then define W_n^i to be the optimal continuation payoff, and show that it is continuous and increasing in p_n^i. We then conclude that σ^* is the only optimal stopping time and that it corresponds to a cutoff strategy.

STEP 1: *The sequence* (W_n^i).

LEMMA 12: $V_n^i - Y_n^i$ *is measurable w.r.t. the pair* (p_n^i, t_n^j).

Thus, $V_n^i - Y_n^i$ coincides (**P**-a.s.) with some function $\eta_n^i(p_n^i, t_n^j)$.

PROOF OF LEMMA 12: We adapt the proof from Neveu (1972). By (3), one has

$$V_{n,k+1}^i - Y_n^i = \max\{0, \delta^n \mathbf{E}[X_n^i \mid \mathcal{G}_n^i] + \mathbf{E}[V_{n+1,k}^i - Y_{n+1}^i \mid \mathcal{G}_n^i]\},$$
$$\text{for every } n \geq 0 \text{ and } k \geq 1.$$

Observe that the expected current payoff, $\mathbf{E}[X_n^i \mid \mathcal{G}_n^i] = q_n^i \mathbf{E}_{\overline{\theta}}[X_n^i] + (1 - q_n^i)\mathbf{E}_{\underline{\theta}}[X_n^i]$ is measurable w.r.t. (p_n^i, t_n^j).

On the other hand, since the sequence (p_n^i, t_n^j) is a Markov chain for $(\mathcal{G}_n^i)$, the r.v. $\mathbf{E}[V_{n+1,k}^i - Y_{n+1}^i \mid \mathcal{G}_n^i]$ is measurable w.r.t. (p_n^i, t_n^j) as soon as $V_{n+1,k}^i - Y_{n+1}^i$ is measurable w.r.t. (p_{n+1}^i, t_{n+1}^j). It thus follows inductively that $V_{n,k}^i - Y_n^i$ is measurable w.r.t. (p_n^i, t_n^j) for each k. The result follows by letting $k \to +\infty$.

Q.E.D.

We define $W_n^i := \delta^n \mathbf{E}[X_n^i \mid \mathcal{G}_n^i] + \mathbf{E}[V_{n+1}^i - Y_{n+1}^i \mid \mathcal{G}_n^i]$. This is player i's optimal continuation payoff if she decides to remain active. One has $V_n^i - Y_n^i = \max\{0, W_n^i\}$. As in the proof of Lemma 12, W_n^i is measurable w.r.t. (p_n^i, t_n^j).

Set $\tau_* := \inf\{k > n : V_k^i = Y_k^i\}$, so that $V_{n+1}^i = \mathbf{E}[Y_{\tau_*}^i \mid \mathcal{G}_{n+1}^i]$. The stopping time τ_* is optimal for player i, assuming she is active at stage n. It follows that $W_n^i = \mathbf{E}[Y_{n\to\tau_*}^i \mid \mathcal{G}_n^i]$, where $Y_{n\to\tau_*}^i = Y_{\tau_*}^i - Y_n^i = \sum_{k=n}^{\tau_*-1} \delta^k X_k^i$ is the sum of payoffs received from stage n up to stage τ_*.

STEP 2: *Regularity properties*. We here state and prove Lemmas 13 and 14.

LEMMA 13: *The r.v. W_n^i has a version $\omega_n^i(p_n^i, t_n^j)$, such that for fixed α, the map $\omega_n^i(\cdot, \alpha)$ is continuous on the support of p_n^i.*

Plainly, it will be sufficient to consider α's such that $\mathbf{P}(t_n^j = \alpha) > 0$. For such α's, the map $\omega_n^i(\cdot, \alpha)$ need not be uniquely defined. Indeed, if the distributions of payoffs are such that the private belief p_1^i cannot fall below, say, some level $\lambda > 0$, the definition of $\omega_1^i(p, \alpha)$ for $p < \lambda$ is completely arbitrary. However, if $\mathbf{P}(t_n^j = \alpha) > 0$, the *restriction* of $\omega_n^i(\cdot, \alpha)$ to the support of p_n^i is uniquely defined.

LEMMA 14: *For fixed α, the map $\omega_n^i(\cdot, \alpha)$ is increasing on the support of p_n^i.*

We will use the following technical lemma, which follows from Lusin's theorem.

LEMMA 15: *Let ν be a probability measure over $\mathbf{R}$, absolutely continuous w.r.t. the Lebesgue measure, and let $B \subseteq \mathbf{R}$ be a Borel set. Then the map $x \in \mathbf{R} \mapsto \nu(x + B)$ is continuous.*

PROOF OF LEMMA 13: W_n^i may also be expressed as

$$\begin{aligned} W_n^i &= \mathbf{E}[Y^i_{n\to\tau_*} \mid p_n^i, t_n^j] \\ &= \sum_{\theta\in\{\underline{\theta},\overline{\theta}\}} \mathbf{P}(\Theta = \theta \mid p_n^i, t_n^j)\mathbf{E}_\theta[Y^i_{n\to\tau^*} \mid p_n^i, t_n^j] \\ &= \sum_{\theta\in\{\underline{\theta},\overline{\theta}\}} \mathbf{P}(\Theta = \theta \mid p_n^i, t_n^j) \sum_{k\geq n} \delta^k\ \theta\, \mathbf{P}_\theta(\tau^* > k \mid p_n^i, t_n^j). \end{aligned}$$

The belief $\mathbf{P}(\Theta = \overline{\theta} \mid p_n^i, t_n^j)$ has a continuous version, q_n^i. It is therefore sufficient to construct a continuous version of $\mathbf{P}_\theta(\tau^* > k \mid p_n^i, t_n^j)$ for each $k \geq n$.

To be concise, we let $\vec{t}^j$ stand for the vector $(t_{n+1}^j, \dots, t_k^j)$ and let $\vec{\alpha} = (\alpha_{n+1}, \dots, \alpha_k)$ denote generic values of $\vec{t}^j$.

Observe first that $\tau^* > k$ if and only if[8] $\eta_m^i(p_m^i, t_m^j) > 0$ for all $n < m \leq k$. For a given α, define $G_m(\alpha) := \{p : \eta_m^i(p, \alpha) > 0\}$ to be those beliefs at which player i remains active at stage m. Thus, on the event $\vec{t}^j = \alpha$, one has $\tau^* > k$ if and only if $p_m^i \in G_m(\alpha_m)$ for all $n < m \leq k$.

By (1), the private belief p_m^i is related to p_n^i through the equality

$$\ln\frac{p_m^i}{1-p_m^i} = \ln\frac{p_n^i}{1-p_n^i} + \ln\frac{f_{\overline{\theta}}(X_n^i)}{f_{\underline{\theta}}(X_n^i)} + \sum_{s=n+1}^{m-1} \ln\frac{f_{\overline{\theta}}(X_s^i)}{f_{\underline{\theta}}(X_s^i)}, \tag{4}$$

[8]Recall that $\eta_m^i(p_m^i, t_m^j) = V_m^i - Y_m^i$ is the optimal continuation payoff at stage m.

so that on the event $\vec{t}^j = \vec{\alpha}$, one has $\tau^* > k$ if and only if

$$\ln \frac{p_n^i}{1-p_n^i} + \ln \frac{f_{\overline{\theta}}(X_n^i)}{f_{\underline{\theta}}(X_n^i)} + \sum_{s=n+1}^{m-1} \ln \frac{f_{\overline{\theta}}(X_s^i)}{f_{\underline{\theta}}(X_s^i)} \in F_m(\alpha_m), \quad \text{for all } n < m \leq k, \tag{5}$$

where $F_m(\alpha_m)$ is the image of the set $G_m(\alpha_m)$ under the map $x \mapsto \ln \frac{x}{1-x}$.

With obvious notations, (5) is in turn equivalent to

$$\ln \frac{f_{\overline{\theta}}(X_n^i)}{f_{\underline{\theta}}(X_n^i)} \in F_m(\alpha_m, X_{n+1}^i, \ldots, X_{k-1}^i) - \ln \frac{p_n^i}{1-p_n^i}, \quad \text{for all } n < m \leq k.$$

Finally, set $F(\vec{\alpha}, x_{n+1}, \ldots, x_{k-1}) := \bigcap_{m=n+1}^{k} F_m(\alpha_m, x_{n+1}, \ldots, x_{k-1})$, so that

$$\begin{aligned} &\mathbf{P}_\theta(\tau_* > k, \vec{t}^j = \vec{\alpha} \mid (p_n^i, t_n^j)) \\ &\quad = \mathbf{P}_\theta(\vec{t}^j = \vec{\alpha} \mid t_n^j) \\ &\qquad \times \mathbf{P}_\theta\left(\ln \frac{f_{\overline{\theta}}(X_n^i)}{f_{\underline{\theta}}(X_n^i)} \in F(\vec{\alpha}, X_{n+1}^i, \ldots, X_{k-1}^i) - \ln \frac{p_n^i}{1-p_n^i} \,\middle|\, p_n^i \right). \end{aligned} \tag{6}$$

Conditional on $\Theta = \theta$, the private belief p_n^i is independent of the future payoffs X_m^i, $n \leq m < k$. Therefore, a version of the conditional probability on the right-hand side of (6) is given by the integral

$$\begin{aligned} &\int_{\mathbf{R}^{k-n}} \mathbf{1}_{\{\ln \frac{f_{\overline{\theta}}(x_n)}{f_{\underline{\theta}}(x_n)} \in F(\vec{\alpha}, x_{n+1}, \ldots, x_{k-1}) - \ln \frac{p_n^i}{1-p_n^i}\}} \prod_{m=n}^{k-1} d\mathbf{P}_\theta(x_m) \\ &\quad = \int_{\mathbf{R}^{k-n-1}} \left\{ \int_{\mathbf{R}} \mathbf{1}_{\{\ln \frac{f_{\overline{\theta}}(x_n)}{f_{\underline{\theta}}(x_n)} \in F(\vec{\alpha}, x_{n+1}, \ldots, x_{k-1}) - \ln \frac{p_n^i}{1-p_n^i}\}} d\mathbf{P}_\theta(x_n) \right\} \\ &\qquad \times \prod_{m=n+1}^{k-1} d\mathbf{P}_\theta(x_m). \end{aligned}$$

Observe now that the inner integral is equal to $\nu(F(\vec{\alpha}, x_{n+1}, \ldots, x_{k-1}) - \ln \frac{p_n^i}{1-p_n^i})$, where ν is the law under $\mathbf{P}_\theta$ of the random variable $\ln \frac{f_{\overline{\theta}}(X_n^i)}{f_{\underline{\theta}}(X_n^i)}$. By assumption, the latter variable has a density, so that ν is absolutely continuous w.r.t. Lebesgue measure. Hence, by Lemma 15, this inner integral is continuous w.r.t. p_n^i. By dominated convergence, the integral in (7) is also continuous w.r.t. p_n^i.

By plugging (7) into the right-hand side of (6), one obtains a version of $\mathbf{P}_\theta(\tau_* > k, \vec{t}^j = \vec{\alpha} \mid p_n^i, t_n^j)$, which is continuous in p_n^i. By summing over $\vec{\alpha}$, one then obtains a version of $\mathbf{P}_\theta(\tau_* > k \mid p_n^i, t_n^j)$, which is continuous in p_n^i, as desired. *Q.E.D.*

PROOF OF LEMMA 14: Given a version of p_n^i, we use (4) to choose a version for $p_m^i = p_m^i(p_n^i, X_n^i, \dots, X_{m-1}^i)$ $(m > n)$. Fix p in the support of p_n^i. Define the stopping time

$$\sigma_p := \inf\{k \geq n+1 : \omega_k^i(p_k^i(p, X_n^i, \dots, X_{k-1}^i), t_k^j) \leq 0\}.$$

Under σ_p, player i behaves as if she had reached stage n with a private belief equal to p, and would play τ^*. Thus, conditional on θ, her continuation payoff $\mathbf{E}_\theta[Y_{n\to\sigma_p}^i \mid \mathcal{G}_n^i]$ does not depend on p_n^i. One version of this continuation payoff is $C(p_n^i, t_n^j; p) := q_n^i(p_n^i, t_n^j)\mathbf{E}_{\overline{\theta}}[Y_{n\to\sigma_p}^i] + (1 - q_n^i(p_n^i, t_n^j))\mathbf{E}_{\underline{\theta}}[Y_{n\to\sigma_p}^i]$, which is continuous and increasing in p_n^i.

Fix the version of τ_* to be $\tau_* = \inf\{k > n : \omega_k^i(p_k^i(p_n^i, X_n^i, \dots, X_{m-1}^i), t_k^j) \leq 0\}$. By construction, one has $C(p, t_n^j; p) = \omega_n^i(p, t_n^j)$. This inequality holds everywhere, and not only $\mathbf{P}$-a.s., since both C and ω_n^i are continuous (see Lemma 13 for the latter). For the same reason, and since $\omega_n^i(p_n^i, t_n^j)$ is the highest continuation payoff, one has $C(p_n^i, t_n^j; p) \leq \omega_n^i(p_n^i, t_n^j)$ everywhere. Since C is increasing, this implies for every p' higher than p in the support of p_n^i and every α,

$$w_n^i(p, \alpha) = C(p, \alpha; p) < C(p', \alpha; p) \leq w_n^i(p', \alpha). \tag{7}$$

We obtained that $w_n^i(\cdot, \alpha)$ is increasing in p, as desired. *Q.E.D.*

STEP 3: *Conclusion*. We here state and prove Lemma 16, which concludes the proof.

LEMMA 16: *The stopping time σ^* is the only optimal solution to $\mathcal{P}$. Moreover, it corresponds to a cut-off strategy.*

PROOF: We start with the first assertion. Let σ be a solution to $\mathcal{P}$. By Lemma 11, $\sigma \geq \sigma_*$. Fix a stage n. By Lemma 14, and since p_n^i has a density, one has $\omega_n^i(p_n^i, t_n^j) < 0$ on the event $\Omega_n := \{\sigma_* = n < \sigma\}$. In particular, $\mathbf{E}[Y_\sigma^i \mathbf{1}_{\Omega_n}] \leq \mathbf{E}[Y_{\sigma^*}^i \mathbf{1}_{\omega_n}]$, with a strict inequality if $\mathbf{P}(\Omega_n) > 0$. Since $\mathbf{E}[Y_\sigma^i] = \mathbf{E}[Y_{\sigma^*}^i]$, one must have $\mathbf{P}(\Omega_n) = 0$ for each n.

We now turn to the second claim. If $\mathbf{P}(t_n^j = \alpha) > 0$, Lemmas 13 and 14 provide us with a version $\omega_n^i(\cdot, \alpha)$ that is continuous and increasing over the support of p_n^i. We extend it to a continuous, increasing, function defined over $[0, 1]$ and define $\pi_n^i(\alpha)$ to be the unique value of p such that $w_n^i(p, \alpha) = 0$. If instead $\mathbf{P}(\tau^j(\alpha)) = 0$, we choose $\pi_n^i(\alpha) \in [0, 1]$ in an arbitrary way.

It is immediate to check that $\sigma^* = \inf\{n : p_n^i \leq \pi_n^i(t_n^j)\}$ ($\mathbf{P}$-a.s.). Hence, σ^* is a cutoff strategy. *Q.E.D.*

REMARK: σ^* is the unique optimal stopping time (up to $\mathbf{P}$-null sets). However, the associated cutoffs need not be uniquely defined, for two reasons.

Consider first α such that $\mathbf{P}(t_n^j = \alpha) > 0$. If the value of $\pi_n^i(\alpha)$, as obtained in the previous proof, falls outside the support of p_n^i, then its value depends on the choice of the extension of w_n^i. However, this indeterminacy is only apparent, as these correspond to beliefs that are reached with probability zero: the corresponding stopping time $\tau^i(\alpha)$ is uniquely defined (up to **P**-null sets).

If $\mathbf{P}(t_n^j = \alpha) = 0$, then any choice for $\pi_n^i(\alpha)$ is admissible, and different choices may yield different stopping times $\tau^i(\alpha)$. Again, this indeterminacy is only apparent, since, against τ^j, the stopping time $\tau^i(\alpha)$ will "never" be used.

If the cutoff p does not belong to the support of p_n^i, then changing the definition of ω_n^i around p would change the cutoff as well. This indeterminacy is only apparent if p corresponds to a belief that is reached with probability zero. In other words, the best-reply is unique, even if there may be different cutoff sequences associated with it.

C: PROOF OF THEOREM 3

The existence of a symmetric equilibrium derives from a standard fixed-point argument. A cutoff strategy of player i is a sequence $(\pi_n^i(k))$ indexed by $n \geq 0$ and $k \in \{\blacktriangle, 1, \dots, n-1\}$, with values in $[0, 1]$. The set Φ of such sequences is compact when endowed with the product topology.

Player i's best-reply map is given by $B(\phi_j) := \{\phi_i \in \Phi : \gamma^i(\phi_i, \phi_j) = \max_\Phi \gamma^i(\cdot, \phi_j)\}$ for each cutoff strategy $\phi_j \in \Phi$. From the analysis of the previous section, $B(\phi')$ is convex-valued.

We now check that the payoff function γ^i is continuous over the space of cutoff profiles. Let a sequence (ϕ_m) of cutoff profiles be given, that converges to ϕ (in the product topology). The realized payoff $r^i(\phi_m)$ converges to $r^i(\phi)$, except possibly if the belief is equal to the cutoff: $p_n^i = \pi_n^i(\alpha)$ for some $n \geq 0$, $\alpha = \blacktriangle, 0, 1, \dots, n-1$, and $i = 1, 2$. Since the law of p_n^i has a density, this event has **P**-measure zero. Since $|r^i(\phi_m)| \leq \sup_{n \in \mathbf{N}} |Y_n^i|$, the dominated convergence theorem applies and $\lim_{m \to \infty} \gamma^i(\phi_m) = \gamma^i(\phi)$.

Since γ^i is continuous, B is upper hemi continuous Since Φ is compact and, by Glicksberg's (1952) generalization of Kakutani's fixed-point theorem, B has a fixed point, ϕ_*. Plainly, the profile (ϕ_*, ϕ_*) is an equilibrium.

D: PROOF OF THEOREM 5

We here prove the qualitative results listed in Theorem 5. Most proofs have to do with the impact of the public information on the posterior belief.

We fix a cutoff strategy ϕ^j of player j. Denote by π_n^j the cutoff used by player j if player i is still active at that stage.

D.1. *Proof of P3*

Player j stops at stage $\tau^j := \inf\{n : p_n^j \leq \pi_n^j\}$. In particular, by Lemma 7, player j tends to stop earlier if the state is $\underline{\theta}$: $\mathbf{P}_{\overline{\theta}}(\tau^j \geq n) \leq \mathbf{P}_{\underline{\theta}}(\tau^j \geq n)$. By (2) this implies that $q_n^i \geq p_n^i$ whenever player j is active: having an active opponent is good news.

If at stage n player i chooses not to watch player j any longer, she faces a one-player problem, in which her continuation payoff is positive once her posterior belief exceeds π_*. In the two-player game, player i has more options if player j is still active, hence player i's continuation payoff is positive as well. Since $q_n^i \geq p_n^i$, whenever her private belief p_n^i exceeds π_*, player i continues. This readily implies that $\pi_n^i(\blacktriangle) \leq \pi_*$: the first part of P3 follows.

We now prove that having an active opponent is always the best possible news on θ. Recall that $Q_n^i(p_n^i, t_n^j)$ is the version of q_n^i given by (2).

LEMMA 17: *One has $Q_n^i(p, \blacktriangle) \geq Q_n^i(p, m)$ for every $m < n$ and $p \in [0, 1]$.*

PROOF: We introduce an auxiliary family of beliefs and set $p_{n,m}^{i,j} := \mathbf{P}(\Theta = \overline{\theta} \mid \mathcal{F}_n^i, \mathcal{F}_m^j)$, $m \leq n$. The belief $p_{n,m}^{i,j}$ is computed by collecting the private information held by the two players at two *possibly different* stages n and m. (A version of) $p_{n,m}^{i,j}$ is given by

$$\frac{p_{n,m}^{i,j}}{1 - p_{n,m}^{i,j}} = \frac{p_n^i}{1 - p_n^i} \times \frac{p_m^j}{1 - p_m^j} \times \frac{1 - p_0}{p_0},$$

hence, it is a continuous and increasing function of p_n^i and p_m^j, that we denote $p_{n,m}^{i,j}(\cdot, \cdot)$.

We also let $Q_{n,m}^i(p_n^i, t_m^j) := \mathbf{P}(\Theta = \overline{\theta} \mid p_n^i, t_m^j)$. By the law of iterated conditional expectations, and since t_m^j is a function of $p_1^j, \ldots, p_m^j$, one has

$$\begin{aligned} Q_{n,m}^i(p_n^i, t_m^j) &= \mathbf{E}\big[\mathbf{P}(\Theta = \overline{\theta} \mid p_n^i, p_1^j, \ldots, p_m^j) \mid p_n^i, t_m^j\big] \\ &= \mathbf{E}[p_{n,m}^{i,j} \mid p_n^i, t_m^j]. \end{aligned}$$

We will prove that the following two inequalities hold:

$$Q_{n,m}^i(p, m-1) < Q_{n,m}^i(p, \blacktriangle) < Q_{n,m+1}^i(p, \blacktriangle). \tag{8}$$

According to the second inequality, at stage n, knowing that player j was active at stage $m + 1$ is better news than knowing she was active at stage m. When iterated, this inequality yields $Q_{n,m}^i(p, \blacktriangle) < Q_{n,n}^i(p, \blacktriangle) = Q_n^i(p, \blacktriangle)$.

According to the first inequality, at stage n, it is better news to learn that player j chose to remain active at stage $m - 1$ than to discover that she dropped out at stage $m - 1$. Once player j drops out at stage m, player j's state cannot possibly change; hence, $Q_{n,n}^i(p, m) = Q_{n,m}^i(p, m)$. The result thus follows from (8).

We start with the first inequality in (8). If $t_m^j = \blacktriangle$, one has $p_m^j > \pi_m^j(\blacktriangle)$. Since $p_{n,m}^{i,j}$ is monotonic in p_m^j, $Q_{n,m}^i(p_n^i, t_m^j) > p_{n,m}^{i,j}(p_n^i, \pi_m^j(\blacktriangle))$. By contrast, if $t_m^j =$

$m-1$, one has $p^j_m \le \pi^j_m(\blacktriangle)$, hence $Q^i_{n,m}(p^i_n, t^j_m) \le p^{i,j}_{n,m}(p^i_n, \pi^j_m(\blacktriangle))$. Combining the two inequalities yields $Q^i_{n,m}(p, m-1) < Q^i_{n,m}(p, \blacktriangle)$, as desired.

We turn to the second inequality in (8). Using once more the law of iterated conditional expectations, one has

$$Q^i_{n,m}(p^i_n, t^j_m) = \mathbf{E}[\mathbf{P}(\Theta = \overline{\theta} \mid p^i_n, t^j_{m+1}) \mid p^i_n, t^j_m] = \mathbf{E}[Q^i_{n,m+1} \mid p^i_n, t^j_m].$$

It follows that $Q^i_{n,m}(p, \blacktriangle)$ is a convex combination of $Q^i_{n,m+1}(p, \blacktriangle)$ and $Q^i_{n,m+1}(p, m)$. By the first inequality, the former is higher than the latter and the result follows. *Q.E.D.*

D.2. *Proof of P2*

P2 follows from the next two lemmas.

LEMMA 18: $\mathbf{P}_{\overline{\theta}}(\tau^i(\blacktriangle) = +\infty) > 0$.

PROOF: Recall that (p^i_n) is a submartingale under $\mathbf{P}_{\overline{\theta}}$, bounded by 1. Therefore, the probability that $p^i_n \le \pi_*$ for some $n \in \mathbf{N}$ is at most $(1-p_0)/(1-\pi_*)$. The first part of P3 implies that $\mathbf{P}_{\overline{\theta}}(\tau^i(\blacktriangle) = +\infty) > 0$: given $\overline{\theta}$, there is positive probability that no player will ever stop. *Q.E.D.*

LEMMA 19: *Let ϕ be an equilibrium. Then* $\lim_{n\to\infty} \pi^i_n(\blacktriangle) = 0$ *for some player i.*

PROOF: Assume that the sequence $\pi^2_n(\blacktriangle)$ does not converge to zero, for otherwise the conclusion already holds.

If player 2 is still active, player 1 has more opportunities than in the one-player problem. Hence, if player 1 drops out when player 2 is still active, she will a fortiori drop out when alone. Therefore, assuming $p^1_n = \pi^1_n(\blacktriangle)$,

$$\frac{q^1_n}{1-q^1_n} = \frac{\mathbf{P}_{\overline{\theta}}(\tau^2(\blacktriangle) \ge n)}{\mathbf{P}_{\underline{\theta}}(\tau^2(\blacktriangle) \ge n)} \times \frac{\pi^1_n(\blacktriangle)}{1-\pi^1_n(\blacktriangle)} \le \frac{\pi_*}{1-\pi_*}.$$

Since $\mathbf{P}_{\overline{\theta}}(\tau^2(\blacktriangle) \ge n) > 0$, to show that $\lim_{n\to\infty} \pi^1_n(\blacktriangle) = 0$ it is then sufficient to show that $\lim_{n\to\infty} \mathbf{P}_{\underline{\theta}}(\tau^2(\blacktriangle) \ge n) = 0$. But this holds since under $\mathbf{P}_{\underline{\theta}}$ the private beliefs $(p^2_n)_n$ form a supermartingale that converges to 0 a.s. *Q.E.D.*

D.3. *Proof of P1*

Let ϕ^i be the unique best reply to ϕ^j, with cutoffs (π^i_n). We will prove that the sequence $(\pi^i_n(\blacktriangle))$ is nonincreasing.

The formal proof involves a long list of inequalities. We provide a detailed sketch, which can be easily transformed into a formal proof. We will prove that player i's optimal continuation payoff (OCP for short) is lower in situation (A) than in situation (E) below:

(A) $p^i_n = p$ and $t^j_n = \blacktriangle$.
(E) $p^i_{n+1} = p$ and $t^j_{n+1} = \blacktriangle$.
This will show that $w^i_n(p, \blacktriangle) \leq w^i_{n+1}(p, \blacktriangle)$ for every p, and the result follows.
We proceed by introducing several situations player i may face, including fictitious ones:

(B) $p^i_n = p$, $t^j_n = \blacktriangle$, and there is an interim stage $n - \frac{1}{2}$, between stages $n-1$ and n, in which *only* player i receives a payoff (but the players make no choices). This situation is purely fictitious.

(C) $p^i_{n+1} = p$, $t^j_n = \blacktriangle$, and, starting from stage n, player i observes the status of player j with a one-stage delay. This situation involves a modified game.

(D) $p^i_{n+1} = p$ and $t^j_n = \blacktriangle$. This is the situation in which player i reaches stage $n+1$ with a private belief p, but has not yet figured out whether player j chose to remain active or not at stage n.

We compare these situations, from the viewpoint of the optimization problem faced by player i.

STEP 1: *Variations* (A) *and* (B). All relevant information contained in past payoffs is summarized in the private belief: it is irrelevant that in (A) and (B) different payoffs and a different number of payoffs lead to the same private belief. Besides, player j receives the same number of observations in both cases, so that the conditional distribution of p^j_n is the same in both cases. Since from stage n on, the two situations coincide, player i's OCP at stage n in both situations is the same.

STEP 2: *Variations* (B) *and* (C). The continuation games faced by player i in situations (B) and (C) are strategically equivalent. Therefore, player i's OCP is the same in both situations.

STEP 3: *Variations* (C) *and* (D). The only difference between the continuation games from stage $n+1$ in the two situations is that in (C), information is delayed for player i. Hence in (C), player i has fewer strategies, so that player i's OCP in (C), does not exceed her expected OCP in (D).

STEP 4: *Variations* (D) *and* (E). In (D), player i has not yet observed the choice made by player j at stage n. Once player i observes player j's choice, two cases may arise: either player j remained active and we reach (E), or she chose to drop out, so that we reach yet another situation:

(F) $p^i_{n+1} = p$ and $t^j_{n+1} = n$.

As a result, player i's expected OCP in situation (D) is a weighted average of her OCP's in situations (E) and (F).

Therefore, to prove that the OCP in situation (D) is at most her OCP in situation (E), it is sufficient to prove that the OCP in situation (F) does not exceed that in situation (E). This holds since (i) player i has more strategies in (E) than in (D) and (ii) by Lemma 17, her posterior belief that the state is $\overline{\theta}$ is higher in (E) than in (D). *Q.E.D.*

REFERENCES

AUMANN, R. J. (1964): "Mixed and Behavior Strategies in Infinite Extensive Games," in *Advances in Game Theory*, Annals of Mathematics Study, Vol. 52, ed. by M. Dresher, L. S. Shapley, and A. W. Tucker. Princeton, NJ: Princeton University Press, 627–650. [1594]

BERGEMANN, D., AND J. VÄLIMÄKI (1997): "Market Diffusion with Two-Sided Learning," *RAND Journal of Economics*, 28, 773–795. [1591]

——— (2000): "Experimentation in Markets," *Review of Economic Studies*, 67, 213–234. [1591]

BOLTON, P., AND C. HARRIS (1999): "Strategic Experimentation," *Econometrica*, 67, 349–374. [1591,1598]

CAPLIN, A., AND J. LEAHY (1994): "Business as Usual, Market Crashes and Wisdom after the Fact," *American Economic Review*, 84, 547–564. [1598]

CHAMLEY, C. (2004): "Delays and Equilibria with Large and Small Information in Social Learning," *European Economic Review*, 48, 477–501. [1592]

CHAMLEY, C., AND D. GALE (1994): "Information Revelation and Strategic Delay in a Model of Investment," *Econometrica*, 62, 1065–1085. [1592]

CHOW, Y. S., AND H. ROBBINS (1963): "On Optimal Stopping Rules," *Zeitschrift Warscheinlichkeitstheorie und Verwandte Gebiete*, 2, 33–49. [1596,1602]

DÉCAMPS, J. P., AND T. MARIOTTI (2004): "Investment Timing and Learning Externalities," *Journal of Economic Theory*, 118, 80–102. [1591]

DIXIT, A. K., AND R. S. PINDYCK (1994): *Investment under Uncertainty*. Princeton, NJ: Princeton University Press. [1592]

FERGUSON, T. (2004): "Optimal Stopping and Applications," available at http://www.math.ucla.edu/~tom/Stopping/Contents.html. [1596,1602]

GIRALDEAU, L.-A., T. J. VALONE, AND J. J. TEMPLETON (2002): "Potential Disadvantages of Using Socially Acquired Information," *Philosophical Transactions of the Royal Society of London, Ser. B*, 357, 1559–1566. [1592]

GLICKSBERG, I. L. (1952): "A Further Generalization of the Kakutani Fixed Point Theorem, with Application to Nash Equilibrium Points," *Proceedings of the American Mathematical Society*, 3, 170–174. [1607]

KELLER, G., S. RADY, AND M. CRIPPS (2005): "Strategic Experimentation with Exponential Bandits," *Econometrica*, 73, 39–68. [1591]

MILGROM, P. (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12, 380–391. [1596]

NEVEU, J. (1972): *Martingales à Temps Discret*, Paris: Masson. [1602,1603]

ROSENBERG, D., E. SOLAN, AND N. VIEILLE (2004): "Social Learning in One-Arm Bandit Problems," Discussion Paper 1396, The Center for Mathematical Studies in Economics and Management Science, Northwestern University. [1597,1600]

SHIRYAEV, A. (1996): *Probability* (Second Ed.). New York: Springer-Verlag. [1601]

VALONE, T. J., AND J. J. TEMPLETON (2002): "Public Information for the Assessment of Quality: A Widespread Social Phenomenon," *Philosophical Transactions of the Royal Society of London, Ser. B*, 357, 1549–1557. [1592]