

A PARTIAL FOLK THEOREM FOR GAMES WITH UNKNOWN PAYOFF
DISTRIBUTIONS

BY THOMAS WISEMAN¹

Repeated games with unknown payoff distributions are analogous to a single decision maker's "multi-armed bandit" problem. Each state of the world corresponds to a different payoff matrix of a stage game. When monitoring is perfect, information about the state is public, and players are sufficiently patient, the following result holds: For any function that maps each state to a payoff vector that is feasible and individually rational in that state, there is a sequential equilibrium in which players experiment to learn the realized state and achieve a payoff close to the one specified for that state.

KEYWORDS: Repeated games, learning, folk theorem, payoff uncertainty.

1. INTRODUCTION

THIS PAPER PRESENTS a model of repeated games with unknown payoff distributions. Given the vector of players' actions, payoffs in the stage game are random variables. Their joint distributions depend on the state of nature, which is chosen at the beginning of the game and is not observed by the players. The players can learn the state of the world (and the payoff matrix) only through coordinated experimentation, by trying different action profiles, observing the resulting payoff realizations, and updating their beliefs about the state. The vectors of actions and realized payoffs are assumed to be publicly observable, so all information about the state is public. This paper presents a partial folk theorem for a class of such games — when players are patient enough, there is a sequential equilibrium in which they learn the state of nature and achieve a payoff close to any feasible, strictly individually rational payoff specified for that state. That result is only a partial folk theorem because it does not address whether or not players can achieve a wider range of *ex ante* expected payoffs in equilibria with incomplete learning.

Games with unknown payoff distributions are analogous to the “multi-armed bandit” problem (described in Rothschild (1974)), in which a single agent must choose repeatedly from a finite set of actions, or “arms.” For each arm, there is a separate, unknown probability distribution over payoffs. The agent's prior beliefs about the payoff distributions induce subjective payoff expectations for each arm, but the arm with the highest subjective expected payoff may not be the best one to choose. The agent may want to sacrifice expected payoff in the short run to gain information about the realized payoff distributions that will help him in the long run. With multiple agents, the situation

is more complicated. An “arm” now corresponds to a vector of action choices. Experimentation has to be coordinated to be effective, since each player must play his specified action. Strategic considerations may also interfere with learning. For example, if a player has a very low equilibrium payoff in one state of the world, it may be in her interest to block experimentation by refusing to play certain actions. She may prefer ignorance to the risk of discovering that the low-payoff state has been realized.

As an example, suppose that two partners must decide repeatedly whether or not to invest in a joint project (whether to “work” or to “shirk”). In each period, a player who works incurs a cost c ; shirking is costless. If both players work, they each receive a random payoff x . If either player shirks, they get nothing. It is efficient for the players to invest only if the mean value of the random payoff x is greater than c , but the only way to find out the mean value is to experiment by playing $(work, work)$.

A second example is the case of duopolists introducing a new product. Every month each firm sets its level of output; the market price that month is a function of both firms’ output, the state of demand, and a random monthly fluctuation. The firms do not know how the product will be received. That is, they do not know the demand curve, although they have beliefs about its possible values. In choosing their output levels, the firms must balance three, possibly conflicting incentives: short-term profit maximization, strategic considerations, and the long-run value of information about the state of demand.

The aim of this paper is to present a partial folk theorem for such games with unknown payoff distributions. Each state of the world corresponds to a payoff matrix whose entries are the expected values of the realized payoff distributions. Each such matrix has its own set of feasible payoffs and its own set of individually rational payoffs

(i.e., where all players get at least their minmax payoff). In this context, the theorem implies that for any function that maps each state of the world to a payoff vector that is feasible and strictly individually rational in that state, there is a sequential equilibrium in which players experiment to learn the realized state and achieve a payoff close (in expectation) to that state's specified payoff. I prove that such a folk theorem holds for games where monitoring is perfect, all information about the state of the world is public, and players are patient.

Such a result is both more and less than a folk theorem. It is more in the sense that it guarantees equilibria that not only produce desired payoffs but also result in learning, so equilibrium payoffs can depend on the state even if the feasible set does not. Even for games where every action has positive expected learning, that learning result is not immediate. For example, players could learn about some entries in the payoff matrix but not others, or learning may occur very gradually. On the other hand, the result is less than a folk theorem in the sense that some games may have equilibria with incomplete learning that yield *ex ante* (that is, before the state is realized) expected payoffs outside the set attainable in learning equilibria and that may even Pareto dominate that set. In an equilibrium without learning, players may receive payoffs above their *expected* minmax payoffs, but below their *actual* minmax payoffs in the realized state – a potentially harsh punishment can deter deviations. I do not explore that possibility here. An additional limitation is that the result shows only that payoffs are achieved approximately (although with arbitrary precision).

In the folk theorem equilibrium, players experiment in the early rounds of the game to learn the payoffs, and then play an equilibrium of the revealed game. One

difficulty is that when payoffs are unknown it may be hard to punish a player for deviating from his equilibrium strategy. The problem is that when the state of the world is uncertain, the set of feasible (expected) payoffs is not necessarily the convex combination of the feasible sets in the different states. Instead, it is a (possibly strict) subset of that convex combination: Action profiles whose payoffs differ from each other in each state may have the same average payoff across states. For example, if action A yields payoff 1 in state i and -1 in state j , and action B gives the opposite, then when beliefs are fifty-fifty both actions produce zero in expectation. However, it turns out that it is sufficient either to punish a deviating player with a low payoff immediately or to learn more about the payoffs and then punish him. It further turns out that a player cannot simultaneously guarantee himself a high payoff and block learning, so effective punishment strategies are possible. In fact, there exists a profile of actions for the other players such that any response either results in learning or yields the punishee a payoff no higher than his lowest minmax payoff across all possible states of the world.

The main tool used in the proof is the notion of the self-generating set of payoffs developed in Abreu, Pearce, and Stacchetti (1986, 1990) and Fudenberg, Levine, and Maskin (1994). For games with unknown payoff distributions, there may be a different set of equilibrium payoffs for each different set of beliefs about the state of the world, so the appropriate concept is self-generating sets of payoff/belief pairs. I construct such a set C , and show that for each pair (v, b) in C , the payoff v is achievable as the outcome of a sequential equilibrium, starting from initial beliefs b . Payoff/belief pairs in the interior of C are achieved using action profiles that maximize the expected squared distance between current beliefs and next period's updated beliefs. In equilibrium, players choose

the maximal learning actions until they reach near-certainty about the state, at the boundaries of C . The set C is constructed so that when beliefs assign a high probability to a state, then the associated payoffs are close to that state's specified payoff.

There are several strands of previous work on learning payoffs in games. Aoyagi (1998) and Bolton and Harris (1999) consider extensions of the multi-armed bandit. Sequential choice ("herding") is studied by Banerjee (1992), Bikhchandani, Hirshleifer, and Welch (1992), and Smith and Sorensen (2000). There is also an extensive literature on repeated games with incomplete information, where strategic interaction has an important role. (See Aumann and Hart (1992).) Kalai and Lehrer (1995) consider repeated games where players are uncertain both about payoffs and about the strategies of their opponents, and prove that play converges to a subjective equilibrium. When players know only their own payoffs, Kalai and Lehrer (1993) show that players eventually play a Nash equilibrium of the full information game. Gossner and Vielle (2003) examine a model similar to the one presented here, except that payoffs are non-stochastic given the state of the world, so that playing an action profile once suffices to learn its payoff with certainty. They derive a folk theorem with full learning for the case of zero discounting.

The structure of the paper is as follows: In Section 2, I formally model games with unknown payoff distributions and present the theorem, and in Section 3 I conclude.

2. GAMES WITH UNKNOWN PAYOFF DISTRIBUTIONS

Model

Define a game \mathbf{G} as follows: There is a finite set \mathbf{W} of possible states of the world; an element of \mathbf{W} is denoted by \mathbf{w} . Let $K \equiv \#(\mathbf{W})$. Each state of the world corresponds to a different stage game $G(\mathbf{w})$. Each $G(\mathbf{w})$ has the same set of expected-utility maximizing players $N = \{1, 2, \dots, N\}$, and the same finite set of actions A_i for each $i \in N$. The set of action profiles is $A = A_1 \times A_2 \times \dots \times A_N$. The set of mixed action profiles is \mathbf{DA} , with element \mathbf{a} . The stage games differ across states only with respect to payoffs, and payoffs depend only on the state and the action profile a . For each $G(\mathbf{w})$, the payoff $u(a, \mathbf{w})$ resulting from action profile a is a random vector drawn from a cumulative distribution function $F(a, \mathbf{w})$ that varies with the state of the world; an independent draw is made from $F(a, \mathbf{w})$ each time a is played. The mean of $u(a, \mathbf{w})$ under $F(a, \mathbf{w})$ is $U(a, \mathbf{w})$. In an abuse of notation, define the expected payoff from mixed action \mathbf{a} in state \mathbf{w} , $U(\mathbf{a}, \mathbf{w})$, as $\sum_{a \in A} \mathbf{p}(a | \mathbf{a}) U(a, \mathbf{w})$, where $\mathbf{p}(a | \mathbf{a})$ is the probability that pure action profile a is the outcome of mixed action \mathbf{a} .

No two states yield the same payoff distribution functions for every action profile, so learning all of the payoff distributions is sufficient to learn the state. No other assumptions are made about the relationships among the payoff distributions $F(a, \mathbf{w})$, either across players or across action profiles. Both action profiles a (that is, the outcome of any randomization in a mixed action) and realized payoffs are observed by all players.

Play proceeds in this way: First, the state of the world is chosen once and for all according to the probability distribution \mathbf{F} , defined on \mathbf{W} . The distribution \mathbf{F} is common knowledge among the players, and it assigns strictly positive probability to each state of

the world. The realized state of the world determines the stage game $G(\mathbf{w})$. The players, who share a common discount factor \mathbf{d} , then repeat $G(\mathbf{w})$ for an infinite number of periods. Their payoffs from \mathbf{G} are the discounted sums of their payoffs in each round, multiplied by $(1 - \mathbf{d})$ to make them comparable to the stage game payoffs. I assume that a public randomization device is available to the players.

A *public history* H^t contains the action profiles chosen and payoffs realized in periods 1 through $t - 1$; H^1 is the null set. In period t , players have *belief* $B^t(H^t)$. The value $B^t(H^t)(\mathbf{w})$ is the subjective probability that the state is \mathbf{w} after history H^t is observed. Initially, players' beliefs are the prior probabilities, F . After each period, players observe actions and payoffs and update their beliefs according to Bayes' rule. All the available information about the state of the world is contained in the public histories, so players share the same beliefs in every period. For each action profile a and beliefs B , let $\Pr(B\mathcal{C} | a, B)$ be the probability that starting from beliefs B the payoffs realized from a will result in beliefs being updated to $B\mathcal{C}$ (The expectation of $B\mathcal{C}$ given B is always B). Let $EU(\mathbf{a}, B) \equiv \sum_{\mathbf{w} \in \Omega} B(\mathbf{w})U(\mathbf{a}, \mathbf{w})$ be the expected payoff from action \mathbf{a} given beliefs B .

For each $i \in N$ and $\mathbf{w} \in \mathbf{W}$, let $m^i(\mathbf{w}) \in \mathbf{DA}$ be the (possibly mixed) action profile that minmaxes Player i when the state of the world is \mathbf{w} . Let $e_i(\mathbf{w})$ denote Player i 's minmax payoff in state \mathbf{w} . That is, $e_i(\mathbf{w}) \equiv \min_{\mathbf{a}_{-i} \in \Delta A_{-i}} \max_{a_i \in A_i} U_i((a_i, \mathbf{a}_{-i}), \mathbf{w})$.

For each $\mathbf{w} \in \mathbf{W}$, let $V(\mathbf{w}) = \{U(a, \mathbf{w}) \in \mathbf{R}^N : a \in A\}$. $V^*(\mathbf{w})$, defined as the convex hull of $V(\mathbf{w})$, is the set of feasible payoffs in state \mathbf{w} . Let $V^{**}(\mathbf{w}) \equiv \{u \in V^*(\mathbf{w}) : u_i > e_i(\mathbf{w}) \text{ for all } i \in N\}$ be the set of feasible payoffs that are strictly individually rational.

Throughout, it will be assumed that $V^{**}(\mathbf{w})$ has dimension N for all $\mathbf{w} \in \mathbf{W}$. (This is the multiple-state extension of the standard full-dimensionality condition for folk theorems.)

Folk Theorem

The solution concept used here is the sequential equilibrium. Following Abreu, Pearce, and Stacchetti (1986, 1990) and Fudenberg, Levine, and Maskin (1994) (from now on, FLM), let $W_B \subseteq \mathbf{R}^N$ be the set of *continuation payoffs* when beliefs are B . The set $W \subseteq \mathbf{R}^N \times \Delta_K$ is defined as the set of all payoff/belief pairs (v, B) such that $B \in \Delta_K$ and $v \in W_B$. A function $w : A \times \Delta_K \rightarrow \mathbf{R}^N$ such that $w(a, B) \in W_B$ is a *continuation payoff function*. It gives continuation payoffs as a function of the current period's action profile and the next period's beliefs (which are updated from current beliefs after payoffs are realized). The *expected continuation payoff* given current beliefs B , $Ew : A \times \Delta_K \rightarrow \mathbf{R}^N$, gives the expected value of the continuation payoff as a function of the action profile and current beliefs. $Ew(a, B)$ is equal to $\int_{B' \in \Delta_K} \Pr(B' | a, B) w(a, B')$. Given a discount factor \mathbf{d} , then, a payoff/belief pair (v, B) is *generated* by W if there exists an action profile $\mathbf{a} \in \mathbf{DA}$ and a continuation payoff function w such that, for all i ,

$$(1) \quad v_i = (1 - \mathbf{d})EU_i((a_i, \mathbf{a}_{-i}), B) + \mathbf{d} \sum_{a \in A} \mathbf{p}(a | (a_i, \mathbf{a}_{-i}))Ew_i(a, B)$$

for all $a_i \in A_i$ such that \mathbf{a} assigns non-zero probability to a_i , and

$$v_i \geq (1 - \mathbf{d})EU_i((a_i, \mathbf{a}_{-i}), B) + \mathbf{d} \sum_{a \in A} p(a | (a_i, \mathbf{a}_{-i})) Ew_i(a, B)$$

for all $a_i \in A_i$.

The first part of Condition 1 means that the combination of this period's payoff from playing \mathbf{a} with the expected continuation payoff yields payoff v . The second part guarantees that all players will choose to play \mathbf{a} . If Condition 1 holds, the action profile \mathbf{a} is said to be *enforceable* by the function w . Call the set of payoff/belief pairs that are generated by W when \mathbf{d} is the discount factor $G(W, \mathbf{d})$. If $W \subseteq G(W, \mathbf{d})$ for some \mathbf{d} , then W is *self-generating* with respect to \mathbf{d} . When the state of the world is known, FLM show that every payoff vector in a self-generating set is the result of a sequential equilibrium. In the same way, if a payoff/belief pair (v, B) is a member of a self-generating set, then v is the expected payoff of a sequential equilibrium when initial beliefs are given by B .

Beliefs B can be written as b , a K -dimensional column vector whose k -th entry is $B(w_k)$. Define $\text{lng}(a, b)$, the *expected learning* of action a given beliefs b , as $\text{lng}(a, b) \equiv \int_{b' \in \Delta_K} \Pr(b' | a, b) (b' - b)^\top (b' - b)$, the expected squared distance between current and updated beliefs. The expected learning of a mixed action is given, straightforwardly, by the weighted average of the expected learnings of the pure actions in its support. Define $A^*(b) \subseteq A$ to be the action profiles that maximize expected learning:

$$A^*(b) = \arg \max_{a \in A} \{\text{lng}(a, b)\}.$$

Maximized learning is strictly positive for any belief b that assigns positive probability to more than one state of the world.

Let $\mathbf{w}(i, b)$ be the state (or one of the states) of the world assigned positive probability by belief b in which Player i 's minmax payoff $e_i(\mathbf{w})$ is lowest. That is,

$$\mathbf{w}(i, b) \in \arg \min_{\mathbf{w} \in \Omega B(\mathbf{w}) > 0} \{e_i(\mathbf{w})\}.$$

The action profile that minmaxes Player i in state $\mathbf{w}(i, b)$ is $m^i(\mathbf{w}(i, b))$. Let $\underline{L}_i(b)$ be defined as

$$\underline{L}_i(b) \equiv \min_{a_i \in A_i} \lg [(a_i, m^i(\mathbf{w}(i, b))), b],$$

and define $\underline{A}_i(b)$ to be

$$\underline{A}_i(b) \equiv \arg \max_{a_i \in A_i} \{EU_i[(a_i, m^i(\mathbf{w}(i, b))), b] : \lg [(a_i, m^i(\mathbf{w}(i, b))), b] = \underline{L}_i(b)\}.$$

That is, $\underline{L}_i(b)$ is the minimal expected learning that Player i can achieve when the other players are playing $m^i(\mathbf{w}(i, b))$ and beliefs are given by b . Among the set of actions that yield learning $\underline{L}_i(b)$, the actions in $\underline{A}_i(b)$ give Player i the highest expected payoff, again given beliefs b . Those definitions will be used in the proof of the folk theorem for supporting boundary points of the self-generating set.

With that background, it is possible to state the folk theorem:

PROPOSITION 1 (Folk Theorem for Games with Unknown Payoff Distributions):

Let $\mathbf{e} > 0$ and payoffs $v^*(\mathbf{w}_1) \in \text{int}(V^{**}(\mathbf{w}_1)), \dots, v^*(\mathbf{w}_K) \in \text{int}(V^{**}(\mathbf{w}_K))$ be given, and let F be a prior belief that assigns strictly positive probability to each state. Then there exists $\underline{\mathbf{d}}(F) < 1$ such that for all $\mathbf{d} > \underline{\mathbf{d}}(F)$, there is a sequential equilibrium E^* that satisfies the following condition: When the realized state is \mathbf{w} , then the expected payoff vector resulting from E^* is within \mathbf{e} of $v^*(\mathbf{w})$, and with probability at least $1 - \mathbf{e} \max\{t : B^t(\mathbf{w}) < 1 - \mathbf{e}\} < \infty$.

PROOF OF PROPOSITION 1: First, I will construct the set C of payoff/belief pairs, which will be shown to be self-generating. For each $k \in \{1, \dots, K\}$, let $b^*(k)$ be the belief that assigns probability 1 to state \mathbf{w}_k . Let d_k be the distance from $v^*(\mathbf{w}_k)$ to the boundary of $V^{**}(\mathbf{w}_k)$, and let $r = \min\{\mathbf{e}/2, d_1/2, \dots, d_K/2\}$. Define $C_{b^*(k)} \subseteq \mathbf{R}^N \times \Delta_K$ as the closed N -ball of radius r centered at $(v^*(\mathbf{w}_k), b^*(k))$ and lying in the hyperplane defined by $b = b^*(k)$:

$$C_{b^*(k)} = \{(v, b) \in \mathbf{R}^N \times \Delta_K : \|v - v^*(\mathbf{w}_k)\| \leq r \text{ and } b = b^*(k)\}.$$

The ball $C_{b^*(k)}$ lies in the interior of $V^{**}(\mathbf{w}_k) \times \{b = b^*(k)\}$. Let $\underline{\Delta}_K$ be the set of beliefs where the state is uncertain: $\underline{\Delta}_K \equiv \{\Delta_K \setminus \bigcup_{k=1}^K b^*(k)\}$. For each $b^\circ \in \underline{\Delta}_K$, let

$$v^*(b^\circ) = \sum_{k=1}^K b_k^0 v^*(\mathbf{w}_k).$$

Let $r_1, r_2 > 0$ be scalars to be chosen below, and let

$$r(b^\circ) = r_1 + r_2(b^\circ - (1/K)\mathbf{1})^\top(b^\circ - (1/K)\mathbf{1}),$$

where $\mathbf{1}$ is the K -dimensional column vector of 1's. Let C_{b° be the closed N -ball of radius $r(b^\circ)$ centered at $(v^*(b^\circ), b^\circ)$ and lying in the hyperplane defined by $b = b^\circ$. Define the set C as the union of the sections C_{b° , so that

$$C = \bigcup_{b \in \Delta_K} C_b.$$

Define C^E , the “ends” of C , as $C^E = \bigcup_{k=1}^K C_{b^{*(k)}}$, and let the complement of C^E in C be $\bar{C} = \bigcup_{\Delta_K} C_b$. Let $\bar{C}_{\text{int}} = \bigcup_{\Delta_K} \text{int}(C_b)$. The radii of the cross-sections of C increase quadratically from a minimum of r_1 at $b = ((1/K), \dots, (1/K))$, where each state is equally likely, to a maximum of $r_1 + r_2[(K-1)/K]$ at each $b^{*(k)}$, where the state is known with certainty. When the number of states K is 2, for example, the cross-section of C is shaped like an hourglass. (See Figure 1.) The quadratically bowed-in sides of the set C make it possible to support points on its boundary, as will be shown in Lemmas 2 and 4.

Choose the scalars r_1 and r_2 to satisfy the following conditions:

$$(2) \quad r_1 + r_2[(K-1)/K] = r.$$

- (3) For every (v, b) in \bar{C} , there exists a simplex with endpoints $p_1(v, b) \in \text{int}(C_{b^{*(1)}})$, $\dots, p_K(v, b) \in \text{int}(C_{b^{*(K)}})$ that contains (v, b) and whose interior lies in the interior of C , except for payoff/belief pair (v, b) itself, if (v, b) is on the boundary of \bar{C} .

Such scalars are guaranteed to exist, because as r_1 approaches r and r_2 approaches 0, the set C approaches convexity.²

Several lemmas will be used to prove Proposition 1. The first demonstrates that when the state is known with certainty, FLM's result can be extended to show that the ends C^E of the set C are locally self-generating. The set C is *locally self-generating* if for every payoff/belief pair $(v, b) \in C$, there is a $\mathbf{d} < 1$ and an open set O that includes (v, b) such that $O \cap C \subseteq G(C, \mathbf{d})$. It will be shown in Lemma 5 that to prove that C is self-generating, it is sufficient to demonstrate that it is locally self-generating.

LEMMA 1: *The set C is locally self-generating at every point in C^E .*

The second lemma demonstrates that all payoff/belief pairs in the interior of C , as well as points on the boundary of each C_b that do not minimize some player's payoff, are locally self-generating using actions that yield the greatest possible expected learning. FLM generate boundary payoffs using an action \mathbf{a} such that $U(\mathbf{a})$ lies outside of the self-generating set. That technique does not work here. Because the set of feasible payoffs may shrink when the state is unknown, there may not be a feasible payoff outside of C_b .

Instead, boundary points can be supported using expected continuation payoffs that lie outside C_b . Because the sides of C are bowed-in, a continuation payoff function can have an expected value outside of C_b , if it lies on a quadratic close to but inside the boundary of C , and players choose an action with positive expected learning. (See Figure 2.)

LEMMA 2: *For every point $(v, b) \in \bar{C}$ such that $v_i > \min\{v_i' : (v', b) \in C_b\}$ for all i , there is a $\mathbf{d} < 1$ (whose value increases as the maximum expected learning and the distance from (v, b) to the boundary of C_b decrease) and an open set O containing (v, b) such that $O \cap C$ is generated by the set $\tilde{C} \equiv \bar{C}_{\text{int}} \cup C^E$. Furthermore, the action profile used to generate $O \cap C$ maximizes expected learning.*

PROOF: Define the quadratic $Q(v^0, M, m) \subseteq \mathbf{R}^N \times \Delta_K$ as follows:

$$Q(v^0, M, m) \equiv \{(\hat{v}, \hat{b}) : \hat{v} = v^0 + (\hat{b} - b)^T M + (\hat{b} - b)^T (\hat{b} - b)m, \hat{b} \in \Delta_K\},$$

where v^0 and m are vectors in \mathbf{R}^N and M is a $K \times N$ matrix. Condition 3, together with the curved sides of the set C , guarantees that for each belief b there exists $\mathbf{e}(b) > 0$ such that the following is true: If (v^0, b) is in \bar{C} and $\|m\| < \mathbf{e}(b)$, then there exists a matrix M such that $Q(v^0, M, m)$ lies within \tilde{C} , except for possibly (v^0, b) . Let $a^* \in A^*(b)$ be an action profile that maximizes expected learning, and construct a continuation payoff function w such that $w(a^*, \hat{b})$ lies on $Q(v^0, M, m)$ for all updated beliefs \hat{b} . The expected value of $(\hat{b} - b)$ is zero, so the expected continuation payoff is

$$\begin{aligned}
Ew(a^*, b) &= \int_{\hat{b} \in \Delta_k} \Pr(\hat{b} | a^*, b) w(a^*, \hat{b}) \\
&= \int_{\hat{b} \in \Delta_k} \Pr(\hat{b} | a^*, b) [v^0 + (\hat{b} - b)^\top M + (\hat{b} - b)^\top (\hat{b} - b)m] \\
&= v^0 + \int_{\hat{b} \in \Delta_k} \Pr(\hat{b} | a^*, b) [(\hat{b} - b)^\top (\hat{b} - b)m] \\
&= v^0 + \text{lng}(a^*, b)m.
\end{aligned}$$

That is, each player's expected continuation payoff is equal to a constant plus a term proportional to the expected learning of the action profile a^* .

Let (v, b) be given. Choose v^0 , m , and M such that

$$(4) \quad v = (1 - \mathbf{d})EU(a^*, b) + \mathbf{d}[v^0 + \text{lng}(a^*, b)m].$$

As $\mathbf{d} \rightarrow 1$, v^0 close enough to v can be chosen so that (v^0, b) lies in the interior of C_b , $Q(v^0, M, m)$ lies within \tilde{C} , and Condition 4 is satisfied. (See Figure 3.) The smaller $\text{lng}(a^*, b)$ is, and the closer (v, b) is to the boundary of C_b , the higher is the necessary \mathbf{d} .

Construct the rest of the continuation payoff function w as follows: For each player i , choose a payoff vector v^i such that $(v^i, b) \in C_b$ and

$$(5) \quad (1 - \mathbf{d})EU_i((a_i, a_{-i}^*), b) + \mathbf{d}v_i^i < v_i \text{ for all } a_i \in A_i.$$

Since $v_i > \min\{v_i' : (v', b) \in C_b\}$, such vectors exist for large enough \mathbf{d} . Again, the closer (v, b) is to the boundary of C_b , the higher the necessary \mathbf{d} . For each action profile a^i that represents a unilateral deviation by player i from a^* , choose a matrix $M(v^i)$ such that $Q(v^i, M(v^i), 0)$ lies within \tilde{C} , and define $w(a^i, \hat{b}) \equiv v^i + (\hat{b} - b)^\top M(v^i)$. Note that $Ew(a^i, b) = v^i$. For all other action profiles a , let $w(a, \hat{b}) \equiv w(a^*, \hat{b})$. Conditions 4 and 5 imply that the continuation payoff function so defined enforces action profile a^* .

Thus, for high enough \mathbf{d} (v, b) is generated by \tilde{C} using any $a^* \in A^*(b)$. In the same fashion, the payoff/belief pair $(v + c, b)$ is generated by continuation payoffs lying along the quadratic $Q(v^0, M, m) + c/\mathbf{d}$. For c close enough to zero, $Q + c/\mathbf{d}$ lies within \tilde{C} . Similarly, $(v, b\epsilon)$ is generated (with maximal learning) by a quadratic near $Q(v^0, M, m)$ that also lies within \tilde{C} for $b\epsilon$ close enough to b : both expected utility and expected learning are continuous in beliefs, and $A^*(b)$ is upper hemicontinuous. *Q.E.D.*

Lemmas 1 and 2 show that nearly all of C is locally self-generating. All that remains are the points (v, b) that minimize some player i 's payoffs in C_b for some $b \in \underline{\Delta}_K$. Supporting such a point may require an expected continuation payoff strictly less than v_i . (That expected continuation payoff can be achieved using continuation payoffs that are quadratic in beliefs, as in the proof of Lemma 2.) But if player i chooses an action that leads to zero expected learning, then his continuation payoff must lie in C_b , and thus cannot be less than v_i , i 's lowest payoff in C_b . Lemma 3 establishes, however, that no player can at the same time reduce learning to zero and guarantee himself a payoff higher than his lowest minmax payoff across all possible states of the world.

LEMMA 3: For every player i and beliefs $b \in \underline{\Delta}_K$, there exists an action profile $\underline{\mathbf{a}}_{-i} \in \Delta A_{-i}$ for the other players such that for all $a_i \in A_i$, either

- (i) $\text{lng}[(a_i, \underline{\mathbf{a}}_{-i}), b] > 0$, or
- (ii) $EU_i[(a_i, \underline{\mathbf{a}}_{-i}), b] \leq e_i(\mathbf{w}(i, b))$.

PROOF: If a mixed action \mathbf{a} has zero expected learning given beliefs b , then players cannot learn about the state by observing realized payoffs after any pure action $a \in \text{supp}(\mathbf{a})$. That implies that the expected payoff to each such a , and thus the expected payoff to \mathbf{a} , is the same in every possible state; that is, in every state that beliefs b assign non-zero probability. Choose $\underline{\mathbf{a}}_{-i}$ to be $m^i(\mathbf{w}(i, b))$, the profile that minmaxes Player i in the state $\mathbf{w}(i, b)$ where Player i 's minmax payoff e_i is lowest. Then for any $a_i \in A_i$ such that $\text{lng}[(a_i, \underline{\mathbf{a}}_{-i}), b] = 0$, the expected payoff to Player i from action profile $(a_i, \underline{\mathbf{a}}_{-i})$ is

$$\begin{aligned}
 EU_i[(a_i, \underline{\mathbf{a}}_{-i}), b] &= \sum_{\mathbf{w} \in \Omega} B(\mathbf{w}) U_i[(a_i, \underline{\mathbf{a}}_{-i}), \mathbf{w}] \\
 &= U_i[(a_i, \underline{\mathbf{a}}_{-i}), \mathbf{w}(i, b)] \\
 &\leq e_i(\mathbf{w}(i, b)). \qquad \qquad \qquad Q.E.D.
 \end{aligned}$$

That lowest minmax payoff $e_i(\mathbf{w}(i, b))$ lies outside of section C_b , so points on the boundary of C that minimize players' payoffs can be supported using either a variant of FLM's technique (if learning is zero) or the quadratic technique of Lemma 2 (if learning is positive). Lemma 4 shows that C is locally self-generating at such points.

LEMMA 4: For all $b \in \underline{\Delta}_K$, the set C is locally self-generating at every point (v, b) on the boundary of C_b such that $v_i = \min\{v_i' : (v', b) \in C_b\}$ for some i .

Together, Lemmas 1, 2, and 4 show that the entire set C is locally self-generating. It remains only to demonstrate that local self-generation implies self-generation, and that the equilibrium results in sufficient learning. Lemma 5 addresses the first issue.

LEMMA 5: If C is locally self-generating, then there is a $\mathbf{d}\hat{c} < 1$ such that C is self-generating with respect to all $\mathbf{d} \in (\mathbf{d}\hat{c}, 1)$.

PROOF: The proof is identical to the first part of FLM's proof of Lemma 4.2, except in showing that if $(v, b) \in G(C, \mathbf{d}\hat{c})$, then it is in $G(C, \mathbf{d})$ for all $\mathbf{d} \in (\mathbf{d}\hat{c}, 1)$: Given $\mathbf{d}\hat{c}$ let \mathbf{a}^* be the action used to generate (v, b) , and let w^* be the continuation payoff function that enforces \mathbf{a}^* . Then define

$$w_{\mathbf{d}}(a, \hat{b}) \equiv \frac{\mathbf{d}'(1-\mathbf{d})}{\mathbf{d}(1-\mathbf{d}')} w(a, \hat{b}) + \frac{\mathbf{d}-\mathbf{d}'}{\mathbf{d}(1-\mathbf{d}')} [v + (\hat{b}-b)^{\top}M],$$

where the $K \times N$ matrix M is constructed so that $(v + (\hat{b}-b)^{\top}M, \hat{b}) \in C$ for all \hat{b} . (See the proof of Lemma 1.) Since each $C_{b\hat{c}}$ is convex, $(w_{\mathbf{d}}(a, \hat{b}), \hat{b}) \in C$ for all $\mathbf{d} \in (\mathbf{d}\hat{c}, 1)$, and it is easy to check that Condition 1 is met with $w_{\mathbf{d}}$, (v, b) , and \mathbf{a}^* . Q.E.D.

PROOF OF PROPOSITION 1, continued: Define \mathbf{e}^* as the greatest distance between the payoffs to any two actions in any two states of the world, and choose \mathbf{D} such that $\mathbf{D} < \min\{\mathbf{e}, \mathbf{e}/2\mathbf{e}^*\}$. Let the prior belief \underline{b} be given, and choose a positive integer T large enough that after T periods of playing the action profile leading to the greatest expected learning, the probability that updated beliefs will assign the true state a weight of $1 - \mathbf{D}$ or higher and will continue to do so forever is at least $1 - \mathbf{e}$. Choose $\mathbf{D}^* \in (0, \mathbf{D})$ such that if updated beliefs assign a state weight $1 - \mathbf{D}^*$ or higher, then they will continue to assign it a weight of at least $1 - \mathbf{D}$ forever with probability at least $1 - \mathbf{e}$. Define the set $C(\mathbf{e})$ as $C(\mathbf{e}) \equiv \{(v, b) \in C : \|v - v^*(b)\| < \mathbf{e}\}$, a “narrow column” running through the center of C . Let $C^*(\mathbf{e})$ be the subset of $C(\mathbf{e})$ where beliefs b assign all states a weight less than $1 - \mathbf{D}^*$.

Lemmas 1, 2, 4, and 5 show that the set C is self-generating for high enough \mathbf{d} . Lemma 2, furthermore, shows that for every point (v, b) in $\overline{C}_{\text{int}}$ there is a $\mathbf{d} < 1$ large enough that (v, b) is supportable using an action that leads to the greatest possible learning and continuation payoffs that themselves lie either in $\overline{C}_{\text{int}}$ or on the edge C^E , where the state is known with certainty. The necessary \mathbf{d} is decreasing in the distance from (v, b) to the boundary of C_b and in the value of maximal expected learning, which is lowest when the state is known with near-certainty. Also, the larger is \mathbf{d} , the closer the vector m in Lemma 2 can be to zero; that is, the closer the quadratic curve containing the continuation payoffs is to being linear. Therefore, it is possible to find a set $0 < \mathbf{e}_1 < \dots < \mathbf{e}_T$ and $\mathbf{d} < 1$ high enough that $(v^*(\underline{b}), \underline{b})$ can be supported with maximal learning using on-equilibrium continuation payoffs that lie in $C(\mathbf{e}_1)$, and any point in $C^*(\mathbf{e}_t)$ can be similarly supported with payoffs in $C(\mathbf{e}_{t+1})$ for $t \leq T - 1$.

Proposition 1 follows: Suppose without loss of generality that the realized state is w_1 . Choose the point $(v^*(\underline{b}), \underline{b})$ in the center of $C_{\underline{b}}$, and consider the following strategy, E^* : In the first period, play the action that supports $(v^*(\underline{b}), \underline{b})$. In the next period, after observing action a and updating beliefs to b , play the action that supports $(w(a, b), b)$, where w is the continuation payoff function used to support $(v^*(\underline{b}), \underline{b})$. Actions in subsequent periods are specified in the same fashion. By construction, no player will want to deviate, and E^* is a sequential equilibrium. When d is high enough, the first period action lies in $A^*(\underline{b})$, and $(w(a_1, b), b)$ lies in $C(e_1)$. If $(w(a_1, b), b) \in C^*(e_1)$, then period 2's action is in $A^*(b)$, and so on for the first T periods. Thus, either players will choose actions to maximize learning for the first T periods, or they will assign belief at least $1 - D^*$ to some state. Within T periods, then, beliefs will stay forever within D of $b^*(1)$ (certainty of state w_1) with probability at least $1 - e$. By construction of the set C , if a belief $b \in C$ is within D of $b^*(1)$, then every payoff $v \in C$ such that $(v \in C) \in C_{b \in C}$ is within $e/2 + Dbe^* < e$ of $v^*(w_1)$. Thus, for a high enough discount factor the equilibrium E^* satisfies Proposition 1. *Q.E.D.*

3. CONCLUSION

This paper presents a form of folk theorem for games with unknown payoff distributions when all information about the state of the world is public. But suppose, returning to the duopolists in the introduction, that each firm observes the market price and its own output, but not its competitor's output. In that case, a firm's private

information is relevant for making inferences about the demand curve, so Proposition 1 fails to hold. More generally, a player might receive a private signal that gives her information both about the state and about which signals the other players are likely to have received, so higher order beliefs become important. A similar folk theorem may hold for some classes of such games. It seems likely, for example, that in games of common interest, like the partnership game in the introduction, players can cooperate to find the efficient outcome even when signals are private. Proving such a result, however, will require an approach different from the dynamic programming techniques used in this paper, which rely on the game's recursive structure; that structure breaks down when players' beliefs diverge. Deriving such a folk theorem is a subject for further research.

Dept. of Economics, The University of Texas at Austin, Austin, TX 78712;
wiseman@eco.utexas.edu.

APPENDIX

PROOF OF LEMMA 1: Let $(v, b^*(k))$ be a point in C^E , where the state of the world is known with certainty. When the state is known and actions are observable, FLM's Conditions 6.2 and 6.3 hold. FLM's Theorem 6.2 then guarantees that for all $v \notin$ near enough to v , $(v \notin b^*(k))$ is generated by $C_{b^*(k)}$ for some $d < 1$, using continuation payoffs $w^*(a)$ that lie in $C_{b^*(k)}$. It is straightforward to modify the proof of Theorem 6.2 to ensure that the continuation payoffs $w^*(a)$ in fact lie in the interior of $C_{b^*(k)}$. For each a , choose a $K \times N$ matrix $M(a)$ such that the interior of the simplex

$$L(w^*(a), M(a)) \equiv \{(\hat{v}, \hat{b}) : \hat{v} = w^*(a) + (\hat{b} - b^*(k))^T M(a), \hat{b} \in \Delta_K\}$$

lies in the interior of C ; by Condition 3, such a construction is possible. Define a continuation payoff function w such that $w(a, \hat{b}) = w^*(a) + (\hat{b} - b^*(k))^T M(a)$ for all action profiles a and all updated beliefs \hat{b} . At belief $b^*(k)$, $w(a, \hat{b})$ reduces to $w^*(a)$, and thus supports $(v, b^*(k))$.

For beliefs $b \notin$ define a new continuation payoff function $w_{\epsilon}(a, \hat{b})$ as

$$w_{\epsilon}(a, \hat{b}) \equiv w(a, \hat{b}) + \frac{1-d}{d} [U(a, \mathbf{w}_k) - EU(a, b \notin)] + [Ew(a, b^*(k)) - Ew(a, b \notin)].$$

When belief $b \notin$ is close enough to $b^*(k)$, then the values of $w_{\epsilon}(a, \hat{b})$ also lie in C , and they generate the payoff/belief pair $(v, b \notin)$. *Q.E.D.*

As a preliminary to the proof of Lemma 4, Lemma 7 establishes that there are supportable points that lie outside of each C_b . Lemma 7 in turn makes use of the following lemma:

LEMMA 6: *Let P be an N -ball, and let p be a point on the boundary of P . Let H be the hyperplane tangent to P at p , and let $H + c$ be a translate of H that intersects P . Let $L(c)$ be the diameter of the $(N - 1)$ -ball formed by the intersection of $H + c$ with P , and let $d(c)$ be the maximum distance from $H + c$ to any point in P lying on the same side of $H + c$ as the point p . As $c \rightarrow \mathbf{0}$, the limit of the ratio $d(c)/L(c)$ is 0.*

PROOF OF LEMMA 6: Without loss of generality, let P be centered at $\mathbf{0}$ with radius 1, and let p be the point $(1, 0, \dots, 0)$. The tangent hyperplane H , then, is given by $x_1 = 1$. The translates of H that intersect P can be parametrized as $x_1 = 1 - z$, with $z \in [0, 2]$. For $z > -1$, $L(z) = 2[1 - (1 - z)^2]^{1/2}$, and $d(z) = z$. As $z \rightarrow 0$, the limit of $d(z)/L(z)$ is 0. *Q.E.D.*

LEMMA 7: *For all $b \in \underline{\Delta}_K$ and for each point $(v, b) \in C_b$ such that $v_i = \min\{v_i' : (v', b) \in C_b\}$ for some i , there exists a payoff/belief pair (v^d, b) satisfying $v_i^d < v_i$ such that there exists an open neighborhood of (v^d, b) that is generated by C for high enough d .*

PROOF OF LEMMA 7: As in Lemma 2, the desired payoff/belief pair (v, b) is supported using continuation payoff functions that lie along a quadratic and that have expectations outside C_b . There will be two cases, according to whether or not Player i

has a response to $m^i(\mathbf{w}(i, b))$, the profile that minmaxes Player i in the state $\mathbf{w}(i, b)$ where Player i 's minmax payoff e_i is lowest, that will result in no learning. That is, according to whether or not $\underline{L}_i(b)$, the minimal learning possible in response to $m^i(\mathbf{w}(i, b))$, exceeds zero. Choose $\underline{\mathbf{a}}_i \in \underline{\mathbf{D}}\underline{A}_i(b)$ to be an action for Player i that gives the highest payoff among those that minimize learning in response to $m^i(\mathbf{w}(i, b))$.

Consider expected continuation payoffs Ew that lie on the hyperplane H defined by $x_i = v_i$. (Note that (H, b) is tangent to C_b at (v, b)). Since lying on H places no restrictions on the values of Ew_j for any $j \neq i$, FLM's Lemma 4.3 shows that for all $\mathbf{d} > 0$, it is possible to enforce profile $[\underline{\mathbf{a}}_i, m^i(\mathbf{w}(i, b))]$ for all players except Player i with expected continuation payoffs Ew^c that lie along any translate $H + c$ of H and are no farther apart than $k(1 - \mathbf{d})/\mathbf{d}$, for some positive constant k . That is,

$$\max_{a, a' \in A} \{ \| Ew^c(a) - Ew^c(a') \| \} < k(1 - \mathbf{d})/\mathbf{d}.$$

Note that $Ew_i^c(a, b)$ must equal $v_i + c$ for all a , because each Ew^c lies on $x_i = v_i + c$. As in the proof of Lemma 1, such expected continuation payoffs result from a continuation payoff function w whose values lie on simplices in C . For each \mathbf{d} , let $H + c(\mathbf{d})$ be the translate of H nearest to v such that the values of $(Ew^{c(\mathbf{d})}, b)$ lie in C_b . As $\mathbf{d} \rightarrow 1$, the expected continuation payoffs can be brought closer and closer together, so $c(\mathbf{d})$ can shrink closer and closer to $\mathbf{0}$.

Now construct an alternative continuation payoff function \hat{w}^c as follows:
 $\hat{w}_j^c(a, \hat{b}) = w_j^c(a, b)$ if $j \neq i$, and $\hat{w}_i^c(a, \hat{b}) = w_i^c(a, b) - (\hat{b} - b)^\top (\hat{b} - b)m$, where $m > 0$ is

a scalar. As in the proof of Lemma 2, choose m small enough that each $\hat{w}^c(a, \hat{b})$ can lie in C whenever the values of (Ew^c, b) lie in C_b . The continuation payoff function \hat{w}^c also enforces the profile $[\underline{\mathbf{a}}_i, m^i(\mathbf{w}(i, b))]$ for all players except Player i , because it gives those players the same continuation payoffs as w^c . Given $\hat{w}_i^c(a, \hat{b})$, Player i 's expected continuation payoff from an action profile is strictly decreasing in the expected learning of the action profile, so for high enough \mathbf{d} Player i will choose the action $\underline{\mathbf{a}}_i$ that maximizes payoff among those that minimize expected learning.

The conclusion of the proof depends on the value of $\underline{L}_i(b)$:

Case 1: $\underline{L}_i(b) > 0$. In this case, even the minimal expected learning is greater than zero, so Player i 's optimal expected continuation payoff when faced with \hat{w}^c is strictly less than $Ew_i^c(a, b) = v_i + c$. As $\mathbf{d} \rightarrow 1$, $c(\mathbf{d})$ shrinks to $\mathbf{0}$. For large enough \mathbf{d} , then, the payoff/belief pair $(v^{c(\mathbf{d})}, b)$ thus supported satisfies $v^{c(\mathbf{d})} < v_i$.

Case 2: $\underline{L}_i(b) = 0$. In this case, Lemma 3 implies that $EU_i[(\underline{\mathbf{a}}_i, m^i(\mathbf{w}(i, b))), b] \leq e_i(\mathbf{w}(i, b))$. That is, there is a feasible payoff outside of C_b . For brevity, let \mathbf{q} denote $EU_i[(\underline{\mathbf{a}}_i, m^i(\mathbf{w}(i, b))), b]$. For any $\mathbf{d} > 0$, action $[\underline{\mathbf{a}}_i, m^i(\mathbf{w}(i, b))]$ and continuation payoffs $\hat{w}^{c(\mathbf{d})}$ generate the payoff/belief pair $(v^{c(\mathbf{d})}, b)$, where the i -th element of $v^{c(\mathbf{d})}$ is

$$v_i^{c(\mathbf{d})} = (1 - \mathbf{d})\mathbf{q} + \mathbf{d} \sum_{a \in A} \mathbf{p}(a | [\underline{\mathbf{a}}_i, m^i(\mathbf{w}(i, b))]) Ew_i^{c(\mathbf{d})}(a, b).$$

Let \mathbf{g} represent $\sum_{a \in A} \mathbf{p}(a | [\underline{\mathbf{a}}_i, m^i(\mathbf{w}(i, b))]) Ew_i^{c(\mathbf{d})}(a, b)$. Since $\mathbf{q} \leq e_i(\mathbf{w}(i, b))$ and

$Ew_i^{c(\mathbf{d})}(a, b) > e_i(\mathbf{w}(i, b))$ for all a (because $Ew^{c(\mathbf{d})}(a, b) \in C_b$), $v_i^{c(\mathbf{d})} < \mathbf{g}$. The distance

between the values $v_i^{c(\mathbf{d})}$ and \mathbf{g} is $(1 - \mathbf{d})|(\mathbf{q} - \mathbf{g})|$. As $\mathbf{d} \rightarrow 1$ and $c(\mathbf{d})$ approaches $\mathbf{0}$, \mathbf{g} converges to v . As $\mathbf{d} \rightarrow 1$, then, the ratio of the difference between $v_i^{c(\mathbf{d})}$ and \mathbf{g} to the length $(k(1 - \mathbf{d})/\mathbf{d})$ of $(H + c(\mathbf{d}), b) \cap C_b$ converges to the positive constant $|(\mathbf{q} - \mathbf{g})| / k$. On the other hand, Lemma 6 shows that as $\mathbf{d} \rightarrow 1$, the ratio of the distance between the point v and the hyperplane $H + c(\mathbf{d})$ to the length of $(H + c(\mathbf{d}), b) \cap C_b$ goes to zero. Thus, for large enough \mathbf{d} the distance between $v_i^{c(\mathbf{d})}$ and \mathbf{g} is greater than the distance between \mathbf{g} (which lies in $(H + c(\mathbf{d}), b) \cap C_b$) and v_i . Therefore, $v_i^{c(\mathbf{d})} < v_i$.

Thus, there exists a payoff/belief pair $(v_i^{c(\mathbf{d})}, b)$ generated by C for large enough \mathbf{d} such that $v_i^{c(\mathbf{d})} < v_i$. Points near enough to $(v_i^{c(\mathbf{d})}, b)$ can be generated using a similar construction. (Because the set $\underline{A}_i(b)$ may not be lower hemicontinuous, enforcing action \underline{a}_i for Player i , even when it no longer minimizes learning, may require adjusting his continuation payoffs in a continuous way as beliefs vary from b .) *Q.E.D.*

PROOF OF LEMMA 4:³ Pick any belief $b \in \underline{\Delta}_K$ and any point $(v, b) \in C_b$ such that $v_i = \min\{v_i' : (v', b) \in C_b\}$ for some i . Let $S(b)$ be the set of supportable payoff/belief pairs described in Lemma 7 that give Player i a payoff less than v_i . Note that given any N -ball B , any point x on the boundary of B , and any point y that is separated from B by the hyperplane tangent to B at x , the point x is a convex combination of y and some interior point of B . Each C_b is an N -ball, and $S(b)$ is separated from C_b by the hyperplane tangent to it at the boundary point (v, b) , so payoff/belief pair (v, b) is a convex combination of a point $(s, b) \in S(b)$ and a point $(v \notin b) \in \text{int}(C_b)$:

$$(v, b) = I(s, b) + (1 - I)(v \zeta b), I \in (0,1).$$

Lemmas 2 and 7 guarantee that there exist open sets around the points (s, b) and $(v \zeta b)$ that are generated by C for some $\mathbf{d} < 1$. Therefore, there is an open neighborhood O around (v, b) such that $O \cap C_b$ is generated by C as well. To generate the point (v, b) , for example, players with probability I use the action profile and continuation payoff function that generates (s, b) , and with probability $(1 - I)$ they use the action profile and continuation payoff function that generates $(v \zeta b)$. *Q.E.D.*

REFERENCES

- ABREU, D., D. PEARCE, AND E. STACCHETTI (1986): "Optimal Cartel Monitoring with Imperfect Information," *Journal of Economic Theory*, 39, 251-269.
- (1990): "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, 58, 1041-1063.
- AOYAGI, M. (1998): "Mutual Observability and the Convergence of Actions in a Multi-Person Two-Armed Bandit Model," *Journal of Economic Theory*, 82, 405-424.
- AUMANN, R. AND S. HART (1992): *Handbook of Game Theory with Economic Applications*. Volume 1. New York: Elsevier Science Publishers.
- BANERJEE, A. (1992): "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107, 797-817.
- BIKHCHANDANI, S., D. HIRSHLEIFER, AND I. WELCH (1992): "A Theory of Fads, Fashions, Custom, and Cultural Change as Information Cascades," *Journal of Political Economy*, 100, 992-1026.
- BOLTON, P., AND C. HARRIS (1999): "Strategic Experimentation," *Econometrica*, 67, 349-374.

FUDENBERG, D., D. LEVINE, AND E. MASKIN (1994): “The Folk Theorem with Imperfect Public Information,” *Econometrica*, 62, 997-1039.

GOSSNER, O. AND VIELLE, N. (2003): “Strategic Learning in Games with Symmetric Information,” *Games and Economic Behavior*, 42, 25-47.

KALAI, E. AND E. LEHRER (1993): “Rational Learning Leads to Nash Equilibrium,” *Econometrica*, 61, 1019-1045.

—— (1995): “Subjective Games and Equilibria,” *Games and Economic Behavior*, 8, 123-163.

ROTHSCHILD, M. (1974): “A Two-Armed Bandit Theory of Market Pricing,” *Journal of Economic Theory*, 9, 185-202.

SMITH, L. AND P. SORENSEN (2000): “Pathological Outcomes of Observational Learning,” *Econometrica*, 68, 371-398.

¹ Thanks to Jeffrey Ely and Juuso Valimaki for suggesting this line of research. I am especially grateful to Jeffrey Ely for many helpful ideas, suggestions, and criticisms. I also want to thank Eduardo Faingold, Thomas Geraghty, Ehud Kalai, George Mailath, Peter Meyer, Dale Stahl, Robert Vigfusson, Asher Wolinsky, and numerous seminar participants for useful comments. Finally, thanks to editor Glenn Ellison and three anonymous referees for their careful reading and detailed suggestions. All remaining errors are my responsibility.

² Note that the set C , although contained within a convex combination of feasible payoffs in different states, may include points (v, b) such that payoff v is not a feasible expected payoff when beliefs are b , as described in the introduction. Any equilibrium that achieves expected payoff v starting from beliefs b , therefore, must involve learning.

³ The proof of Lemma 4 is the only place where the assumption of a public randomization device is used.