Theoretical Economics 7 (2012), 217–239 1555-7561/20120217

A partial folk theorem for games with private learning

Thomas Wiseman
Department of Economics, University of Texas at Austin

The payoff matrix of a finite stage game is realized randomly and then the stage game is repeated infinitely. The distribution over states of the world (a state corresponds to a payoff matrix) is commonly known, but players do not observe nature's choice. Over time, they can learn the state in two ways. After each round, each player observes his own realized payoff (which may be stochastic, conditional on the state) and he observes a noisy public signal of the state (whose informativeness may vary with the actions chosen). Actions are perfectly observable. The result is that 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 patient players learn the realized state with arbitrary precision and achieve a payoff close to the one specified for that state. That result extends to the case where there is no public signal, but instead players receive very closely correlated private signals of the vector of realized payoffs.

Keywords. Repeated games, learning, folk theorem.
JEL classification. C73, D83.

1. Introduction

Consider a repeated stage game in which the sets of players and actions are known, but the distribution of payoffs conditional on action profiles is chosen randomly once and for all at the start of play, and the players do not observe nature's choice. The players have a common prior over the finite set of possible states of the world and they have two ways to learn (directly) the state over time. First, each player observes the profile of actions and his own realized payoff in each period: that payoff is a random variable whose distribution depends on the state. Second, every period there is a noisy public signal of the state. The public signal varies in informativeness across action profiles and stochastically incorporates the information in the privately observed payoffs. A player may also learn about the state indirectly, if the actions of other players reveal their history of private payoffs. This paper, extending the result of Wiseman (2005), establishes for such games a partial folk theorem of the following sort: for any function mapping

Thomas Wiseman: wiseman@austin.utexas.edu
I thank participants at the 2008 Workshop on Recent Advances in Repeated Games at the Stony Brook Game Theory Festival (especially Johannes Hörner, George Mailath, David Miller, Larry Samuelson, and Tristan Tomala) and the 2008 Conference on Strategic Information Acquisition and Transmission, funded by SFB TR 15, as well as seminar audiences and two anonymous referees, for helpful comments. This material is based on work supported by the National Science Foundation under Grant SES-0614654.

DOI: 10.3982/TE913

each state of the world to a payoff vector that is feasible and individually rational in that state, there exists a sequential equilibrium in which patient players achieve the payoff specified for whatever state is realized. (That folk theorem is "partial" because there may be equilibria without learning that yield ex ante—that is, unconditional on the state—expected payoffs outside the set derived here.) That result also holds in the case where there is no public signal, but players receive very closely correlated private signals of the vector of realized payoffs in each period (if there are at least three players and cheap talk is allowed).

One example of such a game is the situation of oligopolists introducing a product into a new market with uncertain demand. Every month, each seller sets a publicly observed price, and sees her own sales and a noisy public indicator of total sales in the market.[1] Another situation that fits the model is co-authors presenting a joint research paper. Each author sees which seminars and conferences the other is invited to, but does not observe the reaction of the audience. Opposing sides of a war-of-attrition-type strike, with asymmetric information about the costs of holding out, may read newspaper accounts of how their opponents are faring. The model also applies to wars between nations or to partners in a joint investment project who observe their own returns precisely and each others' returns only approximately.

These repeated games with unknown payoff distributions and private learning are in some ways analogous to repeated games with imperfect public monitoring. In the latter class of games, focusing on perfect public equilibria (PPE), in which players' strategies depend only on public information, is a fruitful approach to obtain folk theorems (as in Fudenberg et al. 1994). The key insight in establishing the existence of a PPE in their setting is that the set of best responses to a public strategy profile always includes a public strategy: when the other players' future actions depend only on public information, a player loses nothing by ignoring his private information. In the current context, however, a perfect public equilibrium may fail to exist; there may be no public strategy best response to a public strategy profile, for two reasons. First, a player's private information influences his expectation of future public signals and thus his expectations of the future play of his opponents. Second, the public history may call for him to play a sequence of actions that he knows from his private signals yields a payoff below his minmax payoff. For both reasons, it may be costly for a player to ignore his private information, and so any best response must be a private strategy.

More generally, the potential for private and public beliefs to diverge creates a difficulty in deriving a folk theorem of the type desired when players learn privately. (Such a divergence of beliefs is, almost by definition, unlikely, but depending on the signal structure, it may occur with positive probability in equilibrium.) For example, suppose that the public signals that the players use to determine their actions in equilibrium suggest very strongly that nature has chosen state A, but player 1's private payoffs indicate state B even more strongly, so that his private belief assigns probability close to 1 to state B. In that situation, player 1 believes that eventually the public belief will converge to state B

[1] Related work on that topic includes Vives (1989, 2002), Kihlstrom and Vives (1992), Mirman et al. (1994), Kuhn and Vives (1995), Bergemann and Välimäki (2000).

if players continue to experiment, but (i) in the future, equilibrium may specify only actions that yield the same payoffs in all states, so that no further learning occurs, or (ii) the current public belief may put so little weight on state B that the expected time before convergence is very long, even if the equilibrium calls for continued experimentation. Further, in state B the equilibrium actions specified for state A may yield a lower payoff to player 1 than the actions designed to punish player 1 for a deviation in state A yield, and so player 1 deviates. In response, however, the other players may conclude that player 1 must believe in state B, and so the public belief may adjust very far toward state B. But then such a deviation may be profitable for player 1 even when his private information is consistent with state A, if the punishment profile specified for him in state B gives him a higher payoff when the actual state is A than does the on-path profile specified in state A.

In this paper, the way around such complications is to construct equilibrium block strategies, in which actions depend only on recent public signals. For example, if the blocks are 100 periods long, then the actions in period 199 are determined by the public signals from periods 101 through 198 only. In effect, the players coordinate their actions using a public "dummy belief" over states, which is reset to the prior at the beginning of each block. This approach bounds the possible divergence between the private beliefs and the dummy belief. In particular, a patient player prefers not to deviate even if the dummy belief, after a sequence of misleading public signals, calls for an action profile that he believes gives him a very low payoff for the duration of the current block: at the start of the next block, the dummy belief reverts to the prior, and with high likelihood, experimentation in the next block (and all future blocks) reveals the true state and enables the other players either to give him his equilibrium payoff or to effectively punish his deviation. The strategies constructed are public on the equilibrium path, but after deviating, a player uses a private strategy temporarily, as is described in detail later. For that reason, the strategies are neither a "belief-free" equilibrium, as in Hörner and Lovo (2009) and Hörner et al. (2011a), for example, nor a "perfect public ex post" equilibrium, as in Fudenberg and Yamamoto (2010).

An additional difficulty in that construction is that a player being minmaxed may be able to unilaterally block learning: a player may have an action such that regardless of the actions chosen by the other players, the resulting payoff is uninformative about the state. (For example, the payoff to an investment project that requires the participation of all players cannot be learned if one player chooses not to contribute.) If other players need to tailor their punishment actions to the state, such a failure of learning clearly creates a problem for constructing credible threats. The solution, briefly, is to demonstrate that there is a strategy available to the punishing players such that if the player being punished can block learning, he can do so only at the cost of low payoffs in the stage game.

It is straightforward to extend the folk theorem to the case where players have private information about the state of the world before the beginning of the game or where they receive private signals beyond their realized payoffs during the game. The folk theorem

also extends to the case that there are no public signals, but instead players' private signals are very closely correlated (conditional on the state and the action profile). The construction, which requires the additional assumptions that there are at least three players and that cheap-talk communication is possible after each period, uses players' announcements of their private signals in place of public signals. A player who announces a different signal from the others is punished as though he had deviated, and thus he prefers to report truthfully.

Cripps et al. (2008) (henceforth CEMS) consider a problem related to the issue in this paper. In their setting, too, nature selects a state once and for all according to a known prior. At least some of a finite set of players are initially uninformed about nature's choice. In each period, players receive noisy private signals of the state. There is no game, however; the players do not interact with each other. Instead, CEMS study conditions under which the information in these private signals eventually generates "common learnings," that is, the realized state ω becomes common 1-belief (each player believes with probability 1 that each player believes with probability 1 ... that the state is ω) in the limit. One goal of that research is to establish conditions for a folk theorem. The relationship between their work and this paper is discussed briefly in the conclusion.

The literature on repeated games with incomplete information is extensive; see Aumann and Hart (1992) and Aumann et al. (1995). Kalai and Lehrer (1993, 1995) examine the case where players are uncertain both about payoffs and about the strategies of other players. Nearer to this paper, Gossner and Vieille (2003) and Wiseman (2005) establish folk theorems (without discounting and with discounting, respectively) for repeated games with symmetric learning. The most closely related papers are Fudenberg and Yamamoto (2010) and Yamamoto (2010), who study repeated games with imperfect monitoring in which players are initially uncertain both about payoffs and about the monitoring structure; players' private information about their own actions generates asymmetric learning about the state of nature. (Monitoring is public in Fudenberg and Yamamoto 2010, and private in Yamamoto 2010.) They derive conditions on the monitoring structure under which a folk theorem holds.[2] Those conditions rule out the possibility that any player can unilaterally block learning; the *statewise full-rank condition* requires that for any two states, there is an action profile with the property that the resulting public signal distinguishes between the states, even if a single player deviates. As described above, that assumption removes a significant potential difficulty in generating a folk theorem. The key assumption in this paper (Assumption 1) is that an action profile must generate different public signals in different states only if the resulting payoffs are different, that is, if any player can distinguish between states by observing his payoffs, then the public signals also distinguish between them. An example that satisfies Assumption 1 but not the statewise full-rank condition is the joint investment project mentioned above: each period, the project succeeds if and only if both players exert effort. If they do, then the expected return depends on the state of the world. If any player fails to exert effort, then the return is zero in any state.

[2] Fudenberg and Yamamoto (2011b) show that those conditions can be weakened if players have initial private information about the state.

On the other hand, the models in Fudenberg and Yamamoto (2010) and Yamamoto (2010) allow not only imperfect monitoring, but uncertainty about the monitoring structure itself, which are substantial extensions relative to the perfect monitoring of actions assumed in this paper. Another difference is that the proof here is constructive, while Fudenberg and Yamamoto (2010) and Yamamoto (2010) use an extension of the linear programming techniques of Fudenberg and Levine (1994).

Another strand of related literature is on reputation games with two patient players; the difference in those models is that the uninformed player gets no signals of the informed player's type except through the actions of the informed player. Mailath and Samuelson (2006) provide an extensive summary of research in that area; more recent work includes Peski (2008) and Atakan and Ekmekci (2009, 2011, forthcoming). Similarly, Hörner and Lovo (2009) and Hörner et al. (2011a) study repeated games where each player has private information about the state at the beginning of the game, but does not receive any further signals (except for the actions of other players).

The structure of the rest of the paper is as follows. Section 2 presents the model, Section 3 contains the folk theorem and its proof, Section 4 extends the result to the case of private, almost-public signals, and Section 5 concludes.

2. Model

There is a set Ω of K possible states of the world and $N > 1$ expected-utility maximizing players repeating a stage game infinitely. The players have a common discount factor δ. Before the start of play, the state of the world ω is chosen once and for all according to the commonly known distribution Φ, which assigns strictly positive probability to each of the K possible states. In each different state ω, the stage game $G(\omega)$ has different payoffs but the same set of action profiles $A = A_1 \times A_2 \times \cdots \times A_N$, where A_i is the set of actions for player i; let L denote the number of action profiles. A mixed action profile is denoted by α and (mixed) actions are publicly observed.[3] In each period, payoffs are determined as follows: action profile a is played, and then player i's random payoff U_i, which is observed by player i only, is drawn from a distribution $F_i(a, \omega)$ that depends on the action profile a and the state ω. At the same time, a random public signal Z, whose distribution $F_z(a, \omega)$ also varies with a and ω, is realized. Each time action profile a is played, independent draws from $F_z(a, \omega)$ and $F_i(a, \omega)$ are made. The expected payoff from mixed action profile α in state ω, is given by

$$EU_i(\alpha, \omega) = \sum_{a \in A} \alpha(a) E_{F_i(a,\omega)}[U_i],$$

where $\alpha(a)$ is the probability of pure action profile a under mixed profile α. A public randomization device is available to the players.

The private payoffs U_i jointly identify the state probabilistically: for any two states ω and ω', there is some player i and some action profile $a \in A$ such that $F_i(a, \omega)$ and

[3]The implications of relaxing the assumption that mixed strategies rather than just the actions actually played are observable are discussed in Section 5.

$F_i(a, \omega')$ differ on a set of positive measure. (That assumption is without loss of generality; two states that yield the same payoff distributions can be treated as a single state.) The key assumption is that the public signals Z contain (probabilistically) at least the information in the private payoffs.

Assumption 1. *For any action profile $a \in A$ and pair of states $\omega, \omega' \in \Omega$ such that $F_i(a, \omega)$ and $F_i(a, \omega')$ differ for some player i, the distributions $F_z(a, \omega)$ and $F_z(a, \omega')$ also differ.*

Thus, the public signals also identify the state. The importance of Assumption 1 is that it ensures that any player i can be (approximately) minmaxed, even when the state is unknown, as is described in detail in the proof of Proposition 1. The idea, roughly, is that the punishing players start by playing a tentative punishment profile. If player i responds with an action that gives him a high payoff, then the other players (i) observe (probabilistically) the high payoff, (ii) conclude that the state is such that their tentative profile is not effective, and (iii) switch to a new punishment profile. That process continues until an effective punishment profile is found.

No assumptions are made about (i) the informativeness about the state ω of private payoff U_i, conditional on action profile a and public signal Z, (ii) the informativeness about the state ω of public signal Z, conditional on action profile a and the vector of private payoffs U, (iii) the correlation between any two players' payoffs U_i and U_j, conditional on action profile a, public signal Z, and state ω, (iv) the dependence of private payoff U_i on public signal Z, conditional on action profile a and state ω, (v) the relationship between $F_z(a, \omega)$ and $F_z(a', \omega)$ or between $F_i(a, \omega)$ and $F_i(a', \omega)$ across states of the world ω (that is, about whether learning the payoff from action profile a is informative about the payoff from action profile a'), or (vi) whether learning the payoff from action profile a to player i is informative about the payoff from a to any other player j. The model allows for arbitrary correlation of payoffs in each stage across players in any given state of the world, so player i's beliefs about player j's private payoffs and other higher order beliefs are unrestricted. The supports of the distributions F_i and F_z are not required to be the same in each state of the world, so there may be public or private realizations that perfectly reveal the state. In fact, the distributions may be degenerate, so that payoffs or signals are nonstochastic (conditional on the state). Alternatively, if two states do have the same support, players may never be able to learn the state for certain.

Examples of information structures allowed by the model include the following.

Example 1 (Purely public learning). Suppose that each player i's private payoff U_i is independent of the state ω, conditional on action profile a and public signal Z. In that case, private payoffs are uninformative about the state of the world and all learning is public. This situation corresponds to the model in Wiseman (2005). ◊

Example 2 (Noisy public signal of payoffs). Suppose that the public signal Z is an N-dimensional vector and that conditional on the vector of realized payoffs U, Z is independent of the state ω. In particular, the ith component of public signal Z_i is the sum

of each player i's private payoff U_i and a mean-zero noise term ε_i that is independent of ω. Then all available information about the state is contained in the private payoffs: conditional on the vector U, there is no information in the public signal Z (although Z may be informative to an individual player, who observes only his own payoff U_i). ◊

Example 3 (Known own payoffs). Suppose that each private payoff identifies player i's payoff function for each i. That is, whenever the supports of the distributions $F_i(a, \omega)$ and $F_i(a, \omega')$ have a nonempty intersection, then $F_i(a', \omega) = F_i(a', \omega')$ for all $a' \in A$. In this case, players completely learn their own payoffs after a single period, although they may still be uncertain about the payoffs of other players. This setting fits (after the first period) models with single- or multisided reputations (and gradual public learning of players' types). ◊

A *public history* H^t contains the action profiles chosen and public signals realized in periods 1 through $t-1$; H^1 is the null set. A *private history* for player i, h_i^t, includes the information in the public history plus player i's realized payoffs; h_i^1 is again the null set. Let $b_i^t(h_i^t) \in \Delta_K$ denote player i's *private belief* about the state of the world at the start of period t. Bayesian updating of the prior Φ using the information in the private history (including any information about other players' payoffs revealed by their action choices) yields the private belief. In contrast, the public *dummy belief* $B^t(H^t) \in \Delta_K$ is obtained by updating using the information in the public history *except* information about private payoffs implicit in actions. Thus, the dummy belief can be calculated without reference to equilibrium strategies. Similarly, for $s \leq t$, obtain the *period-s-truncated dummy belief* $B^{t\backslash s}(H^{t\backslash s}) \in \Delta_K$ by updating the prior using the *period-s-truncated public history* $H^{t\backslash s}$ from periods s through $t-1$ (again ignoring information about private payoffs); $B^{t\backslash t}(H^{t\backslash t})$ equals the prior Φ. Let $EU(\alpha, b) \equiv \sum_{k=1}^{K} b_k EU(\alpha, \omega_k)$ be the expected payoff vector from action α given beliefs b.

Define $\overline{u}$ as the highest absolute value of expected stage-game payoffs across all players, all action profiles, and all states of the world, so that

$$\overline{u} = \max_{a \in A, \omega \in \Omega, i \in N} |EU_i(a, \omega)|.$$

Define $\overline{U} > \overline{u}$ as $\sqrt{N}\overline{u}$; no vector of expected payoffs can have a length greater than $\overline{U}$.

For each player i and state ω, let $m^i(\omega) \in \Delta A_{-i}$ be the action profile that minmaxes player i in state ω and let $e_i(\omega)$ denote the corresponding minmax payoff; that is,

$$e_i(\omega) \equiv \min_{\alpha_{-i} \in \times_{j \neq i} \Delta A_j} \max_{a_i \in A_i} EU_i((a_i, \alpha_{-i}), \omega).$$

Let $V(\omega)$, defined as the convex hull of the set $\{EU(a, \omega) \in \mathbf{R}^N : a \in A\}$, be the set of feasible payoffs in state ω. Then $V^*(\omega) \equiv \{u \in V(\omega) : u_i > e_i(\omega) \text{ for all } i \in N\}$ is the set of strictly individually rational feasible payoffs. Extending the standard full-dimensionality condition for folk theorems to the case of multiple states, I assume that $V^*(\omega)$ has dimension N for all $\omega \in \Omega$.

3. Results

The main result is 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 exists a sequential equilibrium in which (with very high probability) players achieve (very close to) the payoff specified for the realized state, as long as players are patient enough.

Proposition 1. *Let $\varepsilon > 0$ and payoffs $v^*(\omega_1) \in \operatorname{int}(V^*(\omega_1)), \dots, v^*(\omega_K) \in \operatorname{int}(V^*(\omega_K))$ be given, and let Φ be a prior belief that assigns strictly positive probability to each state. If Assumption 1 holds, then there exists $\underline{\delta} < 1$ such that for all $\delta > \underline{\delta}$, there is a sequential equilibrium that with probability at least $1 - \varepsilon$, conditional on any state ω being realized, yields a payoff vector within ε of $v^*(\omega)$. In equilibrium, each player i's private belief converges to the truth: $\lim_{t\to\infty} b_i^t(h_i^t)[\omega] = 1$ with probability 1.*

The proof is constructive, using elements of Fudenberg and Maskin's (1986) and Gossner and Vieille's (2003) folk theorems and Hörner and Olszewski's (2006) block strategies. The equilibrium uses play "blocks" of $M + T$ periods each. (The values M and T are described below.) A block has a "target payoff" $v(\omega)$ for each state ω. Within each block, players follow strategies that rely only on the history of public signals since the start of the block and, in particular, on the dummy belief truncated at the beginning of the block. The first M periods are used in experimentation to learn the state of the world (since all previous information about the state is ignored). The most likely state $\hat{\omega}$ is identified and then in the remaining $T \gg M$ periods, players choose the action profile that yields the target payoff $v(\hat{\omega})$ in state $\hat{\omega}$. If (i) M is large enough to identify the true state with high probability, (ii) $T/(M + T)$ is close to 1, so that nearly all of the periods within the block are spent playing the target action profile, and (iii) players are very patient, then the expected payoff from the block when the realized state is ω is very close to the target payoff $v(\omega)$.

There are $2N + 1$ types of blocks: an on-equilibrium block, a minmax block for each player, and a postdeviation block for each player. The initial block is on-equilibrium, as are all subsequent blocks until the first deviation. If player i deviates during a block, then a minmax-i block begins in the next period and all subsequent blocks (until another player deviates) are postdeviation-i blocks. The target payoffs in the on-equilibrium block correspond to the desired equilibrium payoffs $v^*(\omega)$. The target payoff for player i in his minmax block is roughly $e_i(\omega)$. (Play in the minmax blocks is somewhat different from play in the other blocks, as described in the proof.) Target payoffs in the postdeviation blocks are chosen so that $v_i^{\text{dev-}i}(\omega) < v_i^*(\omega)$ and $v_i^{\text{dev-}i}(\omega) < v_i^{\text{dev-}j}(\omega)$ for each state ω. That is, the payoff to a deviator is lower than he would get in equilibrium or when punishing another player, regardless of the state. A patient player, then, does not deviate, on or off the equilibrium path, regardless of his private beliefs. Simultaneous deviations are ignored.

In a minmax block, the challenge is that the punished player i may be able to block experimentation, making it impossible for the other players to identify the state and play the appropriate minmax profile. For example, player i may have an action a_i such

that any profile that includes a_i yields a payoff that is independent of the state. The resulting lack of learning is a problem, since a profile that minmaxes player i in one state may give him a high payoff in another. It turns out, though, that there is a strategy for the other players that prevents player i from simultaneously getting a high payoff and blocking learning. Suppose that the other players play $m^i(\omega)$, the profile that minmaxes player i in state ω. If player i's response yields him a payoff above (in expectation) his state-ω minmax $e_i(\omega)$, then necessarily his payoff reveals that the state is not ω. The public signal, by Assumption 1, reveals that information as well, and so the other players learn. The other players, then, can start a minmax-i block by playing the minmax profile $m^i(\omega')$ corresponding to the state ω' in which the minmax payoff $e_i(\omega')$ is lowest, continue playing that profile unless they learn that the state is not ω', then switch to the minmax profile for the state with the second-lowest minmax payoff, and so on. If they follow such a strategy, then, conditional on any state ω being realized, the number of periods in the block in which player i can attain a payoff above his minmax payoff for that state $e_i(\omega)$ is bounded (probabilistically). Thus, if the number of periods in the minmax block is high enough, then player i's average payoff in the block cannot be (much) greater than $e_i(\omega)$.

During a minmax-i block, deviations by player i are ignored. Therefore, the optimal strategy for player i is to play a best response to the equilibrium actions of the other players for the remaining periods of the minmax-i block. Note that as a consequence, the equilibrium strategies are not public strategies. The actions chosen by a player who has not deviated depend only on the public history, but the actions of a player who has deviated in general depend (during the subsequent minmax block) on his private beliefs.

Proof of Proposition 1. For each state $\omega \in \Omega$ and each player i, choose a payoff vector $v^{\text{dev-}i}(\omega) \in \operatorname{int}(V^*(\omega))$ such that $v_i^{\text{dev-}i}(\omega) < v_i^*(\omega)$ and $v_i^{\text{dev-}i}(\omega) < v_i^{\text{dev-}j}(\omega)$ for each $j \neq i$; these feasible, strictly individually rational payoffs are the long-run (that is, after the minmax block) payoffs after a deviation by player i. Without loss of generality, assume that

$$\varepsilon < \min\{1, v_i^*(\omega) - v_i^{\text{dev-}i}(\omega), v_i^{\text{dev-}j}(\omega) - v_i^{\text{dev-}i}(\omega), v_i^{\text{dev-}i}(\omega) - e_i(\omega)\}$$

for all players and states.

Choose a positive integer M to satisfy the following three conditions.

Condition 1. The integer M is divisible by L (the number of pure action profiles) and by K (the number of states).

Condition 2. Conditional on any state ω, updating the prior Φ with the M public signals that result from playing each action profile $a \in A$ M/L times yields a posterior probability that puts weight greater than $1/2$ on ω with probability at least $1 - \varepsilon/(8\overline{U})$.

Condition 3. Take any two states $\omega \neq \omega'$, any player i, and any history of actions in which (i) players other than i always play $m^i(\omega)$ and (ii) in at least M/K periods, player i plays an action $a_i \in A_i$ such that $EU_i((a_i, m^i(\omega)), \omega') > e_i(\omega)$. (It need not be the same

such action in each of the M/K periods.) Then conditional on state ω', updating the prior Φ with the resulting public signals yields a posterior probability that puts weight less than $\varepsilon/2$ on ω with probability at least $(1-\varepsilon/(8\overline{U}))^{1/K}$.

The second condition requires that M periods of experimentation, split equally among all action profiles, are with high probability sufficient to learn the state with great precision, starting from the prior. The third condition says that if the other players play the state-ω minmax profile for player i when the actual state is ω', and player i's response yields an expected (in state ω') payoff above his state-ω minmax, then within M/K periods the players very likely learn very precisely that the state is not ω.

Condition 4. Given M, choose a positive integer T greater than $(4M/\varepsilon)\overline{U}-M$, so that

$$\overline{U}-\left[\frac{M}{M+T}(-\overline{U})+\frac{T}{M+T}\overline{U}\right]<\frac{\varepsilon}{4}.$$

Thus, if action profile a is played for M periods and action profile a' is played for T periods in state ω, then the average payoff (without discounting) in those $M+T$ periods will be close to $U(a',\omega)$, regardless of how far apart $U(a',\omega)$ and $U(a,\omega)$ are.

Condition 5. Given M and T, choose a positive integer $T'\geq T$ such that

$$\left[\frac{M+T}{M+T'+1}(-\overline{U})+\frac{T'-T+1}{M+T'+1}v_i^{\text{dev-}i}(\omega)\right]-\left[\frac{M+1}{M+T'+1}\overline{U}+\frac{T'}{M+T'+1}e_i(\omega)\right]>\varepsilon$$
$$\text{for every player } i \text{ in every state } \omega,$$

so that the average (undiscounted) payoff to player i from T' periods of being minmaxed and $M+1$ periods of any other action profile is lower than his long-run payoff after deviating, even if the payoff from the current postdeviation block happens to be low. (The minmax blocks are constructed below to last for $M+T'$ periods.)

Finally, the value of the discount factor δ is described later.

The equilibrium strategies are based on $2N+1$ types of blocks, as described above. There are an on-equilibrium block, a postdeviation block for each player, and a minmax block for each player.

Within-block strategies. The on-equilibrium block has length $M+T$ periods. Order the action profiles A arbitrarily as $A=\{a^1,\dots,a^L\}$. The players play profile a^1 for the first M/L periods of the block, a^2 for the next M/L periods, and so on for the rest of the first M periods. In a block that begins in period s, that experimentation yields the truncated dummy belief $B^{s+M\backslash s}(H^{s+M\backslash s})$. Let $\hat{\omega}(H^{s+M\backslash s})$ denote the state given the highest probability under $B^{s+M\backslash s}(H^{s+M\backslash s})$. (Ties can be broken arbitrarily.) For the remaining T periods of the block, players play the profile that results in payoff $v^*(\hat{\omega}(H^{s+M\backslash s}))$ in state $\hat{\omega}(H^{s+M\backslash s})$. Within the block, simultaneous deviations are ignored. If player i deviates unilaterally, then the on-equilibrium block ends immediately and a minmax-i block begins in the next period (as described below).

Play in a postdeviation-i block is similar. The only difference is that in the last T periods, the profile yielding payoff $v^{\text{dev-}i}(\hat{\omega}(H^{s+M\backslash s}))$ in state $\hat{\omega}(H^{s+M\backslash s})$ is played, rather than the profile yielding $v^*(\hat{\omega}(H^{s+M\backslash s}))$. The construction of the block is the same; only the target payoffs differ.

The construction of a minmax-i block, which has length $M+T'$ periods, is somewhat different. For convenience, suppose that the K states of the world are indexed such that $e_i(\omega_1) \leq e_i(\omega_2) \leq \cdots \leq e_i(\omega_K)$. That is, player i's minmax payoff is lowest in state ω_1 and increases up through state ω_K. In each period of a minmax-i block that begins in period s, players other than i play the minmax profile $m^i(\omega_k)$ corresponding to the lowest-indexed state ω_k that has probability at least $\varepsilon/2$ under the period-s-truncated dummy belief $B^{t\backslash s}(H^{t\backslash s})$. (For example, they start the block by playing $m^i(\omega_1)$ if state ω_1 has probability $\varepsilon/2$ or higher under the prior Φ.) When there are R periods remaining in the block, player i plays the R-period best response, given his private belief, to the public strategies used by the other players in the rest of the block. (That is, he constructs a best response by backward induction, as though the last period of the minmax block were the end of the game.) If he has multiple best responses, his equilibrium-specified choice among them is common knowledge. Since deviations by player i (who is using a private strategy) may not be detectable, none of his actions is considered a deviation: a simultaneous deviation by players i and j is treated as a unilateral deviation by player j. (That is, the other players do not interpret any action by player i as a deviation.) Among the other players, simultaneous deviations are ignored. If any player $j \neq i$ deviates unilaterally, then the minmax-i block ends immediately and a minmax-j block begins in the next period (as described below).

Transitions between blocks. Play begins with an on-equilibrium block. An on-equilibrium block in which there are no unilateral deviations is followed by another on-equilibrium block, and a postdeviation-i block with no unilateral deviations is followed by another postdeviation-i block. A minmax-i block with no unilateral deviations (by a player other than player i) is followed by a postdeviation-i block.

During an equilibrium block or postdeviation-i block, if player j (possibly $j=i$) deviates unilaterally, then a transition to a minmax-j block occurs in the next period. Finally, in the case of a deviation during a minmax-i block by a player j other than player i, a transition to a minmax-j block occurs in the next period.

Beliefs. On the equilibrium path, each player's private belief $b_i^t(h_i^t)$ is derived by Bayesian updating of the prior Φ using the information in his private history h_i^t. Actions, because they depend only on the public signals, reveal nothing about players' private payoffs. Off equilibrium, a player's decision to deviate is also treated as uninformative. After player i deviates, his actions in the subsequent minmax-i block may depend on his history of private payoffs, and so the other players' private beliefs incorporate that information. (Note that during that time, player i has no incentive to deviate from a short-run best response so as to influence others' beliefs, since their actions depend only on the public history. On the other hand, his choice of a short-run best response reflects any expected benefit from influencing the flow of public signals.)

Payoffs. Condition 2 implies that with probability at least $1-\varepsilon/(8\overline{U})$, the M periods of experimentation at the beginning of an on-equilibrium block identify the true state ω

as the most likely. In that event, the expected (undiscounted) average payoff of the block is within $\varepsilon/4$ of its target payoff $v^*(\omega)$, by Condition 4. If the wrong state is identified, the expected payoff cannot differ from $v^*(\omega)$ by more than twice $\overline{U}$ (the greatest possible magnitude of any payoff vector). Conditional on state ω, therefore, the expected undiscounted average payoff of an on-equilibrium block (in which there are no deviations) is within

$$\left(1-\frac{\varepsilon}{8\overline{U}}\right)\frac{\varepsilon}{4}+\frac{\varepsilon}{8\overline{U}}2\overline{U}<\frac{\varepsilon}{2}$$

of $v^*(\omega)$. Similarly, the expected undiscounted average payoff of a postdeviation-i block, conditional on state ω, is strictly within $\varepsilon/2$ of the target $v^{\text{dev-}i}$.

Now consider the payoff to player i from a minmax-i block in state ω. With probability $1-\varepsilon/(8\overline{U})$ or higher, the number of periods in which player i's expected payoff is above his minmax $e_i(\omega)$ is no greater than M: Condition 3 ensures that the only way player i can prevent the other players from learning the state (and punishing him accordingly) is by playing actions that yield him payoffs no better than $e_i(\omega)$. (For example, if $e_i(\omega)$ is the third-lowest minmax payoff for player i across all states, then players following the strategies for a minmax-i block most likely take no more than $2M/K$ periods of high payoffs for player i to rule out the two states with lower minmaxes.) By an argument similar to the one in the last paragraph, then, Condition 4 (plus the fact that $T'\geq T$) ensures that the expected payoff to player i from a minmax-i block in state ω is strictly within $\varepsilon/2$ of $e_i(\omega)$. The expected payoff to any other player must be at least $-\overline{U}$.

Since the expected (conditional on state ω) undiscounted average payoff of an on-equilibrium block is strictly within $\varepsilon/2$ of its target payoff $v^*(\omega)$, for high enough values of the discount factor δ, the expected discounted average payoff is also within $\varepsilon/2$ of $v^*(\omega)$. Similarly, the expected discounted average payoffs of a postdeviation-i block and a minmax-i block are within $\varepsilon/2$ of their target payoffs $v^{\text{dev-}i}(\omega)$ and $e_i(\omega)$, respectively, when δ is high. Formally, let $\kappa<\varepsilon/2$ be the maximum (across states ω and players i) difference between the expected undiscounted average payoff and the target payoff of any of the three types of blocks. Then a sufficient condition on the value of $\delta<1$ is the following condition.

Condition 6. For any $S\leq M+T'$ and any sequence of payoff vectors $\{x_s\}_{s=1}^S$, where each $x_s\in[-\overline{U},\overline{U}]^N$,

$$\left\|\frac{1-\delta}{1-\delta^S}\sum_{s=1}^S\delta^{s-1}x_s-\frac{1}{S}\sum_{s=1}^S x_s\right\|<\frac{\varepsilon}{2}-\kappa.$$

.

Best-response conditions. The value of the discount factor δ can be chosen so that no player has an incentive to deviate, on or off the equilibrium path, whatever his private belief about the state or his higher order beliefs (about others' beliefs), and whatever the truncated dummy belief. In a minmax-i block, player i has no incentive to deviate: he is playing a short-run best response within the block, and his actions do not affect play after the block. On the equilibrium path, there is no profitable deviation if Condition 6 and the following condition hold.

Condition 7. For every player i in every state ω,

$$\overline{U} + \sum_{t=2}^{M+T'+1} \delta^{t-1}\left[e_i(\omega) + \frac{\varepsilon}{2}\right] + \sum_{t=M+T'+2}^{\infty} \delta^{t-1}\left[v_i^{\text{dev-}i}(\omega) + \frac{\varepsilon}{2}\right] < \sum_{t=1}^{M+T} \delta^{t-1}[-\overline{U}] + \sum_{t=M+T+1}^{\infty} \delta^{t-1}\left[v_i^*(\omega) - \frac{\varepsilon}{2}\right].$$

After a deviation in state ω, player i's long-run payoff is $v_i^{\text{dev-}i}(\omega)$: play switches to a postdeviation-i block after a minmax-i block. Even if (i) by deviating, player i can get the highest possible immediate payoff, (ii) player i expects to get the lowest possible payoff for the rest of the current block (for example, because of a divergence between his private belief and the truncated dummy belief), and (iii) the number of periods remaining in the current block is as high as possible ($M+T$), a patient player prefers not to deviate. Because $v_i^*(\omega) - v_i^{\text{dev-}i}(\omega) > \varepsilon$, Condition 7 is satisfied for high enough values of δ.

Similarly, because $v_i^{\text{dev-}j}(\omega) - v_i^{\text{dev-}i}(\omega) > \varepsilon$, if δ is high enough, then no player $j \neq i$ has a profitable deviation during a postdeviation-i block or a minmax-i block, even if minmaxing player i gives player j the lowest possible payoff. All that remains is to show that player i cannot gain from deviating during a postdeviation-i block. In that case, a deviation does not affect player i's long-run payoff (which is still $v_i^{\text{dev-}i}(\omega)$), but in the short run, the deviation triggers a minmax-i block and lowers his payoff. In particular, the following condition (together with Condition 6) implies that no such deviation is profitable.

Condition 8. For every player i in every state ω,

$$\overline{U} + \sum_{t=2}^{M+T'+1} \delta^{t-1}\left[e_i(\omega) + \frac{\varepsilon}{2}\right] < \sum_{t=1}^{M+T} \delta^{t-1}[-\overline{U}] + \sum_{t=M+T+1}^{M+T'+1} \delta^{t-1}\left[v_i^{\text{dev-}i}(\omega) - \frac{\varepsilon}{2}\right].$$

That is, even if deviating results in the highest possible immediate payoff and not deviating results in the lowest possible immediate payoff for the rest of the block, the overall payoff from deviating is lower, no matter how many periods of the block remain. Condition 5 guarantees that Condition 8 is satisfied for high δ.

Finally, note that since for high δ, the expected payoff from an on-equilibrium block, conditional on any state ω being realized, is within $\varepsilon/2$ of $v^*(\omega)$, so is the expected equilibrium payoff of the repeated game. Since both payoff realizations and public signals are independent and identically distributed across periods (conditional on the state and actions), a central limit theorem implies that the undiscounted average of the first S discounted block payoffs $\overline{v}^S$, (that is,

$$\overline{v}^S = \frac{1}{S}\sum_{s=1}^{S}\left(\frac{1-\delta}{1-\delta^{M+T}} \sum_{t=1}^{M+T} \delta^{t-1} U_{(s-1)(M+T)+t}\right),$$

where U_t is the vector of realized payoffs in period t), converges over time to a normally distributed random vector with a mean that is also within $\varepsilon/2$ of $v^*(\omega)$; the variance

shrinks to zero. Thus, we can choose S^* such that with probability at least $1-\varepsilon$, $\overline{v}^S(\omega)$ is within $3\varepsilon/4$ of $v^*(\omega)$ for all $S \geq S^*$. Therefore, if δ is high enough, then the limiting discounted average of the block payoffs (which is the realized payoff of the repeated game) is with probability at least $1-\varepsilon$ within ε of $v^*(\omega)$. A sufficient condition on the value of $\delta < 1$ is the following condition.

Condition 9. For any sequence of payoff vectors $\{x_s\}_{s=1}^{S^*}$, where each $x_s \in [-\overline{U}, \overline{U}]^N$,

$$\left\| \frac{1-\delta^{M+T}}{1-(\delta^{M+T})^{S^*}} \sum_{s=1}^{S^*} (\delta^{M+T})^{s-1} x_s - \frac{1}{S^*} \sum_{s=1}^{S^*} x_s \right\| < \frac{\varepsilon}{4}.$$

If the discount factor δ is high enough to satisfy Conditions 6–9, then the strategies and beliefs constructed constitute a sequential equilibrium in which, with high probability, the players achieve payoffs close to those specified for the realized state. Since players experiment at the beginning of each on-equilibrium block, their private beliefs converge to the truth with probability 1. □

In the model of Section 2, players get (direct) information about the state of the world only from the public signals and from their own realized payoffs. In many settings, though, players might learn about the state in other ways during the game or before play starts; some of the examples discussed in the Introduction fit that assumption. In such an environment, the strategies constructed in the proof of Proposition 1 are still an equilibrium; the key is that each player, no matter what is the gap between his private belief and the public signals in the current block, expects that the public signals from experimentation in future blocks will identify the true state with high probability. Proposition 2 states that result formally.

Proposition 2. *Suppose that in any realized state ω, each player i in each period $t \in \{0, 1, \ldots\}$ observes a private signal drawn from a distribution $F_{it}(\omega)$. Let $\varepsilon > 0$ and payoffs $v^*(\omega_1) \in \operatorname{int}(V^*(\omega_1)), \ldots, v^*(\omega_K) \in \operatorname{int}(V^*(\omega_K))$ be given, and let Φ be a prior belief that assigns strictly positive probability to each state. If Assumption 1 holds, then there exists $\underline{\delta} < 1$ such that for all $\delta > \underline{\delta}$, there is a sequential equilibrium that with probability at least $1-\varepsilon$, conditional on state ω, yields a payoff vector within ε of $v^*(\omega)$. In equilibrium, each player i's private belief converges to the truth: $\lim_{t\to\infty} b_i^t(h_i^t)[\omega] = 1$ with probability 1.*

4. Private, almost-public signals

In this section, Proposition 1 is extended to the case where there may be no public signal z, but, instead, in each period players receive private signals of the vector of realized payoffs, and those signals are very closely correlated.[4] Formally, the model is modified in two ways.

[4] Mailath and Morris (2002) show that any strict perfect public equilibrium with bounded recall of a game with public monitoring remains an equilibrium under private, almost-public monitoring. The strategies constructed in the proof of Proposition 1 do not have bounded recall, though: the short-run best response of a deviator depends on his private beliefs, and thus on his entire private history.

First, at the end of each period, each player i observes his own realized payoff U_i, as well as a private signal Y_i that reflects the payoffs of all players (including player i). The distribution of Y_i, $F_{y_i}(a, \omega)$, depends on the action profile a and the state ω. Every time action profile a is played, an independent draw from $F_{y_i}(a, \omega)$ is made. For each player i, the private signal Y_i contains (probabilistically) at least the information in the private payoff: for any action profile $a \in A$ and pair of states $\omega, \omega' \in \Omega$ such that $F_i(a, \omega)$ and $F_i(a, \omega')$ differ, the distributions $F_{y_i}(a, \omega)$ and $F_{y_i}(a, \omega')$ also differ. (A motivating special case is that $Y_i = (Y_i^1, \dots, Y_i^N)$, where $Y_i^i \equiv U_i$, is a vector of observations of each player's payoff. A player observes his own payoff exactly, and those of other players with noise.) A player's private signal may or may not be independent (conditional on a and ω) of his realized payoff or of other players' payoffs.

Second, after observing his signal and before the start of the next period, each player i can send a costless, public signal from the support of F_{y_i}. These signals are cheap talk: they are used to communicate private signals, but players are free to announce any signal that they wish. (These cheap-talk signals also are used to let a player specify the action profile that he wants to be played during his "reward period," as described in the proof of Proposition 3.) Player i's private history h_i^t contains the action profiles chosen and public announcements made in periods 1 through $t-1$, as well as player i's realized payoffs and private signals. Recall that a public randomization device is available.

Following Mailath and Samuelson (2006), say that players' private signals are *almost public* when each player, after receiving any private signal, no matter how unlikely, assigns high probability to the other players' having seen the same signal. Note that almost-public signals identify the state, since $F_{y_i}(a, \omega) \neq F_{y_i}(a, \omega')$ whenever $F_i(a, \omega) \neq F_i(a, \omega')$, and F_{y_i} and F_{y_j} are arbitrarily close.

Definition. For $\eta \in [0, 1]$, signals are *η-public* if for any player i, any action profile a, any state ω, any realized payoff u_i, and any private signal y_i, the probability, conditional on the event $U_i = u_i$ and $Y_i = y_i$, that $Y_j = y_i$ for all $j \in \{1, \dots, N\}$ is at least $1 - \eta$.

(Note that in the special case described above, η-public signals must also be η-perfect: player i observes his own payoff U_i, so player j's signal of U_i can match player i's only if it is exactly correct.)

Proposition 1 extends to this environment when the private signals are almost public. Proposition 3 gives the formal statement:

Proposition 3. *Suppose that $N \geq 3$. Let $\varepsilon > 0$ and payoffs $v^*(\omega_1) \in \operatorname{int}(V^*(\omega_1)), \dots, v^*(\omega_K) \in \operatorname{int}(V^*(\omega_K))$ be given, and let Φ be a prior belief that assigns strictly positive probability to each state. If Assumption 1 holds, then there exists $\underline{\delta} < 1$ and $\overline{\eta} > 0$ such that if $\delta > \underline{\delta}$ and signals are $\overline{\eta}$-public, there is a sequential equilibrium that with probability at least $1 - \varepsilon$, conditional on any state ω being realized, yields a payoff vector within ε of $v^*(\omega)$. In equilibrium, each player i's private belief converges to the truth: $\lim_{t\to\infty} b_i^t(h_i^t)[\omega] = 1$ with probability 1.*

The strategies used to prove Proposition 3 are very similar to those in the proof of Proposition 1, with public, cheap-talk announcements of private signals taking the place of public signals as a coordinating device. The additional difficulty is to provide incentives for honest reporting. The solution to that difficulty is to punish when players' reports do not agree. Thus, a player wants to announce whatever he believes that other players are going to announce, and when signals are almost public, he believes that other players see (and announce) the same signals that he observes. The close correlation of players' signals is important in this argument. Without it, a player who assigns high probability to state 1 but sees a signal that is very unlikely in that state may believe that other players probably did not see the same signal. As long as there are at least three players reporting their signals, it is possible to identify (and punish) any single player who makes an announcement that differs from all the others.[5]

Proof of Proposition 3. Choose the vectors of punishment payoffs $v^{\text{dev-}i}$, the integers M (satisfying Conditions 1–3), T (satisfying Condition 4), and T' (satisfying Condition 5), and the value of δ (satisfying Conditions 6–8) just as in the proof of Proposition 1. Also as before, there are $2N+1$ types of blocks: an on-equilibrium block, a postdeviation block for each player, and a minmax block for each player.

Within-block strategies. Just as before, the on-equilibrium block and the postdeviation-i blocks have length $M+T$ periods, and the minmax-i blocks last for $M+T'$ periods (except as described below). At the end of each period (with one exception, also described below), players announce their private signals truthfully. If at least $N-1$ players announce the same signal y in period t, then the *dummy public signal* for period t, $\hat{z}^t$, is equal to y; otherwise, a *multilateral misreport* occurs. (In that case, no dummy public signal is generated, as described below.) A *period-s-truncated dummy public history* $\hat{H}^{t\backslash s}$ contains the action profiles chosen and dummy public signals reported in periods $s+1$ through $t-1$. A player *unilaterally misreports* if he reports a signal y' while the other players unanimously announce a different signal y. Note that either type of misreport may occur in equilibrium. As usual, a player *deviates* if he chooses an action in the stage game other than what is prescribed by the equilibrium strategies. A player who deviates or misreports is said to have committed a *violation*. Violations are punished in the same way that unilateral deviations are in the public-history case. (For that reason, blocks other than on-equilibrium blocks may occur in equilibrium, so it is misleading to label only the one type of block as "on equilibrium." Nevertheless, for the sake of consistency, that nomenclature is maintained.)

Consider first an on-equilibrium block that starts in period s. As long as no multilateral misreports occur, play is just as in the public-history case, treating the dummy public signals as the public signals and treating both types of violations as unilateral deviations. Specifically, the players play profile a^1 for the first M/L periods of the block, play a^2 for the next M/L periods, and so on for the first M periods. That experimentation yields a truncated dummy public history $\hat{H}^{s+M\backslash s}$. For the remaining

[5]This approach is similar to the construction introduced in Renault (2001) and generalized in Renault and Tomala (2004).

T periods of the block, players play the profile that results in payoff $v^*(\hat{\omega}(\hat{H}^{s+M\backslash s}))$ in state $\hat{\omega}(\hat{H}^{s+M\backslash s})$, where $\hat{\omega}(\hat{H}^{s+M\backslash s})$ is the state given the highest probability under $B^{s+M\backslash s}(\hat{H}^{s+M\backslash s})$. If there is a misreport (unilateral or multilateral) or a unilateral deviation in an on-equilibrium block, then the block ends and the next block starts immediately. (The type of the next block is determined according to the transition rules below.) Multilateral deviations are ignored.

Play in a postdeviation-i block is the same as in an on-equilibrium block, except that in the last T periods (barring misreports or unilateral deviations), the profile that yields payoff $v^{\text{dev-}i}(\hat{\omega}(\hat{H}^{s+M\backslash s}))$ in state $\hat{\omega}(\hat{H}^{s+M\backslash s})$ is played, rather than the profile that yields $v^*(\hat{\omega}(\hat{H}^{s+M\backslash s}))$.

For a minmax-i block starting in period s, play is again just as in the public-signal case, using the dummy public signals in the place of actual public signals and treating misreports as deviations. As in the public-signal case, suppose that the K states of the world are indexed such that $e_i(\omega_1) \leq e_i(\omega_2) \leq \cdots \leq e_i(\omega_K)$. As long as no player other than i has committed a violation, then in each period those players play the minmax profile $m^i(\omega_k)$ that corresponds to the lowest-indexed state ω_k that has probability at least $\varepsilon/2$ under $B^{t\backslash s}(\hat{H}^{t\backslash s})$. When there are R periods remaining in the block, player i plays the (or a specified) R-period best response, given his private belief, to the public strategies used by the other players in the rest of the block. Deviations by player i are ignored, as are simultaneous deviations by two or more of the other players. If a single player $j \neq i$ commits a unilateral violation, then the minmax-i block ends immediately and a minmax-j block begins in the next period (as described below). If a multilateral misreport occurs, then the block ends and the next block (a postdeviation-i block, as described below) starts immediately. To give player i the incentive to announce his private signals truthfully, there is potentially a single "reward period" added to the end of the block. In each period of the block that all players announce the same signal, the probability that such a reward period is played increases (from an initial value of zero) by $\Delta = 1/(M + T')$. (The total probability is bounded above by 1, since $M + T'$ is the length of the block.) At the end of the block, public randomization determines whether the reward period is played, and player i announces the action profile α^R to be played in the reward period; α^R is the profile that maximizes player i's expected payoff in the stage game, given his private belief. If the reward period is to occur, then α^R is played. Players do not make an announcement at the end of a reward period.

Transitions between blocks. Transitions are just as in the public-signal case, with two differences: unilateral misreports are treated like deviations and, after a multilateral misreport, the next block starts immediately. Play begins with an on-equilibrium block. An on-equilibrium block in which there are no unilateral violations is followed by another on-equilibrium block, and a postdeviation-i block with no unilateral violations is followed by another postdeviation-i block. A minmax-i block with no unilateral deviations or unilateral misreports (by a player other than player i) is followed by a postdeviation-i block. Those transitions occur immediately if a multilateral misreport occurs; otherwise, they occur after the end of the current block.

During an equilibrium block or postdeviation-i block, if player j (possibly $j = i$) commits a unilateral violation, then a transition to a minmax-j block occurs in the next period.[6] Finally, in the case of a unilateral violation during a minmax-i block by a player j other than player i, a transition to a minmax-j block occurs in the next period.

Beliefs. On the equilibrium path, each player's private belief $b_i^t(h_i^t)$ is derived by Bayesian updating of the prior Φ using the information in his private history h_i^t, including the private signals of the other players, which are believed to be truthfully announced. Because they depend only on the public signals, actions reveal nothing about players' private information. Off equilibrium, any player i's decision to deviate is also treated as uninformative, as is an announcement of a signal outside the support of F_{y_i}. After player i commits a unilateral deviation or misreport, his actions during the subsequent minmax-i block may depend on his private information, and so the other players' private beliefs incorporate that information.

Payoffs and best-response conditions. If the probability that players receive different private signals is zero ($\eta = 0$), then the arguments of the public-signal case ensure that the actions specified above are best responses, that with high probability the players achieve payoffs close to those specified for the realized state, and that private beliefs converge to the truth with probability 1. For small enough values of η, the expected (conditional on state ω) discounted average payoff of an on-equilibrium block, a postdeviation-i block, and a minmax-i block are still within $\varepsilon/2$ of $v^*(\omega)$, $v^{\text{dev-}i}(\omega)$, and $e_i(\omega)$, respectively, when δ is high, so the same arguments apply. (The behavior of a player i during a minmax-i block may change when η is slightly positive rather than 0, because he may have been playing only a weak best response, but in all other cases, the best-response conditions were satisfied with strict inequalities.) All that remains is to show that truthfully reporting private signals is a best response.

During an on-equilibrium block, for example, a player's announcement can influence future play in three ways. First, it affects whether a transition to the next block occurs immediately (as it does after a misreport). Second, it affects the type of future blocks by influencing whether a misreport occurs and which kind. Third, if another player unilaterally commits a violation in the future, then afterward his play in the short run depends on his private beliefs and thus on past announcements. With regard to the first effect, the greatest possible (undiscounted, expected) gain from a false report is the maximum difference between two payoffs in the stage game times the number of periods in the block, $2\overline{U}(M + T)$. The cost from the second effect is that unless one of the other players receives a signal different from his own (an event that occurs with probability no greater than η), player i's false report results in a unilateral misreport and the resulting punishment: a lower continuation payoff. The third effect has a negligible effect on expected payoffs for small values of η, because misreports become very rare and unilateral deviations are off-path (for any η) starting from any history. Thus, when η is low enough, Conditions 6 and 7, which ensure that player i has no incentive to deviate from his prescribed action during an on-equilibrium block, also guarantee

[6]The condition that $N \geq 3$ is used here. Note that if $N = 2$, then whenever the two players report different signals, both have misreported unilaterally.

that truthfully reporting private signals is optimal. Similarly, Conditions 6 and 7 also deter false reporting by player i during postdeviation-j blocks and minmax-j blocks, and Conditions 6 and 8 suffice during postdeviation-i blocks.

During a minmax-i block, unilateral misreports by player i do not affect the type of future blocks, so future punishment cannot deter false reporting by player i. Instead, the reward period gives player i the incentive to report truthfully. The continuation payoff from future blocks when η is small is close to $v^{\text{dev-}i}(\omega)$, conditional on state ω. Thus, player i's expected payoff from α^R (the profile, to be played in the reward period, that maximizes his expected payoff in the stage game) strictly exceeds his expected continuation payoff when the next block starts, whatever his private beliefs are. Let d denote the difference. Player i gains in expectation Δd at the end of the current block whenever his report matches the reports of the other players (and there are no misreports later in the block), so truthful reporting yields an expected gain of at least $\delta^{M+T'}(1-\eta)^{M+T'}\Delta d$. Alternatively, player i can gain from a false report only if at least one other player j receives a signal different from player i's: in that case, a false report can generate either a unilateral misreport by player j (so that future punishment is transferred to player j) or a multilateral misreport that ends the current block early. However, the event that some other player receives a different signal occurs with probability no greater than η. Thus, for δ and small enough η, truthful reporting is optimal.

Thus, when $N \geq 3$ and signals are η-public for low enough η, then the strategies and beliefs constructed constitute a sequential equilibrium in which with high probability the players achieve payoffs close to those specified for the realized state, and private beliefs converge to the truth with probability 1. □

As was the case with public signals, the strategies constructed above still constitute an equilibrium if players receive additional private signals about the state of the world before or during play.

Proposition 4. *Suppose that $N \geq 3$ and suppose that in any realized state ω, each player i in each period $t \in \{0, 1, \ldots\}$ observes a private signal drawn from a distribution $F_{it}(\omega)$. Let $\varepsilon > 0$ and payoffs $v^*(\omega_1) \in \text{int}(V^*(\omega_1)), \ldots, v^*(\omega_K) \in \text{int}(V^*(\omega_K))$ be given, and let Φ be a prior belief that assigns strictly positive probability to each state. If Assumption 1 holds, then there exists $\underline{\delta} < 1$ and $\overline{\eta} > 0$ such that if $\delta > \underline{\delta}$ and signals are $\overline{\eta}$-public, there is a sequential equilibrium that with probability at least $1-\varepsilon$, conditional on state ω, yields a payoff vector within ε of $v^*(\omega)$. In equilibrium, each player i's private belief converges to the truth: $\lim_{t\to\infty} b_i^t(h_i^t)[\omega] = 1$ with probability 1.*

5. Summary and discussion

This paper presents a form of folk theorem for repeated games with unknown payoffs when players receive both private and public signals. The environment is a generalization of the one in Wiseman (2005), and Proposition 1 is a stronger result than Theorem 1 in Wiseman (2005). (The result in the earlier paper establishes only that expected equilibrium payoffs are close to the targets, while Proposition 1 ensures that with high probability the realized payoffs are close.) Thus, this paper provides an alternative proof of Theorem 1 in Wiseman (2005).

The strategies described in Sections 2 and 3 rely on the assumption that the mixed actions used by players are observable only in the construction of the minmax blocks. If, instead, only the actions played are observed, then the folk theorems continue to hold for payoffs above the pure-strategy minmax payoffs, rather than the mixed-strategy minmax payoffs $e_i(\omega)$. Alternatively, it seems feasible to restore the result for the mixed-strategy minmax case using the arguments of Fudenberg and Maskin (1986, Section 6) and Gossner (1995), although the extension is not immediate.

As mentioned in the Introduction, CEMS study common learning. To focus on that topic, they strip away from their environment strategic considerations, as well as the ability for a player to learn other players' private signals through their actions and the ability of a player to try to manipulate that learning. Those complications do arise in this paper. The endogeneity of the private signals (they depend on actions) is another challenge. The availability of repeated-game punishments and incentives, however, greatly simplifies the task of identifying sufficient conditions for a folk theorem, relative to CEMS's environment. On the other hand, Proposition 1 says nothing about whether common learning occurs in the equilibrium construction or, more generally, about whether common learning is a necessary or sufficient condition for this kind of folk theorem.

It seems likely that the folk theorem in this paper (or some version of it) can be extended to settings where the state of the world changes over time or where the public signals are less revealing.

A dynamic state might evolve according to a Markov process, with players possibly unable to observe when a transition takes place. More generally, the transition and its observability might depend on the players' actions. Given the "repeated sampling" feature of the equilibrium constructed in this paper, it should be straightforward to modify that equilibrium to apply to the case of dynamic states. One complication is the relationship between the frequency at which the state changes to the speed at which players can learn the state through public signals: if the former is too high relative to the latter, then the folk theorem may fail. (Dutta 1995, Fudenberg and Yamamoto 2011a, Hörner et al. 2011b, and Peski and Wiseman 2012 derive folk theorems for classes of such stochastic games for the case where state transitions are publicly observed.)

Finally, it would be interesting to weaken the assumption that the public signal contains at least as much information (statistically) as all of the private payoffs taken collectively. It may be the case, then, that there are two distinct states of the world in which the distribution of the public signal is the same for a given action profile, even though one or more players get different payoffs from that profile in the different states. Without the assumption, however, the construction of the minmax blocks fails. Player i may get a high payoff in response to the state-ω_1 minmax profile without the other players ever realizing it. Roughly, it may be that player i learns the state, but the other players do not—an outcome that clearly limits the scope of punishment (and thus of the payoffs achievable in equilibrium). To take advantage of that asymmetry in information, however, player i must not reveal the state through his actions. His best response to the other players' punishment strategy, though, may vary with the state; there is thus a tension between the desire to use his private information and the desire to conceal it.

		Player 2	
		L	R
Player 1	T	1	2
	B	0	0

State 1

		Player 2	
		L	R
Player 1	T	3	0
	B	3	2

State 2

FIGURE 1. Player 1's payoffs in the two states.

Consider the following two-state, two-player example, in which player 1 can distinguish between states and player 2 cannot. Figure 1 shows the payoffs of player 1 in each state. In state 1, player 2 minmaxes player 1 with action L, resulting in a minmax payoff of 1. In state 2, R is the minmax profile and 2 is the minmax payoff. If player 2 is uncertain about the state and tries to punish player 1 using the strategy from the minmax block constructed in the previous section, then he begins by playing L, the minmax profile in state 1 (the state with the lower minmax payoff). Player 1's best response to L is T in both states, however, so her action does not reveal the state. Thus, in state 2, this strategy give player 1 a payoff of 3, which is not an effective punishment. Suppose, however, that instead player 2 punishes by beginning with action R. Player 1's best response depends on the state: it is T in state 1 and B in state 2. If he responds by playing T in both states, then in state 2, he gets 0, which is below his minmax. If he responds with B in both states, then in state 1, he gets 0, which is below his minmax. To best respond, then, he must condition his action on the state, thus revealing the state to player 2 and enabling her to punish effectively.

Similar situations arise in the study of repeated two-player zero-sum games with incomplete information on one side. The analysis of, for example, Aumann et al. (1995) (in particular, the notion of Blackwell 1956 approachability) can be applied to help determine the lower bound of feasible punishments for each player in each state.[7] (Hart 1985 is also relevant.)

Both extensions are subjects for future research.

REFERENCES

Atakan, Alp E. and Mehmet Ekmekci (2009), "A two-sided reputation result with long run players." Unpublished paper. [221]

Atakan, Alp E. and Mehmet Ekmekci (2011), "Reputation in the long-run with imperfect monitoring." Unpublished paper, Northwestern University. [221]

Atakan, Alp E. and Mehmet Ekmekci (forthcoming), "Reputation in long-run relationships." *Review of Economic Studies*. [221]

Aumann, Robert J. and Sergiu Hart, eds. (1992), *Handbook of Game Theory*, volume 1. North Holland, Amsterdam. [220]

Aumann, Robert J., Michael Maschler, and Richard E. Stearns (1995), *Repeated Games With Incomplete Information*. MIT Press, Cambridge, Massachusetts. [220, 237]

[7] Thanks to Johannes Hörner and Tristan Tomala for suggesting this link.

Bergemann, Dirk and Juuso Välimäki (2000), "Experimentation in markets." *Review of Economic Studies*, 67, 213–234. [218]

Blackwell, David (1956), "An analog of the minmax theorem for vector payoffs." *Pacific Journal of Mathematics*, 6, 1–8. [237]

Cripps, Martin W., Jeffrey C. Ely, George J. Mailath, and Larry Samuelson (2008), "Common learning." *Econometrica*, 76, 909–933. [220]

Dutta, Prajit (1995), "A folk theorem for stochastic games." *Journal of Economic Theory*, 66, 1–32. [236]

Fudenberg, Drew and David K. Levine (1994), "Efficiency and observability with long-run and short-run players." *Journal of Economic Theory*, 62, 103–135. [221]

Fudenberg, Drew, David K. Levine, and Eric S. Maskin (1994), "The Folk Theorem with imperfect public information." *Econometrica*, 62, 997–1039. [218]

Fudenberg, Drew and Eric S. Maskin (1986), "The Folk Theorem in repeated games with discounting or with incomplete information." *Econometrica*, 54, 533–554. [224, 236]

Fudenberg, Drew and Yuichi Yamamoto (2010), "Repeated games where the payoffs and monitoring structure are unknown." *Econometrica*, 78, 1673–1710. [219, 220, 221]

Fudenberg, Drew and Yuichi Yamamoto (2011a), "The Folk Theorem for irreducible stochastic games with imperfect public monitoring." *Journal of Economic Theory*, 146, 1664–1683. [236]

Fudenberg, Drew and Yuichi Yamamoto (2011b), "Learning from private information in noisy repeated games." *Journal of Economic Theory*, 146, 1733–1769. [220]

Gossner, Olivier (1995), "The folk theorem for finitely repeated games with mixed strategies." *International Journal of Game Theory*, 24, 95–107. [236]

Gossner, Olivier and Nicolas Vieille (2003), "Strategic learning in games with symmetric information." *Games and Economic Behavior*, 42, 25–47. [220, 224]

Hart, Sergiu (1985), "Nonzero-sum two-person repeated games with incomplete information." *Mathematics of Operations Research*, 10, 117–153. [237]

Hörner, Johannes and Stefano Lovo (2009), "Belief-free equilibria in games with incomplete information." *Econometrica*, 77, 453–487. [219, 221]

Hörner, Johannes, Stefano Lovo, and Tristan Tomala (2011a), "Belief-free equilibria in games with incomplete information: Characterization and existence." *Journal of Economic Theory*, 146, 1770–1795. [219, 221]

Hörner, Johannes and Wojciech Olszewski (2006), "The Folk Theorem for games with private almost-perfect monitoring." *Econometrica*, 74, 1499–1544. [224]

Hörner, Johannes, Takuo Sugaya, Satoru Takahashi, and Nicolas Vieille (2011b), "Recursive methods in discounted stochastic games: An algorithm for $\delta \to 1$ and a Folk Theorem." *Econometrica*, 79, 1277–1318. [236]

Kalai, Ehud and Ehud Lehrer (1993), "Rational learning leads to Nash equilibrium." *Econometrica*, 61, 1019–1045. [220]

Kalai, Ehud and Ehud Lehrer (1995), "Subjective games and equilibria." *Games and Economic Behavior*, 8, 123–163. [220]

Kihlstrom, Richard E. and Xavier Vives (1992), "Collusion by asymmetrically informed firms." *Journal of Economics and Management Strategy*, 1, 371–396. [218]

Kuhn, Kai Uwe and Xavier Vives (1995), *Information Exchange Among Firms and Their Impact on Competition*. European Commission, Luxembourg. [218]

Mailath, George J. and Stephen Morris (2002), "Repeated games with almost-public monitoring." *Journal of Economic Theory*, 102, 189–228. [230]

Mailath, George J. and Larry Samuelson (2006), *Repeated Games and Reputations*. Oxford University Press, Oxford. [221, 231]

Mirman, Leonard J., Larry Samuelson, and Edward E. Schlee (1994), "Strategic information manipulation in duopolies." *Journal of Economic Theory*, 62, 363–384. [218]

Peski, Marcin (2008), "Repeated games with incomplete information on one side." *Theoretical Economics*, 3, 29–84. [221]

Peski, Marcin, and Thomas Wiseman (2012), "A folk theorem for stochastic games with infrequent state changes." Unpublished paper, University of Texas at Austin. [236]

Renault, Jérôme (2001), "3-player repeated games with lack of information on one side." *International Journal of Game Theory*, 30, 221–245. [232]

Renault, Jérôme and Tristan Tomala (2004), "Learning the state of nature in repeated games with incomplete information and signals." *Games and Economic Behavior*, 47, 124–156. [232]

Vives, Xavier (1989), "Technological competition, uncertainty, and oligopoly." *Journal of Economic Theory*, 48, 386–415. [218]

Vives, Xavier (2002), "Private information, strategic behavior, and efficiency in Cournot markets." *RAND Journal of Economics*, 33, 361–376. [218]

Wiseman, Thomas (2005), "A partial Folk Theorem for games with unknown payoff distributions." *Econometrica*, 73, 629–645. [217, 220, 222, 235]

Yamamoto, Yuichi (2010), "Individual learning and belief-free equilibria in repeated games." Unpublished paper, Harvard University. [220, 221]

Submitted 2010-12-16. Final version accepted 2011-4-29. Available online 2011-4-29.