Folk Theorems with Bounded Recall under (Almost) Perfect Monitoring[*]

George J. Mailath[†]   Wojciech Olszewski[‡]

May 30, 2008

Abstract

A strategy profile in a repeated game has bounded recall L if play under the profile after two distinct histories that agree in the last L periods is equal. Mailath and Morris (2002, 2006) proved that any strict equilibrium in bounded-recall strategies of a game with full support public monitoring is robust to all perturbations of the monitoring structure towards private monitoring (the case of *almost-public monitoring*), while strict equilibria in unbounded-recall strategies are typically not robust.

We prove that the perfect-monitoring folk theorem continues to hold when attention is restricted to strategies with bounded recall and the equilibrium is essentially required to be strict. The general result uses calendar time in an integral way in the construction of the strategy profile. If the players' action spaces are sufficiently rich, then the strategy profile can be chosen to be independent of calendar time. Either result can then be used to prove a folk theorem for repeated games with almost-perfect almost-public monitoring. **Key Words:** Repeated games, bounded recall strategies, folk theorem, imperfect monitoring. **JEL codes:** C72, C73.

[†]University of Pennsylvania, `gmailath@econ.upenn.edu`

[‡]Northwestern University, `wo@northwestern.edu`

1 Introduction

The folk theorem is not positive economics. Like mechanism design, the folk theorem delineates which payoffs are consistent with intertemporal incentives (equilibrium behavior). Little is known about what restrictions on behavior are "reasonable," "plausible," or have any predictive power. Nonetheless, our understanding of the restrictions implied by intertemporal incentives till the mid 80's was based on explicitly constructed strategy profiles, such as Fudenberg and Maskin's (1986) folk theorem which used the "simple penal codes" of Abreu (1988).

Beginning with Abreu, Pearce, and Stacchetti (1986, 1990) and Fudenberg, Levine, and Maskin (1994), research in repeated games shifted its focus from the structure of repeated game strategies to that of the set of payoffs. However, recent work in private monitoring has returned the focus to the structure of behavior as well as payoffs.

Intertemporal incentives arise when public histories coordinate continuation play. But what if histories are private, but only barely so (i.e., are "almost" public)? Can we still provide intertemporal incentives through the coordination of continuation play? Is behavior robust to the introduction of private monitoring? This is a stronger question than whether the folk theorem is robust to the introduction of private information, which focuses on payoffs and not behavior. The weaker question allows the equilibria to depend on the details of the perturbed monitoring structure.

A strategy profile in a repeated game has L-bounded recall if play under the profile after two distinct histories that agree in the last L periods is equal. Mailath and Morris (2002, 2006) proved that any strict equilibrium in bounded-recall strategies of a game with full support public monitoring is robust to all perturbations of the monitoring structure towards private monitoring (the case of *almost-public monitoring*), while strict equilibria in unbounded-recall strategies are typically *not* robust. For some parameterizations of the imperfect public monitoring repeated prisoners' dilemma, Cole and Kocherlakota (2005) show that the set of PPE payoffs achievable by bounded recall strongly symmetric profiles is degenerate, while the set of strongly symmetric PPE payoffs is strictly larger.

We prove that the perfect-monitoring folk theorem continues to hold when attention is restricted to strategies with bounded recall and the equilibrium is essentially required to be strict. As observed by Mailath and Samuelson (2006), the proof of this result in Mailath and Morris (2002) is fundamentally flawed. The general result uses calendar time in an integral way in the construction of the strategy profile. If the players' action spaces

are sufficiently rich, then the strategy profile can be chosen to be independent of calendar time. Either result can then be used to prove a folk theorem for repeated games with almost-perfect almost-public privata monitoring.

Our interest in bounded-recall strategies arises primarily because of their role in determining the robustness of equilibria of repeated games to private monitoring. In contrast, the existing literature typically views bounded recall as a way of modeling bounded rationality. For example, Aumann (1981) mentions bounded recall and finite automata as two ways of modeling bounded rationality in the context of repeated games. A number of papers investigate the asymptotic behavior of the set of equilibrium payoff vectors in repeated games with no discounting, allowing the recall of all players to grow without bound. The characterization results typically resemble the folk theorem (see, for example, Lehrer (1994)). However, if the recalls of distinct players grow at distinct rates, the minimal payoffs depend on the relative rates of divergence across players. Players with long recall (who are sometimes called in this literature "strong") can correlate their own, or other players' actions in a manner that is concealed from some of their shorter-recall opponents. As a result, the payoffs of those "weaker" opponents fall below their minmax levels under independent actions (see, for example, Lehrer (1994), Gilboa and Schmeidler (1994), or Bavly and Neyman (2003)).

Assuming discounting and perfect monitoring, Barlo, Carmona, and Sabourian (forthcoming) establish the subgame-perfect folk theorem for games with rich action spaces by using strategies with one-period memory. The idea is that if the action spaces are sufficiently rich, players are able to encode entire histories in single stage-game actions. Hörner and Olszewski (2006b) establish the folk theorem in finite-recall strategies without the assumption that the action spaces are rich, and even under imperfect public monitoring.[1] They divide time horizon into blocks in which players play strategies similar to ones used by Fudenberg, Levine, and Maskin (1994), and design after each block a "communication phase" in which players encode the continuation payoff vector to be achieved in the following block. It is essential for the equilibria constructed by those authors to assume that the length of recall increases with the discount factor, and that players are indifferent between sending several distinct messages in the "communication phases." This indifference requires that the strategy profiles depend upon the fine details of the monitoring structure.

---

[1]Note, however, that Hörner and Olszewski (2006b) assume the existence of a public correlating device.

2 Preliminaries

In the stage game, player $i = 1, \ldots, n$ chooses action a_i from a finite set A_i. A profile of actions is a vector $a \in A = \prod_{i=1}^{n} A_i$. Player i's payoff from the action profile a is denoted $u_i(a)$, and the profile of payoffs $(u_1(a), \ldots, u_n(a))$ is denoted $u(a)$. For each i, player i's (pure action) minmax payoff $\underline{v}_i^p$ is given by

$$\underline{v}_i^p \equiv \min_{a_{-i}} \max_{a_i} u_i(a_i, a_{-i}) \equiv \max_{a_i} u_i(a_i, \hat{a}_{-i}^i) \equiv u_i(\hat{a}^i),$$

so that $\hat{a}^i$ is an action profile that minmaxes player i.[2] The payoff $\underline{v}_i^p$ is the lowest payoff that the other players can force on player i in the stage game (using pure actions). The set of stage game payoffs generated by pure action profiles is

$$\mathcal{F} \equiv \{v \in \mathbb{R}^n : \exists a \in A \text{ s.t. } g(a) = v\},$$

while the set of feasible payoffs is

$$\mathcal{F}^\dagger \equiv \text{co}\,\mathcal{F},$$

where $\text{co}\,\mathcal{F}$ is the convex hull of $\mathcal{F}$. Finally, the set of strictly (pure action) individually rational and feasible payoffs is

$$\mathcal{F}^{\dagger p} \equiv \{v \in \mathcal{F}^\dagger : v_i > \underline{v}_i^p, \quad \forall i\}.$$

We assume throughout that the interior of the set $\mathcal{F}^{\dagger p}$ is non-empty.

Given an n-dimensional vector v and a number $\varepsilon > 0$ we denote by $B_\varepsilon(v)$ the open ball of radius ε centered at v in $\mathbb{R}^n$, i.e., $B_\varepsilon(v) = \{v' \in \mathbb{R}^n : \|v - v'\| < \varepsilon\}$.

We begin with infinitely repeated games with *perfect monitoring*. In each period $t = 0, 1, \ldots$, the stage game is played, with the action profile chosen in period t publicly observed at the end of that period. The period t history is $h^t = (a^0, \ldots, a^{t-1}) \in A^t$, where a^s denotes the profile of actions taken in period s, and the set of histories is given by

$$\mathcal{H} = \cup_{t=0}^{\infty} A^t,$$

where we define the initial history to the null set $A^0 = \{\varnothing\}$. A strategy σ_i for player i is a function $\sigma_i : \mathcal{H} \rightarrow A_i$.

---

[2] If the minmaxing profile is not unique, choose one arbitrarily.

Every pure strategy profile has an automaton representation $(\mathcal{W}, w^0, f, \tau)$, where $\mathcal{W}$ is the set of states, w^0 is the initial state, $f : \mathcal{W} \to A$ is the output function, and $\tau : \mathcal{W} \times A \to \mathcal{W}$ is the transition function (Mailath and Samuelson, 2006, Section 2.3). For a fixed automaton $(\mathcal{W}, w^0, f, \tau)$, we say a state w is *accessible from* w^0 if there exists a history $h^t = (a^0, a^1, \ldots, a^{t-1})$ such that $w = \tau(w^0, h^t) \equiv \tau(\cdots \tau(\tau(w^0, a^0), a^1), \ldots, a^{t-1})$. Finally, we say that two states w and $\tilde{w}$ are *reachable in the same period* if there exists t and two histories h^t and $\tilde{h}^t$ such that $w = \tau(w^0, h^t)$ and $\tilde{w} = \tau(w^0, \tilde{h}^t)$.

Definition 1 *A strategy profile σ has* bounded recall of length L *(more simply, L-bounded recall) if for all $t \geq 0$, all $h^t, \hat{h}^t \in \mathcal{H}$, and all $h^L \in \mathcal{H}$,*

$$\sigma(h^t h^L) = \sigma(\hat{h}^t h^L). \tag{1}$$

We then have:[3]

Lemma 1 (Mailath and Morris (2006, Lemma 3)) *The strategy profile represented by the automaton $(\mathcal{W}, w^0, f, \tau)$ has L-bounded recall if and only if for all $w, w' \in \mathcal{W}$ reachable in the same period and for all histories $h^L \in \mathcal{H}$,*

$$\tau(w, h^L) = \tau(w', h^L). \tag{2}$$

Bounded recall strategies are potentially calendar time dependent. For example, a profile that plays a in even periods and $a' \neq a$ in odd periods irrespective of history has bounded recall (of zero length). Our constructions take advantage of the calendar time dependence allowed for in bounded recall strategies. A stronger condition (*bounded memory*) also requires calendar time independence, so that the histories h^t and $\hat{h}^t$ in (1) can be of different length (or equivalently, w and w' need not be reachable in the same period in (2)). The profile yielding alternating a and a' does not have bounded memory of *any* length.

Players share a common discount factor $\delta < 1$, and payoffs in the repeated game are evaluated as the average discounted value. Given a discount factor δ, we denote by $\mathcal{E}(\delta)$ the set of subgame-perfect equilibrium payoffs. The set of payoffs obtained in some subgame-perfect equilibrium in L-bounded recall strategies is denoted $\mathcal{E}^L(\delta)$. Given an automaton representation $(\mathcal{W}, w^0, f, \tau)$, denote player i's average discounted value from play that begins in state w by $V_i(w)$.

---

[3]See Mailath and Samuelson (2006, Lemma 13.3.1) for a proof.

Definition 2 *The strategy profile σ represented by the automaton $(\mathcal{W}, w^0, f, \tau)$ is a* strict subgame perfect equilibrium *if for all states w accessible from w^0, the action profile $f(w)$ is a strict Nash equilibrium of the normal form game described by the payoff function $g^w : A \to \mathbb{R}^n$, where*

$$g^w(a) = (1 - \delta)u(a) + \delta V(\tau(w, a)). \tag{3}$$

The profile is patiently strict *if there exists $\varepsilon > 0$ and $\bar{\delta} < 1$ such that for all w, i, and $a_i \neq f_i(w)$, and all $\delta \in (\bar{\delta}, 1)$,*

$$\frac{g_i^w(f(w)) - g_i^w(a_i, f_{-i}(w))}{1 - \delta} > \varepsilon.$$

Equivalently, σ is a strict subgame perfect equilibrium if every one-shot deviation is *strictly* suboptimal. If $f(w)$ were simply required to be a (possibly non-strict) Nash equilibrium of the game g^w for all accessible w, then we would have subgame perfection (Mailath and Samuelson, 2006, Proposition 2.4.1). We caution the reader that this use of *strict* is a slight abuse of language, since player i is indifferent between σ_i and any deviation from σ_i that leaves the outcome path unchanged. This use is motivated by its use in public monitoring games (see footnote 14 and Mailath and Samuelson (2006, Definition 7.1.3)).

Definition 3 *The strategy profile σ represented by the automaton $(\mathcal{W}, w^0, f, \tau)$ is a* pseudo-strict subgame perfect equilibrium *if for all states w accessible from w^0, the action profile $f(w)$ is a Nash equilibrium of the normal form game described by the payoff function $g^w : A \to \mathbb{R}^n$ given in (3), and if for all $a_i \neq f_i(w)$ for some w accessible from w^0 satisfying*

$$g_i^w(f(w)) = g_i^w(a_i, f_{-i}(w)), \tag{4}$$

we have

$$\tau(w, f(w)) = \tau(w, (a_i, f_{-i}(w))). \tag{5}$$

The profile is patiently pseudo-strict *if there exists $\varepsilon > 0$ and $\bar{\delta} < 1$ such that for all w, i, and $a_i \neq f_i(w)$ for which (4) fails, and all $\delta \in (\bar{\delta}, 1)$,*

$$\frac{g_i^w(f(w)) - g_i^w(a_i, f_{-i}(w))}{1 - \delta} > \varepsilon.$$

The set of payoffs obtained in some patiently pseudo-strict subgame perfect equilibrium in L-bounded recall strategies is denoted $\mathcal{E}^{L,s}(\delta)$.

If $a_i \neq f_i(w)$ satisfies (4), then from (5), player i's stage game payoffs from $f(w)$ and $(a_i, f_{-i}(w))$ are equal. That is, one-shot deviations that are not strictly suboptimal yield the same stage-game payoffs and imply the same continuation play.

We are interested in bounded recall strategies because pseudo-strict perfect public equilibrium of a repeated game with public monitoring in bounded recall strategies is robust to private monitoring, and essentially only such strict PPE are robust (Mailath and Morris, 2002, 2006). Once we have a perfect monitoring folk theorem in bounded recall strategies (and pseudo-strict equilibria), this immediately yields a folk theorem for almost-perfect highly correlated private monitoring games. Patient pseudo-strictness guarantees that the degree of approximation of the pubic and private monitoring games is independent of the degree of patience (see Section 6).

3   Two Players and an Easy First Theorem

We first discuss the two player case. The proof of Mailath and Samuelson (2006, Proposition 13.6.1) immediately implies the following two-player bounded-recall *pure-action* folk theorem:

Theorem 1 *Suppose $n = 2$. For all strictly individually rational $a \in A$ (i.e., $u(a) \in \mathcal{F}^{\dagger p}$), there exists $\bar{\delta} \in (0, 1)$ and $L < \infty$ such that for all $\delta \in (\bar{\delta}, 1)$, there is a patiently strict subgame perfect equilibrium in L-bounded recall (in fact, bounded memory) strategies with outcome a in every period.*

From the proof of Theorem 1 in Fudenberg and Maskin (1986), for any strictly individually rational a, there exists $\bar{\delta} \in (0, 1)$ and L such that, for all $\delta \in (\bar{\delta}, 1)$, mutual minmax $(\hat{a}_1^1, \hat{a}_2^2)$ for L periods followed by a return to a is sufficient to deter any unilateral deviation from both a and mutual minmax. A bounded memory profile with the identical unilateral incentives specifies a if in the last L periods, either a was played in every period, or mutual minmax was played in every period, and mutual minmax otherwise.

The critical observation is that because a is strictly individually rational, every unilateral deviation from a and from mutual minmax results in a profile distinct from both a and mutual minmax (i.e., is "immediately detected"). Hence, such a deviation can be met with L periods of mutual minmax without recourse to information from earlier periods.

In order to cover all strictly individually rational payoffs, it is natural to consider replacing a in the strategy profile with a cycle of action profiles whose average payoff approximates the target payoff. Since we need no

6

longer have the property of "immediate detection" of unilateral deviations, however, such a replacement need not yield an equilibrium. Consequently, there is no simple extension of Theorem 1 to cover all strictly individually rational payoffs. Consider the repeated prisoners' dilemma (with actions C and D, and mutual minmax DD), and a payoff obtained from the convex combination $\frac{1}{2} \circ CD + \frac{1}{2} \circ DC$. Suppose $L = 2$ is sufficient to strictly deter a deviation. The cycle $CD, DC, CD, \dots$ achieves the desired payoff, and requires only a two period memory to implement. On the cycle, a deviation by player 1 to D in some period gives a history ending in DC, DD, which should be followed by two periods of DD, with the cycle only resuming once three periods of DD have occurred.[4]

Consider a profile with 3-bounded recall and a history ending in DC, DD, CD. While the profile specifies DD (since the last 3 periods are inconsistent with the cycle, and so mutual minmax should sill be played), player 2 will optimally unilaterally deviate to C, giving a history whose last 3 periods are DD, CD, DC, and so the cycle resumes. This failure of the profile at histories ending in DC, DD, CD causes the profile to unravel, since such a history is obtained by player 1 unilaterally deviating after a history ending in CD, DC, DD to C, rather than playing D, which in turn is obtained by a unilateral deviation by player 1 on the cycle (at a history ending in DC, CD, DC).

The problem identified in the previous paragraph is a familiar one. It underlies, for example, the anti-folk theorem in the overlapping generations model of Bhaskar (1998). However, bounded recall gives us some flexibility in the design of the profile. In particular, we can specify a priori certain periods as *announcement* periods which the players use to "announce" whether play should follow the cycle, or punish. Such a "trick" allows us to easily obtain a partial folk theorem, and for arbitrary number of players. Specifically, any feasible payoff that Pareto dominates a static Nash equilibrium can be supported in bounded recall strategies. (Recall that mutual minmax is a Nash equilibrium in the prisoners' dilemma.)

Theorem 2 *Suppose a^N is a strict Nash equilibrium of the stage game. For all $v \in \mathcal{F}^\dagger$ with $v_i > u_i(a^N)$ and for all $\varepsilon > 0$, there exists $\bar{\delta} \in (0, 1)$ and $L < \infty$ such that for all $\delta \in (\bar{\delta}, 1)$, there is a patiently strict subgame perfect equilibrium in L-bounded recall strategies with discounted average payoff within ε of v.*

---

[4]The "three" comes from the deviation (one period) and two periods of mutual minmax punishment. If the cycle resumes after two periods of DD, then the punishment is only of length 1.

Proof. Let $\tilde{a} \in A$ be an action profile satisfying $\tilde{a}_i \neq a_i^N$ for at least two players. For T sufficiently large, there is a cycle of actions $h^T \equiv (a^1, \ldots, a^{T-1}, \tilde{a}) \in A^T$ whose average payoff is within $\varepsilon/2$ of v. As a matter of notation, $a^T = \tilde{a}$.

Consider the automaton with states $\{w(k, \ell) : k \in \{0, 1\}, \ell \in \{1, \ldots, T\}\}$, initial period $w^0 = w(0, 1)$, output function $f(w(0, \ell)) = a^\ell$ and $f(w(1, \ell)) = a^N$ for all ℓ, and transition function,

$$\tau(w(0, \ell), a) = \begin{cases} w(0, \ell+1), & \text{if } \ell \leq T-1 \text{ and } a = a^\ell, \\ w(1, \ell+1), & \text{if } \ell \leq T-1 \text{ and } a \neq a^\ell, \\ w(0, 1), & \text{if } \ell = T \text{ and } a = \tilde{a}, \\ w(1, 1), & \text{if } \ell = T \text{ and } a \neq \tilde{a}, \end{cases}$$

and

$$\tau(w(1, \ell), a) = \begin{cases} w(1, \ell+1), & \text{if } \ell \leq T-1, \\ w(0, 1), & \text{if } \ell = T \text{ and } a = \tilde{a}, \text{ and} \\ w(1, 1), & \text{if } \ell = T \text{ and } a \neq \tilde{a}. \end{cases}$$

Under the automaton, the cycle h^T is played every T periods. The automaton has T-bounded recall, because in any period $t = kT$ for any integer k, the automaton is in either state $w(0, T)$ or in state $w(1, T)$, and the transition in both cases are the same. Period $t = kT$ is an announcement period: The cycle is played if, and only if, both players choose their part of $\tilde{a}$; the profile specifies that $\tilde{a}$ is played in that period only if the cycle had been played in the previous T periods (beginning in period $kT - 1$).

Note that a player i cannot unilaterally prevent Nash reversion, since the other players will choose a_{-i}^N in the announcement period. Finally, patient strictness is immediate from Nash reversion for δ sufficiently high. ∎

Remark 1 The approximation in Theorem 2 is due to the announcement period, and not to the use of a cycle per se. Fix $\varepsilon > 0$ sufficiently small that $v_i > v_i^N + 2\varepsilon$ for all i. From Fudenberg and Maskin (1991) (or see Mailath and Samuelson (2006, Lemma 3.7.2)), there is an infinite outcome path $h \in A^\infty$ with value v and whose discounted average payoff at any time t is within ε of v. Suppose that in every period in h, both players choose an action distinct from their Nash action a_i^N. Then, specifying the period t action profile in h if the period $t - 1$ action profile was played and a^N otherwise trivially yields a patiently strict subgame equilibrium in 1-bounded recall strategies. However, in general, the outcome path will not

satisfy such a condition, nor need the action profiles at any regular dates $t = kT$ satisfy such a condition. ∎

Remark 2 At a slight cost of complexity, Nash reversion can be replaced by, after a deviation, play Nash till the second announcement period, followed by a resumption of the cycle h^T. Not only does the formal description of the automaton become a little complicated, but care must now be taken in the specification of the cycle that the continuation value never drop too low (which can be handled similarly to Fudenberg and Maskin (1991)). Mutual minmax cannot be used in place of the Nash equilibrium to obtain a full folk theorem for two players, since incentives must then be provided to the players to carry out the mutual minmax, and this appears impossible when we restrict the punishment triggers to occur every T periods.

Interestingly, it appears to be easier to provide such incentives when there is more flexibility in the specification of the end date of the punishment. Using that freedom, Barlo, Carmona, and Sabourian (2008) prove a folk theorem for two player games.[5] While the profile has bounded memory (and not just bounded recall), the construction is considerably more complicated than that of our proof of Theorem 2. ∎

4 Player-Specific Punishments

With more than two players, some notion of player-specific punishments is needed to obtain a folk theorem.

Definition 4 *A payoff v allows player-specific punishments if there exists a collection $\{v^i\}_{i=1}^n$ of payoff vectors $v^i \in \mathcal{F}^{\dagger p}$, such that*

$$v_i > v_i^i \quad and \quad v_i^j > v_i^i, \quad \forall j \neq i.$$

A payoff v allows pure-action player-specific punishments if $v = u(a(0))$ for some $a(0) \in A$, and $v^j = u(a(j))$ for some $a(j) \in A$ and all $j = 1, \ldots, n$.

---

[5]That paper also proves a folk theorem for more than two players, under some "confusion-proof" conditions that are related to the immediate detection property we discussed above. We only learned of Barlo, Carmona, and Sabourian (2008) when the final draft of this paper was being prepared.

Suppose payoffs $v^0 = u(a(0))$ allow pure-action player-specific punishments. The standard construction of a subgame perfect profile with payoffs v^0 is to use a simple strategy profile. In this profile, any unilateral deviation (by player i say) results in the deviator being minmaxed by $\hat{a}^i$ for a finite number of periods, after which a^i is played (unless there is a further unilateral deviation), and multilateral deviations are ignored.[6] The profile's automaton description has the set of states

$$\widetilde{\mathcal{W}} = \{w(d) : 0 \le d \le n\} \cup \{w(i,t) : 1 \le i \le n,\ 0 \le t \le L-1\},$$

initial state $w^0 = w(0)$, output function $\widetilde{f}(w(d)) = a(d)$, and $\widetilde{f}(w(i,t)) = \hat{a}^i$ for $0 \le t \le L-1$, and transition function

$$\widetilde{\tau}(w(d), a) = \begin{cases} w(j,0), & \text{if } a_j \ne a_j(d),\ a_{-j} = a_{-j}(d), \\ w(d), & \text{otherwise,} \end{cases}$$

and

$$\widetilde{\tau}(w(i,t), a) = \begin{cases} w(j,0), & \text{if } a_j \ne \hat{a}^i_j,\ a_{-j} = \hat{a}^i_{-j} \text{ for } j \ne i, \\ w(i,t+1), & \text{otherwise,} \end{cases}$$

where $w(i,L) \equiv w(i)$. We denote this automaton by $\widetilde{\mathcal{A}}$.

Lemma 2 *For L sufficiently large and δ sufficiently close to 1, the profile induced by $\widetilde{\mathcal{A}}$ is a patiently pseudo-strict subgame perfect equilibrium.*

Proof. The proof of Mailath and Samuelson (2006, Proposition 3.4.1, statement 1) shows that $\widetilde{\mathcal{A}}$ describes a pseudo-strict subgame perfect equilibrium for sufficiently large L and δ close to 1. If either i has a unique best reply to $\hat{a}^i_{-i}$ or the profile restarts the punishment of i after i deviates in the state $w(i,t)$, then the induced equilibrium is strict.

It remains to verify patient pseudo-strictness. We denote player i's average discounted value from play beginning in state w under $\widetilde{\mathcal{A}}$ by $\widetilde{V}_i(w)$. Consider first the normal form game with payoffs $\widetilde{g}^w$ from (3), i.e.,

$$\widetilde{g}^w(a) = (1-\delta)u(a) + \delta\widetilde{V}(\widetilde{\tau}(w,a)),$$

---

[6]Since player i has no myopic incentive to deviate from a *fixed* $\hat{a}^i$, it is obviously not necessary to restart i's minmaxing cycle after a further unilateral deviation by i. On the other hand, in most settings it does no harm, and many presentations (such as Mailath and Samuelson (2006)) restart the punishment in this case as well. However, as explained in footnote 8, it is easier to ignore i's deviations from $\hat{a}^i$ in the subsequent development, and so we do so here.

for $w = w(i,t)$, and a deviation by player $j \neq i$. We have, for $a_j \neq \widetilde{f}_j(w)$,

$$\widetilde{g}_j^w(\hat{a}^i) - \widetilde{g}_j^w(a_j, \hat{a}_{-j}^i) = (1-\delta)(u_j(\hat{a}^i) - u_j(a_j, \hat{a}_{-j}^i)$$
$$+ \delta(\widetilde{V}_j(w(i,t+1)) - \widetilde{V}_j(w(j,0))).$$

But

$$\widetilde{V}_j(w(i,t+1)) - \widetilde{V}_j(w(j,0)) = (1-\delta^{L-t})(u_j(\hat{a}^i) - \underline{v}_j^p)$$
$$+ (\delta^{L-t} - \delta^L)(v_j^i - \underline{v}_j^p) + \delta^L(v_j^i - v_j^j),$$

which is strictly positive for large δ, and bounded away from zero as $\delta \to 1$ (since $u_j(a(i)) > u_j(a(j))$). Suppose $\varepsilon > 0$ is sufficiently small that $2\varepsilon < v_j^i - v_j^j$. Then, for δ sufficiently large, for all $a_j \neq f_j(w)$,

$$\widetilde{g}_j^w(\widetilde{f}(w)) - \widetilde{g}_j^w(a_j, \widetilde{f}_{-j}(w)) > \varepsilon \tag{6}$$

for $w = w(i,t)$ and $j \neq i$. Since $\widetilde{V}(w(d)) = v^d$, this incentive constraint reflects the loss of long-run value of at least $v_j^i - v_j^j$.

A similar calculation shows that (6) also holds for $w = w(d)$ and $d \neq j$.

We do not need to check player j's incentives at $w(j,t)$, since j is myopically optimizing and the transition is independent of j's action.

Finally, for $w = w(j)$, for player j we have

$$\frac{\widetilde{V}_j(w(j)) - \widetilde{V}_j(w(j,0))}{(1-\delta)} = \frac{(1-\delta^L)(v_j^j - v_j^p)}{(1-\delta)}$$
$$\to L(v_j^j - v_j^p) \qquad \text{as} \qquad \delta \to 1.$$

If L satisfies

$$\max_a u_i(a) - v_i^i + 2\varepsilon < L(v_i^i - \underline{v}_i^p) \tag{7}$$

for all i, then for δ sufficiently large, for all $a_j \neq f_j(w(j))$,

$$\frac{\widetilde{g}_j^w(\widetilde{f}(w)) - \widetilde{g}_j^w(a_j, \widetilde{f}_{-j}(w))}{1-\delta} > \varepsilon.$$

Note that this last constraint only reflects a change of behavior in a finite number of periods, since long-run value of v_i^i is unchanged by the deviation. ∎

Because multilateral deviations are ignored, $\widetilde{\mathcal{A}}$ typically does not have bounded recall: For example, a unilateral deviation by i in $w(0)$ eventually

11

leads to $w(i)$, and so $a(i)$, while a multilateral deviation from $a(0)$ to $a(i)$ in every period keeps the automaton in $w(0)$. Potentially more of a problem is that a unilateral deviation by i in state $w(0)$ may yield the same action profile as a unilateral deviation by j in $w(k)$.

For a fixed automaton $(\mathcal{W}, w^0, f, \tau)$, we can view the set of action profiles $\{f(w) : w \in \mathcal{W}\}$ as the set of *intended* profiles. Say that a unilateral deviation by i at w is *immediately detectable* if a_{-i} uniquely identifies the action profile $f(w)$, independent of a_i. In such a case, when $a_i \neq f_i(w)$, we should be able to treat a as the result of a unilateral deviation by i. If the action space is sufficiently rich that all unilateral deviations at every state are immediately detectable, then we can modify the profile in its treatment of multilateral deviations to obtain bounded recall. Note that this is a little delicate, since for example the action profile $(a_i, a_{-i}(j)) \neq a(j)$ is both a unilateral deviation from $a(j)$, as well as a potentially multilateral deviation from $a(k)$, and so must be treated differently than some other multilateral deviation.

Lemma 3 *Suppose $n \geq 3$, $v^0 = u(a(0))$ for some $a(0) \in A$, and that v^0 allows pure action player specific punishments (so that $v^j = u(a(j))$ for some $a(j)$ and for all $j = 1, \ldots, n$). Suppose moreover, that the action profiles $\{a(d) : d = 0, \ldots, n\}$ with $\{\hat{a}^i : i = 1, \ldots, n\}$ are all distinct, player by player, that is, for all $j = 1, \ldots, n$,*

$$\left| \{a_j(d) : d = 0, \ldots, n\} \cup \{\hat{a}^i_j : i = 1, ..., n\} \right| = 2n + 1. \qquad (8)$$

Then there exists $\bar{\delta} < 1$, $L < \infty$, and an L-bounded-memory strategy profile with outcome path $a(0)$ in every period, such that for $\delta \in (\bar{\delta}, 1)$, the L-bounded-memory strategy profile is a patiently pseudo-strict subgame perfect equilibrium.

Proof. Since there are three or more players, (8) implies that every unilateral deviation from an action profile in $\{a(d) : d = 0, \ldots, n\} \cup \{\hat{a}^i : i = 1, \ldots, n\}$ is immediately detectable (in the sense described just before the statement of the lemma). This allows us to define the transitions so that apart from action profiles that minmax a player, the automaton has one-period recall.

As for $\widetilde{\mathcal{A}}$, choose L sufficiently large that (7) is satisfied. The new automaton has set of states

$$\mathcal{W} = \widetilde{\mathcal{W}} \cup \{w(i, L) : 1 \leq i \leq n\},$$

initial state $w^0 = w(0)$, an output function that agrees with $\widetilde{f}$ on $\widetilde{\mathcal{W}}$ and specifies $f(w(i, L)) = a(i)$, and finally, transition function

$$
\tau(w(d), a) = \begin{cases}
w(j, 0), & \text{if } a_j \neq a_j(d), \ a_{-j} = a_{-j}(d) \\
& \qquad \text{or } a_j \neq \hat{a}^k_j, \ a_{-j} = \hat{a}^k_{-j} \text{ some } j \neq k, \\
w(j, 1), & \text{if } a_{-j} = \hat{a}^j_{-j}, \\
w(d), & \text{if } a = a(d), \\
w(0), & \text{otherwise,}
\end{cases}
$$

$$
\tau(w(i, L), a) = \begin{cases}
w(j, 0), & \text{if } a_j \neq a_j(d), \ a_{-j} = a_{-j}(d) \\
& \qquad \text{or } a_j \neq \hat{a}^k_j, \ a_{-j} = \hat{a}^k_{-j} \text{ some } j \neq k, \\
w(j, 1), & \text{if } a_{-j} = \hat{a}^j_{-j}, \ j \neq i \\
w(i, L), & \text{if } a_{-i} = \hat{a}^i_{-i}, \\
w(d), & \text{if } a = a(d), \\
w(0), & \text{otherwise,}
\end{cases}
$$

and, finally, for $t \leq L - 1$,

$$
\tau(w(i, t), a) = \begin{cases}
w(j, 0), & \text{if } a_j \neq a_j(d), \ a_{-j} = a_{-j}(d) \\
& \qquad \text{or if } a_j \neq \hat{a}^k_j, \ a_{-j} = \hat{a}^k_{-j} \text{ some } j \neq k, \\
w(j, 1), & \text{if } a_{-j} = \hat{a}^j_{-j}, \ j \neq i \\
w(i, t+1), & \text{if } a_{-i} = \hat{a}^i_{-i}, \\
w(d), & \text{if } a = a(d), \\
w(0), & \text{otherwise.}
\end{cases}
$$

The verification that the automaton has bounded recall is straightforward. As we indicated before describing the automaton, except for action profiles satisfying $a_{-i} = \hat{a}^i_{-i}$ for some i, the automaton has one-period recall: Irrespective of the current state, after the action profile $a(d)$, the automaton immediately transits to the state $w(d)$; after a unilateral deviation by j from $a(d)$ or from $\hat{a}^k$, $k \neq j$, the automaton immediately transits to the state $w(j, 0)$; and after any other profile satisfying $a_{-i} \neq \hat{a}^i_{-i}$ for all i, the automaton immediately transits to the state $w(0)$. Finally, after an action profile satisfying $a_{-i} = \hat{a}^i_{-i}$ for some i, the automaton transits to a state $w(i, t)$, with the value of t determined by the previous state. Subsequent $a_{-i} = \hat{a}^i_{-i}$ increment the counter t, till $t = L$.

Consider now a T-length history, with a^T being the last period action profile. If $a^T_{-i} \neq \hat{a}^i_{-i}$ for all i, then the current state is determined from the

previous paragraph. Suppose now that there is some i for which $a_{-i}^L \neq \hat{a}_{-i}^i$, and let $\ell = \max\{t : a_{-i}^t \neq \hat{a}_{-i}^i\}$; note that $\ell < T$. Then, the current state of the automaton is given by $w(i, t')$, where $t' = \min\{\ell, L\}$. Thus, action profiles in the history more than L periods in the past are irrelevant, and the automaton has L bounded recall.

Finally, since the new automaton induces the same initial outcome path as $\widetilde{\mathcal{A}}$, as well as inducing the same outcome path after any unilateral deviation as $\widetilde{\mathcal{A}}$, it is patiently pseudo-strict. ∎

Remark 3 The richness condition (8) is stronger than necessary. It is of course enough that every unilateral deviation from an action profile in $\{a(d) : d = 0, \ldots, n\} \cup \{\hat{a}^i : i = 1, \ldots, n\}$ be immediately detectable (in the sense described just before the statement of the lemma).

A natural conjecture is that immediate detection condition is in fact unnecessary. Consider the repeated "epoch" game, where an *epoch* is a block of T periods. Since, by choosing T sufficiently large, we can guarantee that the immediate detection condition holds for appropriately specified T length cycles, we can apply the construction in Lemma 3 to the repeated "epoch" game. The flaw in this argument is that the resulting profile may not be an equilibrium. In particular, consider the following possibility: A player unilaterally deviates in the first period of an epoch, followed by another unilateral deviation (by either the same or a different player). From the epoch viewpoint, this is akin to a multilateral deviation and so is effectively ignored by the profile. Consequently, the construction in Lemma 3 does not imply that such deviations are suboptimal.

Similarly, while it is possible to relax the assumption that v^0 and the associated player-specific punishments can be implemented in single pure action profiles, the immediate detection condition becomes more demanding, since all unilateral deviations must be detected immediately (as before). ∎

5 Perfect Monitoring Folk Theorem under Bounded Recall

In this section, we prove a general perfect monitoring folk theorem under bounded recall.

Theorem 3 *Suppose $n \geq 3$ and $\mathcal{F}^{\dagger p}$ has nonempty interior. For all $v \in \mathcal{F}^{\dagger p}$, for all $\varepsilon > 0$, there exists $\bar{\delta} < 1$ and $L < \infty$, such that for all $\delta \in (\bar{\delta}, 1)$,*

$$B_\varepsilon(v) \cap \mathcal{E}^{L,s}(\delta) \neq \varnothing.$$

We prove the pure action version first, which requires a slightly stronger form of player-specific punishments:

Definition 5 *A payoff v allows strong player-specific punishments if there exists a collection $\{v^i\}_{i=1}^n$ of payoff vectors $v^i \in \mathcal{F}^{\dagger p}$, such that*

$$v_i^j > v_i > v_i^i, \quad \forall j \neq i. \tag{9}$$

A payoff v allows pure-action strong player-specific punishments if $v = u(a(0))$ for some $a(0) \in A$, and $v^j = u(a(j))$ for some $a(j) \in A$ and all $j = 1, \ldots, n$.

Since $\mathcal{F}^{\dagger p}$ has nonempty interior, as usual, any $v^0 \in \text{int}\mathcal{F}^{\dagger p}$ allows strong player-specific punishments, though typically not in pure actions.

Lemma 4 *Suppose $n \geq 3$ and v allows strong pure-action player-specific punishments. Then the conclusion of Theorem 3 holds.*

Proof. In order to deal with the issues raised in remark 3, and similar to the proof of Theorem 2, we modify $\widetilde{\mathcal{A}}$ by introducing an *announcement* phase of length $2n + 1$ that occurs every $T > 2n + 1 + L$ periods.[7] In the announcement phase, the players effectively announce the new initial state for the automaton from $\mathcal{W}^* \equiv \{w(d) : d = 0, \ldots, n\} \cup \{w(i,0) : i = 1, \ldots, n\}$, and then in the following *normal* phase, play according to the automaton with that announced initial state. (This use of *state* will be justified in the next paragraph.) At the end of the normal phase, a new state has been determined (according to the above transition function), which is then announced in the next announcement phase (with $w(i,0)$ announced if the state reached is $w(i,t)$ for any $t = 0, \ldots, L - 1$).[8] We will show that this profile has bounded recall of length $T + (2n + 1)$.

The set of states in the modified automaton $\widehat{\mathcal{A}} \equiv (\widehat{\mathcal{W}}, \widehat{w}^0, \widehat{f}, \widehat{\tau})$ are $\widehat{\mathcal{W}} \equiv \mathcal{W} \times \{1, \ldots, T\}$, with initial state $\widehat{w}^0 = (w(0), 1)$. For states (w, r) with

---

[7] While the idea of using an announcement phase to announce states was inspired by Hörner and Olszewski (2006b), the details of the announcement phase are very different, reflecting our need to obtain pseudo-strict incentives everywhere.

[8] Since the announcement phase does not distinguish between $w(i,0)$ and $w(i,t)$ for $t > 0$, the underlying profile needs to ignore deviations by i from $\hat{a}^i$; see footnote 6.

$r = 1, \ldots, T - 2n - 1$, the specified behavior agrees with that under our earlier automaton, so that $\widehat{f}(w, r) = \widetilde{f}(w)$. The transitions are given by

$$
\widehat{\tau}((w, r), a) = \begin{cases} (\widetilde{\tau}(w, a), r + 1), & \text{if } r \leq T - 2n - 2, \text{ or if } r = T - 2n - 1 \\ & \text{and } \widetilde{\tau}(w, a) \neq w(i, t) \text{ for any } i \text{ and } t, \\ (w(i, 0), T - 2n), & \text{if } r = T - 2n - 1 \text{ and } \widetilde{\tau}(w, a) = w(i, t) \\ & \text{for some } i \text{ and } t. \end{cases}
$$

That is, within the normal phase, the modified automaton behaves as the original and counts down the periods. We will often refer to w as a state or as the *current state*, with the index r implicit. In particular, $w \in \mathcal{W}^*$ is a potential initial state, and $(w, 1)$ is an initial state for the normal phase.

Behavior under $\widehat{\mathcal{A}}$ will have the property that if $(w, T - 2n)$ is the state at the beginning of the announcement phase, then $(w, 1)$ is the state at the beginning of the next normal phase.

Due to the introduction of the announcement phase, a larger value of L is needed: For fixed $\varepsilon > 0$, choose L sufficiently large that, for all i,

$$
(2n + 2)(\max_a u_i(a) - \min_a u_i(a)) + 3\varepsilon < L(v_i^i - \underline{v}_i^p). \tag{10}
$$

Claim 1 (Incentives in the normal phase) *Suppose T satisfies,*

$$
\frac{2n + 1}{T} \left[\max_a u_i(a) - \min_a u_i(a) \right] < \varepsilon/4, \qquad \forall i. \tag{11}
$$

Let $\widehat{g}_i^{(w,r)}$ denote the payoffs of the normal form game (3), i.e.,

$$
\widehat{g}^{(w,r)}(a) = (1 - \delta)u(a) + \delta \widehat{V}(\widehat{\tau}((w, r), a)),
$$

where $\widehat{V}_i(w, r)$ is player i's payoff from state (w, r) under $\widehat{\mathcal{A}}$. There exists $\bar{\delta} \in (0, 1)$ such that for all $\delta \in (\bar{\delta}, 1)$, $w \in \mathcal{W}$, $r \in \{1, 2, \ldots, T - 2n - 1\}$, j, and $a_j \neq \widehat{f}_j(w, r)$, if $w \neq w(j, t)$ for any t, then,

$$
\frac{\widehat{g}_j^{(w,r)}(\widehat{f}(w, r)) - \widehat{g}_j^{(w,r)}(a_j, \widehat{f}_{-j}(w, r))}{1 - \delta} \geq 2\varepsilon.
$$

If $w = w(j, t)$ for some t, then

$$
\widehat{\tau}((w, r), \widehat{f}(w, r)) = \widehat{\tau}((w, r), (a_j, \widehat{f}_{-j}(w, r))).
$$

16

Proof. Observe first that for a sufficiently patient player i, $\widehat{V}_i(w, r)$, is within $\varepsilon/2$ of $\widetilde{V}_i(w)$ for all $w \in \mathcal{W}^*$ (recall that $\widetilde{V}_i(w)$ denotes player i's average discounted value from play beginning in state w under $\widetilde{\mathcal{A}}$). Then, it is immediate from (6) that for large δ,[9] for all $a_j \neq \widehat{f}_j(w, r)$,

$$\widehat{g}_j^{(w,r)}(\widehat{f}(w, r)) - \widehat{g}_j^{(w,r)}(a_j, \widehat{f}_{-j}(w, r)) > \varepsilon$$

for $w = w(i, t)$ and $j \neq i$, and for $w = w(d)$ and $d \neq j$.

Finally, for $w = w(j)$, for player j, the incentive to deviate can be bounded by noting that deviating can contribute at most $2n + 2$ periods of benefit (the current period, plus the impact on the announcement phase), and so

$$\widehat{g}_j^{(w,r)}(a(j)) - \widehat{g}_j^{(w,r)}(a_j, a_{-j}(j)) \geq$$
$$(1 - \delta^{2n+2})(\min_a u_j(a) - \max_a u_j(a))$$
$$+ \delta^{2n+2}(1 - \delta^L)(v_j^j - \underline{v}_j^p)),$$

and so

$$\lim_{\delta \to 1} \frac{\widehat{g}_j^{(w,r)}(a(j)) - \widehat{g}_j^{(w,r)}(a_j, a_{-j}(j))}{1 - \delta} \geq$$
$$(2n + 2)(\min_a u_j(a) - \max_a u_j(a)) + L(v_j^j - \underline{v}_j^p)),$$

which exceeds 3ε from (10). ∎

Each period $r = T - 2n, T - 2n + 1, \ldots, T$ of the announcement phase corresponds to one of the potential initial states, $w \in \mathcal{W}^*$ in order $w(0), w(1), \ldots, w(n), w(1, 0), \ldots, w(n, 0)$.[10] For the first $n + 1$ periods and for each player i, we arbitrarily identify one action in A_i as YES (or Y), and another action as NO (N), with the remaining actions having no meaning. For each i, in the period corresponding to the state $w(i, 0)$, and for each player $j \neq i$, we again arbitrarily identify one action a_j^Y in A_j as YES, and another action

---

[9]The bound on δ is tighter than that yielding (6) since, in states $w(i, t)$, players may minmax i for $2L - 1$ periods. This occurs if the L periods of minmaxing i do not end before the end of the normal phase.

[10]Once $2n$ of the states have been rejected, the remaining state is effectively announced by default, and so announcing all $2n + 1$ states is redundant, but simplifies the description.

a_j^N as NO; for player i we identify every stage-game best reply for i to the profile a_{-i}^Y as YES, and some other action as NO.

At the beginning of the announcement phase, there is a new current state w corresponding to the state $(w, T - 2n - 1)$ (resulting from the previous $T - 2n - 1$ periods). The $2n + 1$ periods of the announcement phase encode w as players use the actions identified as YES and NO to announce w, with deviations triggering appropriate continuations. There is a subtlety however, since when $w = w(i, 0)$, player i cannot be disciplined in the current announcement phase and so will myopically best reply to the behavior of the other players. Consequently, we cannot rely on unanimous announcements to announce the state and we sometimes use a plurality of $n - 1$ YES's or $n - 1$ NO's to indicate the status of a state. The strengthening of player-specific punishments allow us to deter any deviation leading to only $n - 2$ YES's or $n - 2$ NO's by using v^0 as a continuation value after such a deviation.

In each period of the announcement phase, given a current state players are prescribed to answer truthfully whether the current state is the state corresponding to that period. If the current state is $w(i, 0)$, then in the period corresponding to that state, player i may have multiple myopic best replies and so more than one action identified as YES. We will not resolve this ambiguity; this does not affect any incentives since continuation play is independent of the player i's action in that period.

It remains to describe how the current state is determined, and verify the bounded recall nature of its determination. We begin with the easy case of more than 3 players. We first classify all action profiles into the six (possibly non-exclusive) classes listed in Figure 1.

Starting with the initial period of the announcement phase, call the initial state the current state. Given a current state, and an action profile for that period, we identify a new state. For example, if in a period, all players announce YES, the new state is the state corresponding to that period, independent of the current state. Similarly, if only one player does not announce YES in that period, the new state is the punishment state for that player. If the action profile is in class 5, then the new state is the current state, unless this is the last period of the announcement phase *and* class 5 profiles have been observed in every period of the announcement phase, in which case the new state is $w(0)$. At the end of the announcement phase, the state is the *announced* state.

Note that every $2n + 1$ sequence of action profiles in the announcement phases leads to an announced state *independent* of the initial state.

Claim 2 (Incentives in the announcement phase) *Suppose*

class	profile	ending state if current state independent
1	$|\{a_i = Y\}| = n$	announcement period
2	$|\{a_i = Y\}| = n - 1,\ a_j \neq Y$	$w(j, 0)$
3	$|\{a_i = N\}| = n - 2$	$w(0)$
4	$|\{a_i = N\}| = n - 1,\ a_j \neq N$	$w(j, 0)$
5	not in the above classes	

Figure 1: The classes of action profiles when $n \geq 4$. In class 1, the new state is the state corresponding to the current period. If the action profile is in class 5, then the new state is the current state, unless this is the last period of the announcement phase and class 5 profiles have been observed in every period of the announcement phase, in which case the new state is $w(0)$.

$n \geq 4$ and T satisfies (11). There exists $\bar{\delta} \in (0, 1)$ such that for all $\delta \in (\bar{\delta}, 1)$, for all $w \in \mathcal{W}^*$, $r \in \{T - 2n, T - 2n + 1, \ldots, T\}$, j, and $a_j \neq \widehat{f}_j(w, r)$, if $w \neq w(j, 0)$, then,

$$\frac{\widehat{g}_j^{(w,r)}(\widehat{f}(w, r)) - \widehat{g}_j^{(w,r)}(a_j, \widehat{f}_{-j}(w, r))}{1 - \delta} \geq 2\varepsilon.$$

If $w = w(j, 0)$, then

$$\widehat{\tau}((w, r), \widehat{f}(w, r)) = \widehat{\tau}((w, r), (a_j, \widehat{f}_{-j}(w, r))).$$

Proof. Suppose the current state is $w(d)$ for some d. Under $\widehat{\mathcal{A}}$, in any period not corresponding to $w(d)$, all players are supposed to choose $a_i = N$. A unilateral deviation, by j say, yields an action profile a with $|\{a_i = N\}| = n - 1$, and so the new current state is $w(j, 0)$. For player j, the potential benefit of such a deviation is largest when $w = w(j)$ (since there is no loss of long-run value in this case; otherwise such a deviation results in a loss of long-run value of at least $v_j - v_j^j$). The benefit can be bounded by noting that deviating can contribute at most $2n + 1$

19

periods of benefit (the maximum impact in the announcement phase), and so the benefit is no more than

$$(1 - \delta^{2n+1})(\max_a u_j(a) - \min_a u_j(a)) + \delta^{2n+1}(1 - \delta^L)(\underline{v}_j^p - v_j^j)),$$

and the normalized (by $(1 - \delta)^{-1}$) benefit is then no more than

$$(2n + 1)(\max_a u_j(a) - \min_a u_j(a)) + L(\underline{v}_j^p - v_j^j)) < -3\varepsilon$$

(from (10)).

Similarly, in the period corresponding to $w(d)$, all players are supposed to choose $a_i = Y$. A unilateral deviation by j yields a profile in class 2, and new current state $w(j, 0)$. As above, such a deviation is not profitable.

Suppose now the current state is $w = w(k, 0)$ for some k. Under $\widehat{\mathcal{A}}$, in any period not corresponding to $w(k, 0)$, all players other than k (who is myopically optimizing) are supposed to choose $a_i = N$. A unilateral deviation, by $j \neq k$ say, yields an action profile a with $|\{a_i = N\}| \geq n - 2$. From Figure 1, the resulting new current state is either $w(0)$, when $a_k \neq N$, or $w(j, 0)$, when $a_k = N$. Note that such a deviation by j must result in a loss of long-run value, since $v_j^k > v_j^0, v_j^j$.

Finally, in the period corresponding to $w(k, 0)$, all players (including k, for whom every myopic best reply to $a_{-k} = Y$ is categorized as Y) are supposed to choose $a_i = Y$. A unilateral deviation by $j \neq k$ is clearly strictly suboptimal, since it leads to the current state $w(j, 0)$, while a unilateral deviation by k to an action $a_k \neq Y$ does not alter the current state and is strictly suboptimal. ∎

Matters are more delicate for three players, since classes 2 and 3 in Figure 1 overlap. For more than three players, profiles in class 2 and in class 3 lead to distinct current states. Consider the action profile YNY in the period corresponding to the state $w(1)$. This may be the result of a unilateral deviation by player 2 from the current state $w(1)$, for which the appropriate new state is $w(2, 0)$. On the other hand, if the current state is $w(i, 0)$ for $i = 1$ or 3 , then player i will myopically optimize, and we cannot rule out the possibility that player i's action Y is a myopic best reply to the action profile NN of the other two players. Consequently, YNY may be the result

class	profile	current state	ending state
1	$\|\{a_i = Y\}\| = 3$	w	$w(d)$
2A	$\|\{a_i = Y\}\| = 2, a_j = N, j \neq d$	$w(k,0),\ k \neq j$	$w(0)$
		otherwise	$w(j,0)$
2B	$\|\{a_i = Y\}\| = 2, a_j \neq Y$, not in class 2A	w	$w(j,0)$
3	$\|\{a_i = Y\}\| \leq 1, \|\{a_i = N\}\| = 1$	w	$w(0)$
4	$\|\{a_i = N\}\| = 2,\ a_j \neq N$	w	$w(j,0)$
5	not in the above classes	w	w

Figure 2: State determination in the period corresponding to state $w(d)$ when $n = 3$.

of a unilateral deviation by player $j \in \{1,3\}$, $j \neq i$, in the current state $w(i,0)$, for which the appropriate new state is $w(j,0)$.

The idea is to treat separately state transitions in the periods corresponding to states $w(d)$ from those corresponding to states $w(k,0)$, and (since we have strong player-specific punishments) use $w(0)$ in potentially ambiguous situations. The state transitions in the periods corresponding to states $w(d)$ are given in Figure 2.

Turning to the state transitions in periods corresponding to states $w(k,0)$, for profiles in classes 2 and 4, the state is updated as follows: If the current state is $w(d)$ and player i unilaterally deviates, the new state is $w(i,0)$. If the current state is $w(i,0)$ and player $j \neq i$ unilaterally deviates,[11] the new state is $w(0)$, and deviations by player i are ignored. This specification yields for periods corresponding to $w(i,0)$ and profiles in classes 2 and 4, an ending state independent of the current state (see Figure 3).

It remains to describe the ending state when the action profile is in class 5 in a period corresponding to a state $w(k,0)$. Fix an initial state w at the beginning of the announcement phase, and a history of four action profiles in the periods corresponding to $w(0), w(1), w(2),$ and $w(3)$. Recall that states are announced in the order $w(0), w(1), w(2), w(3), w(1,0), w(2,0),$ and finally $w(3,0)$. We say the ending state in the period corresponding to $w(3)$

_____

[11]For this specification to be consistent with incentives, it is necessary that the choice of YES actions for player i in this period is not arbitrary, being every myopic best reply.

class	profile	ending state
1	$\lvert\{a_i = Y\}\rvert = 3$	$w(k,0)$
2a	$\lvert\{a_i = Y\}\rvert = 2$ and $a_k \neq Y$	$w(k,0)$
2b	$\lvert\{a_i = Y\}\rvert = 2$ and $a_k = Y$	$w(0)$
3	$\lvert\{a_i = Y\}\rvert \leq 1, \lvert\{a_i = N\}\rvert = 1$	$w(0)$
4	$\lvert\{a_i = N\}\rvert = 2,\ a_j \neq N$	$w(j,0)$
5	not in the above classes	

Figure 3: State determination in the period corresponding to state $w(k,0)$ when $n = 3$. See the text for ending state description when the action profile is in class 5.

is *not fixed* by the history if, under the transitions in Figure 2, it depends on w. (Any history not fixing the ending state consists of profiles in classes 2A and 5 only). If the action profile is in class 5, then the new state is the current state, unless this is the last period of the announcement phase, class 5 profiles have been observed in the previous two periods (i.e., corresponding to $w(1,0)$ and $w(2,0)$), and the ending state in the period corresponding to $w(3)$ is not fixed, in which case the new state is $w(0)$.[12] With this specification, the current state at the end of the announcement phase is a function only of the $2n + 1$ action profiles chosen during the announcement phase.

Claim 3 (Incentives in the announcement phase) *Claim 2 also holds for $n = 3$.*

Proof. Suppose the current state is $w(d)$ for some d. Under $\widehat{\mathcal{A}}$, in any period not corresponding to $w(d)$, all players are supposed to choose $a_i = N$. A unilateral deviation, by j say, yields an action profile a with $\lvert\{a_i = N\}\rvert = 2$, and so the new current state is $w(j,0)$, and the deviation is not profitable (as before). Similarly, in the period corresponding to $w(d)$, a unilateral deviation by j yields an action profile a with $\lvert\{a_i = Y\}\rvert = 2$, a new current state of $w(j,0)$ (from class 2A or 2B), and the deviation

---

[12]This specification "works" because, under the profile $\widehat{\mathcal{A}}$, an announcement period corresponding to $w(j,0)$ follows any history that does not fix the ending state.

is not profitable.

Suppose the current state is $w(k,0)$ for some k. In the periods corresponding to $w(d)$, note first that if there is no unilateral deviation, then since $a_k = N$ may not be myopically optimal, the action profile under $\widehat{\mathcal{A}}$ may be in class 4, leading to a "new" current state of $w(k,0)$. A unilateral deviation by $j \neq k$ leads to either the state $w(0)$ (when $a_k \neq N$, through either class 2A or class 3) or $w(j,0)$ (when $a_k = N$), a suboptimal change.

In the periods corresponding to $w(i,0)$, $i \neq k$, a unilateral deviation by $j \neq k$ leads to either the state $w(0)$ (when $a_k \neq N$, through either class 2b or class 3)[13] or $w(j,0)$ (when $a_k = N$), a suboptimal change. As before, we need not consider deviations by k. Finally, in the period corresponding to $w(k,0)$, a unilateral deviation by $j \neq k$ leads to the state $w(0)$ (through class 2b), a suboptimal change. ∎

The proof of the lemma is completed by noting that the payoffs under $\widehat{\mathcal{A}}$, $\widehat{V}(w(0),1)$, is within ε of $\widetilde{V}(w(0)) = v^0 = v$. ∎

Proof of Theorem 3. Since $\mathcal{F}^{\dagger p}$ has nonempty interior, every $v \in \mathcal{F}^{\dagger p}$ allows strong player specific punishments $\{v^i\}_{i=1}^n$. Choose $\eta > 0$ sufficiently small that (9) holds for all payoffs in η-neighborhoods of v and $\{v^i\}_{i=1}^n$. For T' sufficiently large, there exist T'-length histories h and h^i for $i = 1, \ldots, n$ whose average payoffs are within η of v and v^i for $i = 1, \ldots, n$, respectively. We modify the automaton $\widetilde{\mathcal{A}}$ from Section 4 so that each state $w(d)$ is replaced by T' states that cycle through the appropriate finite history as long as no deviation has occurred. As there, any unilateral deviation by i results in a transition to $w(i,0)$. The proof of Lemma 4 now completes the argument with obvious minor modifications. ∎

Remark 4 Except a non-generic set of stage games, every player has a single myopic best reply to any action profile of his opponents. Thus, the equilibria described in the proof of Theorem 3 are patiently strict for all but a non-generic collection of stage games.

---

[13] In period $w(2,0)$, NYY can't lead to $w(1,0)$, because player 2 may be better off in $w(1,0)$ than in $w(3,0)$, and Y may be a myopic best reply for 3 to NN in that period.

If we allow for the possibility of public randomization, i.e. players observe public i.i.d. draws from the uniform distribution on $[0, 1]$, then a straightforward modification of our construction yields patiently strict equilibria for all games. At histories at which a player i has several best replies, let the strategy prescribe one of them, arbitrarily chosen. Recall that player i may have several best replies only in those periods in which he is minmaxed. It suffices to replace the profile $\hat{a}^i$ in the very last period of a minmaxing block with a lottery between $\hat{a}^i$ and the action profile that yields player i the highest possible stage-game payoff. The probability of the latter action profile is set to strictly decrease with the number of deviations of player i (within the current block) from the prescribed action to another best reply to $\hat{a}^i_{-i}$.

∎

6 Private Monitoring Games

In this section we show that the perfect monitoring folk theorem is robust to the introduction of a private monitoring, as long as it is highly correlated. In contrast to Hörner and Olszewski (2006a,b) and other recent work on private monitoring games, the strategy profiles are independent of the details of the private monitoring. In other words, *behavior* in the folk theorem is robust to the introduction of private monitoring.

We model the correlated nature of the private monitoring as follows. We first perturb the game with perfect monitoring into a game with public monitoring, and then perturb towards private monitoring. In order to get such a strong robustness, it is important that the private monitoring not be conditionally independent.

A *game with public monitoring* has a public signal y drawn from a finite set Y, with probability $\rho(y \mid a)$. Ex ante payoffs are given by $u_i : \prod_j A_j \to \mathbb{R}$. (Player i's ex post payoffs are a function of the public signal and i's action only, so that the payoffs do not contain additional information beyond that of the public signal.) A *public* pure strategy has an automaton representation $(\mathcal{W}, w^0, f, \tau)$, with $f : \mathcal{W} \to A$ and $\tau : \mathcal{W} \times Y \to \mathcal{W}$. Note that a game with perfect monitoring is a game with public monitoring, where we take $Y = A$ and $\rho(y \mid a) = 1$ if and only if $y = a$. The definition of bounded recall (Definition 1) applies to public strategies once histories are taken to be public, i.e., $h^t \in Y^t$.

Given an automaton $(\mathcal{W}, w^0, f, \tau)$, denote i's average discounted value from play that begins in state w by $V_i(w)$. An automaton induces a *pub-*

lic perfect equilibrium (or *PPE*) if for all states $w \in \mathcal{W}$, $f(w)$ is a Nash equilibrium of the normal form game with payoffs $g^w : A \to \mathbb{R}^n$, where

$$g^w(a) = (1 - \delta)u(a) + \delta \sum_y V(\tau(w, y))\rho(y \mid a).$$

The PPE is *strict* if $f(w)$ is a strict equilibrium of g^w for all w.[14]

A *game with private monitoring* has a private signal $z_i \in Z_i$ for each player, with the vector $z \equiv (z_1, \ldots, z_n) \in Z \equiv Z_1 \times \cdots \times Z_n$ drawn according to a joint probability distribution $\pi(z \mid a)$. Ex ante payoffs are given as before by $u_i : \prod_j A_j \to \mathbb{R}$. (Player i's ex post payoffs are now a function of the private signal and i's action only.)

Definition 6 *A private monitoring distribution* (Z, π) *is* ε-close *to a full support public monitoring distribution* (Y, ρ) *if*

1. $Z_i = Y$ *for all* i, *and*

2. *for all* $y \in Y$ *and all* $a \in A$

$$|\pi(z_i = y, \forall i \mid a) - \rho(y \mid a)| > 1 - \varepsilon.$$

Observe that any strategy for player i in a repeated game with public monitoring trivially also describes a strategy in the repeated game with private monitoring satisfying $Z_i = Y$. It is thus meaningful to ask if a PPE of a repeated game with public monitoring induces a Nash (or sequential) equilibrium of close-by games with private monitoring. Not only is it meaningful, but a weak notion of robustness surely requires that a PPE induce a Nash equilibrium in sufficiently close-by games with private monitoring.

Mailath and Morris (2006) introduce a more general notion of what it means for a private monitoring distribution to be close to a public monitoring distribution. This notion allows for more private signals than public, but preserves the critical features of Definition 6. In particular, any strategy from the public monitoring game induces a well-defined strategy in the private monitoring game, and it is still meaningful to ask if a PPE of a public monitoring game induces an equilibrium in the private monitoring game. The central result in Mailath and Morris (2006) is the following: Fix essentially any strict PPE that does not have bounded recall. Then, for *any*

---

[14]If ρ has *full support* (i.e., $\rho(y \mid a) > 0 \ \forall y \in Y, a \in A$), then a PPE is strict if and only if each player strictly prefers his public strategy to every other public strategy (Mailath and Samuelson, 2006, Corollary 7.1.1).

private monitoring sufficiently close to public monitoring that also satisfies a richness condition,[15] the strategy profile in the private monitoring game is *not* a Nash equilibrium.

In contrast, any strict PPE that does have bounded recall induces a Nash equilibrium in *all* close-by games with private monitoring.[16]

This then raises the question of whether bounded recall is a substantive restriction. For some parameterizations of the imperfect public monitoring repeated prisoners' dilemma, Cole and Kocherlakota (2005) show that the set of PPE payoffs achievable by bounded recall strongly symmetric profiles is degenerate, while the set of strongly symmetric PPE payoffs is strictly larger.

However, at least for games with almost-perfect almost public monitoring, Theorem 3 implies that bounded recall is not a substantive restriction.

A game with full support public monitoring is η-*perfect* if $Y = A$ and

$$\rho(a \mid a) > 1 - \eta.$$

Clearly, any patiently pseudo-strict subgame perfect equilibrium of the perfect monitoring game induces a patiently pseudo-strict PPE of η-perfect public monitoring games, for η sufficiently small. We then have as an implication of Theorem 3 and Mailath and Samuelson (2006, Lemma 13.5.6):

Theorem 4 *Suppose $n \geq 3$ and v is a strictly individually rational payoff. For all $\beta > 0$, there exists $\bar{\delta} < 1$, and $\eta > 0$ such that such that for all η-perfect full support public monitoring distributions (Y, ρ), there exists $\varepsilon > 0$ such that for all private monitoring distributions ε-close to (Y, ρ), for all $\delta \in (\bar{\delta}, 1)$, there is a sequential equilibrium of the private monitoring repeated game with payoffs within β of v.*

In the absence of patient pseudo-strictness, the order of quantifiers would need to be reversed, so that the bound on the closeness of the private monitoring distributions, ε, would depend on δ, and become increasingly severe as $\delta \to 1$ (Mailath and Samuelson, 2006, Section 13.5). This is an undesirable confounding of time preferences with accuracy in the monitoring.

---

[15]The condition is weaker than, but of the spirit of, a requirement that for all public signals, there are private signals with different ordinal rankings of the odds ratios over actions.

[16]While Mailath and Morris (2002, 2006) discuss only strict equilibria, the extension to pseudo-strict equilibria is immediate.

References

ABREU, D. (1988): "On the Theory of Infinitely Repeated Games with Discounting," *Econometrica*, 56(2), 383–396.

ABREU, D., D. PEARCE, AND E. STACCHETTI (1986): "Optimal Cartel Equilibria with Imperfect Monitoring," *Journal of Economic Theory*, 39(1), 251–269.

——— (1990): "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, 58(5), 1041–1063.

AUMANN, R. J. (1981): "Survey of Repeated Games," in *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgnestern*, ed. by R. A. et al., pp. 11–42. Bibliographisches Institut, Zurich.

BARLO, M., G. CARMONA, AND H. SABOURIAN (2008): "Bounded Memory with Finite Action Spaces," Sabanci University, Universidade Nova de Lisboa and University of Cambridge.

——— (forthcoming): "Repeated Games with One-Period Memory," *Journal of Economic Theory*.

BAVLY, G., AND A. NEYMAN (2003): "Online Concealed Correlation by Boundedly Rational Players," Discussion Paper #336, the Hebrew University of Jerusalem.

BHASKAR, V. (1998): "Informational Constraints and the Overlapping Generations Model: Folk and Anti-Folk Theorems," *Review of Economic Studies*, 65(1), 135–149.

COLE, H. L., AND N. R. KOCHERLAKOTA (2005): "Finite Memory and Imperfect Monitoring," *Games and Economic Behavior*, 53(1), 59–72.

FUDENBERG, D., D. K. LEVINE, AND E. MASKIN (1994): "The Folk Theorem with Imperfect Public Information," *Econometrica*, 62(5), 997–1039.

FUDENBERG, D., AND E. MASKIN (1986): "The Folk Theorem in Repeated Games with Discounting or Incomplete Information," *Econometrica*, 54, 533–554.

——— (1991): "On the Dispensability of Public Randomization in Discounted Repeated Games," *Journal of Economic Theory*, 53(2), 428–438.

GILBOA, I., AND D. SCHMEIDLER (1994): "Infinite Histories and Steady Orbits in Repeated Games," *Games and Economic Behavior*, 6, 370–399.

HÖRNER, J., AND W. OLSZEWSKI (2006a): "The Folk Theorem for Games with Private Almost-Perfect Monitoring," *Econometrica*, 74(6), 1499–1544.

——— (2006b): "How Robust is the Folk Theorem with Imperfect Public Monitoring?," Northwestern University, mimeo.

LEHRER, E. (1994): "Many Players with Bounded Recall in Infinite Repeated Games," *Games and Economic Behavior*, 7, 390–405.

MAILATH, G. J., AND S. MORRIS (2002): "Repeated Games with Almost-Public Monitoring," *Journal of Economic Theory*, 102(1), 189–228.

——— (2006): "Coordination Failure in Repeated Games with Almost-Public Monitoring," *Theoretical Economics*, 1, 311–340.

MAILATH, G. J., AND L. SAMUELSON (2006): *Repeated Games and Reputations: Long-Run Relationships.* Oxford University Press, New York, NY.