



Penn Institute for Economic Research  
Department of Economics  
University of Pennsylvania  
3718 Locust Walk  
Philadelphia, PA 19104-6297  
[pier@econ.upenn.edu](mailto:pier@econ.upenn.edu)  
<http://economics.sas.upenn.edu/pier>

## *PIER Working Paper 10-007*

“Folk Theorems with Bounded Recall under (Almost) Perfect Monitoring”  
Third Version

by

George J. Mailath and Wojciech Olszewski

<http://ssrn.com/abstract=1564163>

# Folk Theorems with Bounded Recall under (Almost) Perfect Monitoring\*

George J. Mailath<sup>†</sup>      Wojciech Olszewski<sup>‡</sup>

March 2, 2010<sup>§</sup>

## Abstract

We prove 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 a consequence, the perfect monitoring folk theorem is shown to be *behaviorally* robust under almost-perfect almost-public monitoring. That is, the *same* specification of behavior continues to be an equilibrium when the monitoring is perturbed from perfect to highly-correlated private.

**Key Words:** Repeated games, bounded recall strategies, folk theorem, imperfect monitoring.

**JEL codes:** C72, C73.

---

\*We thank the National Science Foundation, grants #SES-0648780 (Mailath) and #SES-0453061 (Olszewski), for financial support. The first draft of this paper was written in spring 2007 while the first author was on the faculty at and the second author was visiting Yale University. Both thank the Cowles Foundation and the Department of Economics at Yale University for their hospitality. We thank the audience at the Conference in Honor of John Nash's 80<sup>th</sup> Birthday, June 2008, for helpful comments, and Hamid Sabourian for an extremely useful conversation and for sharing Barlo, Carmona, and Sabourian (2008) with us. We also thank the associate editor and referees for exceptionally helpful and perceptive comments.

<sup>†</sup>University of Pennsylvania, [gmailath@econ.upenn.edu](mailto:gmailath@econ.upenn.edu)

<sup>‡</sup>Northwestern University, [wo@northwestern.edu](mailto:wo@northwestern.edu)

<sup>§</sup>First circulating draft, May 30, 2008

# 1 Introduction

Intertemporal incentives arise when public histories coordinate continuation play.<sup>1</sup> 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?

While in general the answer is no, Mailath and Morris (2002, 2006) identify bounded recall as a sufficient and essentially necessary property for an equilibrium to be robust. A strategy profile in a repeated game has *bounded recall* if play under the profile after two distinct histories that agree in the last  $L$  periods (for some fixed  $L$ ) is equal. Mailath and Morris (2002, 2006) proved that any strict perfect public equilibrium (PPE) 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 PPE in unbounded-recall strategies are typically *not* robust.

Not only is bounded recall almost necessary and sufficient for behavioral robustness to private monitoring, the restriction can be substantive: 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.<sup>2</sup> Our result implies that the folk theorem is *behaviorally* robust under almost-perfect almost-public monitoring. That is, consider any strictly individually rational and feasible payoff. There is a specification of equilibrium behavior with a payoff close to this target payoff in the perfect monitoring game that is *robust*: the *same* specification of behavior continues to be an equilibrium when the monitoring is perturbed from perfect to highly-correlated private. It is worth noting that such a result *requires* that the private monitoring be sufficiently correlated. In particular, such a result cannot hold under conditionally-independent

---

<sup>1</sup>Recent work (Ely and Välimäki (2002); Piccione (2002); Ely, Hörner, and Olszewski (2005)) has found so-called “belief-free” equilibria where this is not true. However, these equilibria rely on significant randomization, and it is not clear if such equilibria can be purified (Bhaskar, Mailath, and Morris, 2008).

<sup>2</sup>As observed by Mailath and Samuelson (2006, Remark 13.6.1), the proof of this result in Mailath and Morris (2002) is fundamentally flawed.

private-monitoring.<sup>3</sup>

Our general result (Theorem 4) uses calendar time in an integral way in the construction of the strategy profile. In particular, in the equilibria we construct there is an *announcement* phase that occurs every  $T$  periods. The idea is that play after the previous announcement phase is encoded by the chosen actions, yielding bounded recall. Calendar time is needed so that players know when they are in an announcement phase. If the players' action spaces are sufficiently rich, then the strategy profile can be chosen to be independent of calendar time (Barlo, Carmona, and Sabourian (2009), discussed below; see also Lemma 3). The announcement phase is a substitute for explicit communication, which we do not separately allow. Note that we do *not* rule out communication per se, since our assumptions on stage game payoffs do not preclude some of the actions being cheap-talk messages. However, consistent with our concern with robustness to imperfections in monitoring, unlike Compte (1998) and Kandori and Matsushima (1998), we have not separately allowed players a communication channel that is immune to such imperfections.

Our interest in bounded-recall strategies arises 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) and Sabourian (1998)). 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 (called “strong” in this literature) 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)).

---

<sup>3</sup>Matsushima (1991) shows that if the private monitoring is conditionally independent, then essentially the only pure-strategy equilibria are repetitions of stage-game Nash equilibria. Hörner and Olszewski (2006) show that there are equilibria in close-by games with (even conditionally-independent) private monitoring. However, the fine details of the behavior specified in these equilibria depend on the specifics of the perturbation in the monitoring.

Assuming discounting and perfect monitoring, Barlo, Carmona, and Sabourian (2009) establish the subgame-perfect folk theorem for games with rich action spaces by using strategies with stationary one-period recall. The idea is that if the action spaces are sufficiently rich (for example, convex), players are able to encode entire histories in single stage-game actions. In independent work,<sup>4</sup> Barlo, Carmona, and Sabourian (2008) prove a bounded recall folk theorem for finite action games when either there are two players (and the set of feasible and individually rational payoffs has nonempty interior) or there are more than two players and the stage game satisfies some “confusion-proof” conditions introduced in Barlo, Carmona, and Sabourian (2009).<sup>5</sup> Hörner and Olszewski (2009) establish the folk theorem in bounded recall strategies without the assumption that the action spaces are rich, and even under imperfect public monitoring.<sup>6</sup> 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, and so their folk theorem is not behaviorally robust.

## 2 Preliminaries

In the stage game, player  $i = 1, \dots, n$  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  $(u_1(a), \dots, u_n(a))$  is denoted  $u(a)$ . For each  $i$ , player  $i$ 's (pure action) minmax payoff  $\underline{v}_i^p$  is given by

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

---

<sup>4</sup>We learned of Barlo, Carmona, and Sabourian (2008) when the final draft of this paper was being prepared.

<sup>5</sup>These conditions are related to (but weaker than) the immediate detection property we discuss in Section 4

<sup>6</sup>Note, however, that Hörner and Olszewski (2009) assume the existence of a public correlating device.

so that  $\hat{a}^i$  is an action profile that minmaxes player  $i$ .<sup>7</sup> The payoff  $v_i^p$  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

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

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

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

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

We begin with infinitely repeated games with *perfect monitoring*. In each period  $t = 0, 1, \dots$ , 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  $h^t = (a^0, \dots, a^{t-1}) \in A^t$ , 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  $A^0 = \{\emptyset\}$ . A strategy  $\sigma_i$  for player  $i$  is a function  $\sigma_i : \mathcal{H} \rightarrow A_i$ .

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

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

*A bounded recall strategy profile is a profile with  $L$ -bounded recall for some finite  $L$ .*

Bounded recall strategies are potentially calendar time dependent. For example, the 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

---

<sup>7</sup>If the minmaxing profile is not unique, choose one arbitrarily.

recall strategies. *Stationary bounded recall* also requires calendar time independence, so that the histories  $h^t$  and  $\hat{h}^t$  in (1) can be of different lengths. The profile yielding alternating  $a$  and  $a'$  is not stationary.<sup>8,9</sup>

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,  $f : \mathcal{W} \rightarrow A$  is the output function, and  $\tau : \mathcal{W} \times A \rightarrow \mathcal{W}$  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  $h^t = (a^0, a^1, \dots, a^{t-1})$  such that  $w = \tau(w^0, h^t) \equiv \tau(\dots \tau(\tau(w^0, a^0), a^1), \dots, a^{t-1})$ . 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  $\tilde{w} = \tau(w^0, \tilde{h}^t)$ .

We then have:<sup>10</sup>

**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  $h^L \in \mathcal{H}$ ,*

$$\tau(w, h^L) = \tau(w', h^L). \quad (2)$$

*It has stationary  $L$ -bounded recall if and only if (2) holds for all  $w$  and  $w'$ , including states that need not be reachable in the same period.*

Players share a common discount factor  $\delta < 1$ , and payoffs in the repeated game are evaluated as the average discounted value. 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  $V_i(w)$ .

**Definition 2** *The strategy profile  $\sigma$  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  $g^w : A \rightarrow \mathbb{R}^n$ , where*

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

---

<sup>8</sup>The nomenclature in this area is not settled. Our use of bounded recall follows Hart and Mas-Colell (2006) and Mailath and Morris (2006). Kalai and Stanford (1988), Sabourian (1998), and Mailath and Morris (2002) use *bounded memory* for stationary bounded recall, while Hart and Mas-Colell (2006) use *bounded memory* for profiles that have a finite state space automaton representation. Lehrer's (1994) use of bounded recall imposes stationarity.

<sup>9</sup>To see that alternating between  $a$  and  $a'$  does not have stationary bounded recall (of any length) consider the histories that consist of actions  $a$  only. Without knowing the calendar time, the player does not know whether to play  $a$  and  $a'$ .

<sup>10</sup>See Mailath and Samuelson (2006, Lemma 13.3.1) for a proof.

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

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

It is immediate that  $\sigma$  is a strict subgame perfect equilibrium if, and only 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  $\sigma_i$  and any deviation from  $\sigma_i$  that leaves the outcome path unchanged. This use is motivated by its use in public monitoring games (see footnote 28 and Mailath and Samuelson (2006, Definition 7.1.3)).

Patient strictness (introduced in Mailath and Morris (2002)) is a demanding requirement, since it requires the *same* strategy profile be an equilibrium for all sufficiently large  $\delta$ . It is worth noting that many of the profiles considered in examples (such as grim trigger) are patiently strict. Kalai and Stanford (1988) consider a weaker version, which they term *discount robust subgame perfect equilibrium*, in which the equilibrium is robust to small perturbations of the discount factor.

While the notions of strict and patiently strict are sufficient for the analysis of generic normal form games (see Remark 2), our analysis uses slightly weaker notions to cover all games.

**Definition 3** *The strategy profile  $\sigma$  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  $g^w : A \rightarrow \mathbb{R}^n$  given in (3), and if for all  $a_i \neq f_i(w)$  for  $w$  accessible from  $w^0$  satisfying*

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

we have

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

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

$$\frac{g_i^w(f(w)) - g_i^w(a_i, f_{-i}(w))}{1 - \delta} > \varepsilon. \quad (6)$$



In a pseudo-strict equilibrium, 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 equilibria of a repeated game with public monitoring in bounded recall strategies are robust to private monitoring, and essentially only such strict PPE are robust (Mailath and Morris, 2002, 2006).<sup>11</sup> In particular, a perfect monitoring folk theorem in bounded recall strategies (and pseudo-strict equilibria) is *behaviorally* robust to almost-perfect highly correlated private monitoring perturbations (Theorems 5 and 6). Importantly, patient pseudo-strictness guarantees that the degree of approximation of the public and private monitoring games is independent of the degree of patience.

The cases of two players and more than two players differ, and we consider them in turn.

At the risk of stating the obvious, we allow players to deviate to non-bounded recall strategies. That is, a (patiently pseudo-strict) subgame perfect equilibrium in bounded recall strategies is a strategy profile that is a (patiently pseudo-strict) subgame perfect equilibrium. In contrast, the literature that has focused on bounded recall as a model of bounded rationality often restricts deviations to bounded recall strategies as well (i.e., the recall of the players is bounded). In our setting, the equilibria we study are equilibria also when players' recall is unbounded.<sup>12</sup>

### 3 Two Players and an Easy First Theorem

We begin with 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.,  $u(a) \in \mathcal{F}^{\text{ip}}$ ), there exists  $\bar{\delta} \in (0, 1)$  and a stationary bounded recall strategy profile that is, for all  $\delta \in (\bar{\delta}, 1)$ , a patiently strict subgame perfect equilibrium with outcome  $a$  in every period.*

<sup>11</sup>Mailath and Morris (2002, 2006) discuss only (patiently) strict equilibria. We discuss the extension to (patiently) pseudo-strict equilibria in Section 6.

<sup>12</sup>This need not be true in general: Renault, Scarsini, and Tomala (2007) describe a bounded recall equilibrium in private strategies in a game with imperfect monitoring that is not an equilibrium once players can condition on the entire history.

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 stationary bounded recall profile 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 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 two period recall 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.<sup>13</sup>

Consider a profile with stationary 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 still 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  $DC, CD, DC$ ).

---

<sup>13</sup>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.

The problem identified in the previous paragraph is a familiar one.<sup>14</sup> 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 for an 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 min-max is a Nash equilibrium in the prisoners’ dilemma,<sup>15</sup> and so this allows us to approximate the payoff from  $\frac{1}{2} \circ CD + \frac{1}{2} \circ DC$ .) In Section 5, we use announcement phases to prove a general folk theorem in bounded recall strategies for more than two players.

**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 a bounded recall strategy profile that is, for all  $\delta \in (\bar{\delta}, 1)$ , a patiently strict subgame perfect equilibrium with discounted average payoff within  $\varepsilon$  of  $v$ .*

**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  $h^T \equiv (a^1, \dots, a^{T-1}, \tilde{a}) \in A^T$  whose average payoff is within  $\varepsilon/2$  of  $v$ . As a matter of notation,  $a^T = \tilde{a}$ .

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

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

<sup>14</sup>Barlo, Carmona, and Sabourian (2008) also use this example to illustrate the same point.

<sup>15</sup>Barlo and Carmona (2004) show that all subgame perfect equilibrium outcomes of the discounted repeated prisoners dilemma can be obtained by bounded recall subgame perfect equilibrium strategies.

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, all 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  $\delta$  sufficiently high. ■

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  $h^T$ . 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 is 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 (assuming  $\mathcal{F}^{\dagger p}$  nonempty interior). While the profile has stationary bounded recall (and not just bounded recall), the construction is considerably more complicated than that of our proof of Theorem 2.<sup>16</sup>

**Theorem 3 (Barlo, Carmona, and Sabourian (2008))** *Suppose  $n = 2$ . For all  $v \in \mathcal{F}^{\dagger p}$ , for all  $\varepsilon > 0$ , there exists  $\bar{\delta} \in (0, 1)$  and a stationary bounded recall strategy profile that is, for all  $\delta \in (\bar{\delta}, 1)$ , a patiently strict subgame perfect equilibrium with discounted average payoff within  $\varepsilon$  of  $v$ .*

## 4 Player-Specific Punishments

With more than two players, some notion of player-specific punishments is needed to obtain a folk theorem. In this section, we define player-specific punishments and the associated automaton  $\tilde{\mathcal{A}}$  which plays a central role in the main result.

---

<sup>16</sup>For a simple proof that does not assume  $\mathcal{F}^{\dagger p}$  has nonempty interior (and showing patient strictness), see Appendix A.1.

**Definition 4** A payoff  $v^0$  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

$$v_i^0 > v_i^i \quad \text{and} \quad v_i^j > v_i^i, \quad \forall j \neq i.$$

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

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.<sup>17</sup> The profile's automaton description has set of states

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

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

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

and

$$\tilde{\tau}(w(i, t), a) = \begin{cases} w(j, 0), & \text{if } a_j \neq \hat{a}_j^i, \quad a_{-j} = \hat{a}_{-j}^i \text{ for } j \neq i, \\ w(i, t + 1), & \text{otherwise,} \end{cases}$$

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

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

**Proof.** The proof of [Mailath and Samuelson \(2006, Proposition 3.4.1, statement 1\)](#) shows that  $\tilde{\mathcal{A}}$  describes a pseudo-strict subgame perfect equilibrium

<sup>17</sup>Since player  $i$  has no myopic incentive to deviate from  $\hat{a}^i$ , 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 the proof of [Mailath and Samuelson \(2006, Proposition 3.4.1\)](#)) restart the punishment in this case as well. However, as explained in footnote 20, it is easier to ignore  $i$ 's deviations from  $\hat{a}^i$  in the subsequent development, and we do so here.

for sufficiently large  $L$  and  $\delta$  close to 1. If every  $i$  has a unique best reply to  $\hat{a}_{-i}^i$ , then the induced equilibrium is strict.

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

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

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

$$\begin{aligned} \tilde{g}_j^w(\hat{a}^i) - \tilde{g}_j^w(a_j, \hat{a}_{-j}^i) &= (1 - \delta)(u_j(\hat{a}^i) - u_j(a_j, \hat{a}_{-j}^i)) \\ &\quad + \delta(\tilde{V}_j(w(i, t + 1)) - \tilde{V}_j(w(j, 0))). \end{aligned}$$

But

$$\begin{aligned} \tilde{V}_j(w(i, t + 1)) - \tilde{V}_j(w(j, 0)) &= (1 - \delta^{L-t-1})(u_j(\hat{a}^i) - v_j^p) \\ &\quad + (\delta^{L-t-1} - \delta^L)(v_j^i - v_j^p) + \delta^L(v_j^i - v_j^j), \end{aligned}$$

which is strictly positive for large  $\delta$ , and bounded away from zero as  $\delta \rightarrow 1$  (since  $v_j^i > v_j^j$ ). Suppose  $\varepsilon > 0$  is sufficiently small that  $2\varepsilon < v_j^i - v_j^j$ . Then, for  $\delta$  sufficiently large, for all  $a_j \neq \tilde{f}_j(w)$ ,

$$\tilde{g}_j^w(\tilde{f}(w)) - \tilde{g}_j^w(a_j, \tilde{f}_{-j}(w)) > \varepsilon \quad (7)$$

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

A similar calculation shows that (7) 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

$$\begin{aligned} \frac{\tilde{V}_j(w(j)) - \tilde{V}_j(w(j, 0))}{(1 - \delta)} &= \frac{(1 - \delta^L)(v_j^j - v_j^p)}{(1 - \delta)} \\ &\rightarrow L(v_j^j - v_j^p) \quad \text{as} \quad \delta \rightarrow 1. \end{aligned}$$

If  $L$  satisfies

$$\max_a u_i(a) - v_i^i + 2\varepsilon < L(v_i^i - v_i^p) \quad (8)$$

for all  $i$ , then for  $\delta$  sufficiently large, for all  $a_j \neq \tilde{f}_j(w(j))$ ,

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

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

Because multilateral deviations are ignored,  $\tilde{\mathcal{A}}$  typically does not have bounded recall: For example, a unilateral deviation by  $i$  in  $w(0)$  eventually 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 some games at least, this problem will not arise. In that case, a relatively straightforward modification of  $\tilde{\mathcal{A}}$  yields a bounded recall strategy profile. 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  $a_i$ .<sup>18</sup> 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 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 deviations.

**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  $j = 1, \dots, n$ ). Suppose moreover, that the action profiles  $\{a(d) : d = 0, \dots, n\}$  with  $\{\hat{a}_j^i : i = 1, \dots, n\}$  are all distinct, player by player, that is, for all  $j = 1, \dots, n$ ,*

$$|\{a_j(d) : d = 0, \dots, n\} \cup \{\hat{a}_j^i : i = 1, \dots, n\}| = 2n + 1. \quad (9)$$

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

**Proof.** See appendix. ■

---

<sup>18</sup>Barlo, Carmona, and Sabourian's (2009) notion of *confusion-proof*, while weaker, has a similar interpretation. The details differ, reflecting their use of the richness of the space of action profiles to keep track of behavior; see their paper for details.

**Remark 1** Condition (9) is stronger than necessary. It is enough that every unilateral deviation from an action profile in  $\{a(d) : d = 0, \dots, n\} \cup \{\hat{a}^i : i = 1, \dots, n\}$  be immediately detectable (in the sense described just before the statement of the lemma).

A natural conjecture is that the immediate detectability condition is 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).

The immediate detectability condition can be trivially satisfied when the stage game has connected action spaces and continuous payoffs for each player, since every action profile can be approximated by a continuum of distinct action profiles. Using this observation, Barlo, Carmona, and Sabourian (2009) prove a folk theorem in one-period bounded recall strategies. ■

## 5 Perfect Monitoring Folk Theorem under Bounded Recall

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

**Theorem 4** *Suppose  $n \geq 3$ . For all  $v \in \text{int}\mathcal{F}^{\dagger p}$  and  $\varepsilon > 0$ , there exists  $\bar{\delta} \in (0, 1)$  and a bounded recall strategy profile that is, for all  $\delta \in (\bar{\delta}, 1)$ , a patiently pseudo-strict subgame perfect equilibrium with discounted average payoff within  $\varepsilon$  of  $v$ .*

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



**Definition 5** A payoff  $v^0$  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

$$v_i^j > v_i^0 > v_i^i, \quad \forall j \neq i. \quad (10)$$

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

Note that every  $v^0 \in \text{int}\mathcal{F}^{\dagger p}$  allows strong player-specific punishments, though typically not in pure actions. We first prove the result for  $n \geq 4$ , and then present the modification necessary for  $n = 3$ .

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

**Proof.** In order to deal with the issues raised in Remark 1, we modify  $\tilde{\mathcal{A}}$  by introducing an *announcement* phase of length  $2n + 1$  that begins every  $T > 2n + 1 + L$  periods.<sup>19</sup> In the announcement phase, the players effectively announce the new initial state for the automaton from  $\mathcal{W}^* \equiv \{w(d) : d = 0, \dots, n\} \cup \{w(i, 0) : i = 1, \dots, n\}$ , 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 transition function in  $\tilde{\mathcal{A}}$ ), which is then announced in the next announcement phase (with  $w(i, 0)$  announced if the state reached is  $w(i, t)$  for any  $t = 0, \dots, L - 1$ ).<sup>20</sup> We show that this profile has bounded recall of length  $T + (2n + 1)$ .

The set of states in the modified automaton  $\hat{\mathcal{A}} \equiv (\hat{\mathcal{W}}, \hat{w}^0, \hat{f}, \hat{\tau})$  are  $\hat{\mathcal{W}} = \hat{\mathcal{W}}^N \cup \hat{\mathcal{W}}^A$ , where  $\hat{\mathcal{W}}^N \equiv \tilde{\mathcal{W}} \times \{1, \dots, T - 2n - 1\}$  is the set of states for the normal phase and  $\hat{\mathcal{W}}^A \supset \mathcal{W}^* \times \{T - 2n\} \times \{\dagger\}$  is the set of states for the announcement phase to be described subsequently. The initial state is  $\hat{w}^0 = (w(0), 1)$ .

•**THE NORMAL PHASE:** In the normal phase, the specified behavior agrees with that under our earlier automaton, so that  $\hat{f}(w, r) = \tilde{f}(w)$  for all  $(w, r) \in$

<sup>19</sup>While the idea of using an announcement phase to announce states was inspired by Hörner and Olszewski (2009), the details of the announcement phase are different, reflecting our need to obtain pseudo-strict incentives everywhere.

<sup>20</sup>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 17.

$\widehat{\mathcal{W}}^N$ . The transitions from states in  $\widehat{\mathcal{W}}^N$  are given by

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

That is, within the normal phase, the modified automaton behaves as the original and counts down the periods.

We often refer to  $w \in \widetilde{\mathcal{W}}$  as the *current state*, with the index  $r$  implicit. Moreover, if  $(w, 1)$  is the initial state for a normal phase, we often say  $w$  is the *initial state* for that normal phase.

Behavior under  $\widehat{\mathcal{A}}$  has the property that if  $(w, T - 2n, \dagger)$  is the state at the beginning of the announcement phase, then  $(w, 1)$  is the initial state of the next normal phase. Note that the normal phase is sufficiently long,  $T - 2n - 1 > L$ , that a unilateral deviation does not trigger, under the profile, permanent minmaxing of the deviator (which would be inconsistent with equilibrium incentives). A deviation  $L$  periods before the end of a normal phase yields the maximum number of minmaxing periods,  $2L - 1$ .

Due to the introduction of the announcement phase, a larger value of  $L$  is needed: Fix  $\varepsilon > 0$  satisfying  $\varepsilon < \min_{i \neq j} \{(v_j^i - v_j^j)/2\}$  and choose  $L$  sufficiently large that, for all  $i$ ,<sup>21</sup>

$$(2n + 3)(\max_a u_i(a) - \min_a u_i(a)) + 2\varepsilon < L(v_i^i - v_i^p). \quad (11)$$

•THE ANNOUNCEMENT PHASE: Unfortunately, the description of the announcement phase is unavoidably more complicated than the normal phase.

The set of states is given by

$$\widehat{\mathcal{W}}^A = \left( \mathcal{W}^* \times \{T - 2n + 1, \dots, T\} \right) \cup \left( \mathcal{W}^* \times \{T - 2n, \dots, T\} \times \{\dagger\} \right),$$

so that  $\mathcal{W}^* \times \{T - 2n\} \times \{\dagger\} \subset \widehat{\mathcal{W}}^A$ . We refer to states with the  $\dagger$ -flag as simply *flagged*, and states without it as *unflagged*. The flag means that the current state cannot be determined without knowing the state at the beginning of the announcement phase. In particular, the state at the beginning

<sup>21</sup>This bound works for three or more players. The term  $(2n + 3)$  can be replaced by  $(2n + 2)$  if there are least four players.

of the announcement phase is flagged; once the current state is unflagged, the flag will not reappear within an announcement phase.

As for the normal phase, when the automaton is in state  $(w, r)$  or  $(w, r, \dagger) \in \widehat{\mathcal{W}}^A$ , by current state we mean  $w$ . The integer  $r$  keeps track of progress through the announcement phase. Given a state  $\widehat{w} \in \widehat{\mathcal{W}}^A$ , we sometimes denote the current state by  $w(\widehat{w})$ . For much of the discussion, a state's flag status can be ignored.

Each period  $r = T - 2n, T - 2n + 1, \dots, T$  of the announcement phase corresponds to one of the potential initial states (for the following normal phase),  $w \in \mathcal{W}^*$  in order  $w(0), w(1), \dots, w(n), w(1, 0), \dots, w(n, 0)$ . 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.

Since  $\mathcal{F}^{\dagger p}$  has nonempty interior, for each  $i$ , there is an action profile  $(a_i^N, a_{-i}^Y)$  with the property that  $a_i^N$  is not a myopic best reply to  $a_{-i}^Y$ . In the period corresponding to the state  $w(i, 0)$ , and for each player  $j \neq i$ , we identify the action  $a_j^Y$  in  $A_j$  as YES, and some other action  $a_j^N$  as NO; for player  $i$  we identify every stage-game best reply for  $i$  to the profile  $a_{-i}^Y$  as YES, and the action  $a_i^N$  as NO.

At the beginning of the announcement phase, there is a new current state  $w$  (more precisely,  $(w, T - 2n, \dagger)$ ), 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 to communicate  $w$  truthfully. 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 strong 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.

●●THE OUTPUT FUNCTION IN THE ANNOUNCEMENT PHASE: 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, except for player  $i$  when the current state is  $w(i, 0)$ . When the current state is  $w(i, 0)$  and the current period corresponds to  $w(i, 0)$ , player  $i$  is prescribed action  $Y$ , the stage-game best reply for  $i$  to the profile  $a_{-i}^Y$ . If the current period does not correspond to  $w(i, 0)$ , the specification of player  $i$ 's action depends on the current state's flag status. If the current

class	action profile	current state	ending state	flag status
1	$ \{i : a_i = Y\}  = n$	$w$	$w(d)$	unflagged
2	$ \{i : a_i = Y\}  = n - 1, a_j \neq Y$	$w$	$w(j, 0)$	unflagged
3	$ \{i : a_i = N\}  = n - 2$	$w$	$w(0)$	unflagged
4	$ \{i : a_i = N\}  = n - 1, a_j \neq N$	$w$	$w(j, 0)$	unchanged
5	not in the above classes	$w$	$w$	unchanged

Figure 1: State determination in the period corresponding to state  $w(d) \in \mathcal{W}^*$  when  $n \geq 4$ . For class 4 action profiles, the ending state is flagged if and only if the current state is flagged.

state is flagged, and the current period corresponds to one of the states  $w(0), w(1), \dots, w(n), w(1, 0), \dots, w(i - 1, 0)$ , then player  $i$  is prescribed to play some myopic best response to the profile  $a_{-i}^N$ . Similarly, player  $i$  is prescribed to play some myopic best response to the profile  $a_{-i}^Y$  if the state is unflagged.<sup>22</sup> In all other cases (i.e., when the current period corresponds to one of the states  $w(i + 1, 0), \dots, w(n, 0)$  and the state is flagged), player  $i$  is prescribed play action  $N$ .

It remains to describe how the current state is determined and verify the bounded recall nature of its determination.

●●STATE TRANSITIONS IN THE ANNOUNCEMENT PHASE: Given a current state, and an action profile for that period, we identify a new state from Figures 1, 2, and 3.

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. In both cases, the state is unflagged.

While class 4 action profiles allow the ending state to be unflagged (since it is determined independently of the current state, and so of the beginning state of the announcement phase), the state transitions do not do this. For periods other than the last period of the announcement phase, the state

<sup>22</sup>We need not resolve the ambiguity if player  $i$  has multiple myopic best replies, since incentives are unaffected (continuation play is independent of player  $i$ 's action when the current state is  $w(i, 0)$ ).

class	action profile	current state	ending state	flag status
1	$ \{i : a_i = Y\}  = n$	$w$	$w(k, 0)$	unflagged
2	$ \{i : a_i = Y\}  = n - 1, a_j \neq Y$	$w$	$w(j, 0)$	unflagged
3	$ \{i : a_i = N\}  = n - 2$	$w$	$w(0)$	unflagged
4A	$ \{i : a_i = N\}  = n - 1, a_j \neq N, j > k$	$w$	$w(j, 0)$	unchanged
4B	$ \{i : a_i = N\}  = n - 1, a_j \neq N, j \leq k$	$w$	$w(j, 0)$	unflagged
5	not in the above classes	$w$	$w$	unchanged

Figure 2: State determination in the period corresponding to state  $w(k, 0) \in \mathcal{W}^*$ ,  $k < n$  when  $n \geq 4$ . For class 4A action profiles, the ending state is flagged if and only if the current state is flagged.

class	action profile	current state	ending state	flag status
1	$ \{i : a_i = Y\}  = n$	$w$	$w(n, 0)$	unflagged
2	$ \{i : a_i = Y\}  = n - 1, a_j \neq Y$	$w$	$w(j, 0)$	unflagged
3	$ \{i : a_i = N\}  = n - 2$	$w$	$w(0)$	unflagged
4	$ \{i : a_i = N\}  = n - 1, a_j \neq N$	$w$	$w(j, 0)$	unflagged
5	not in the above classes	$\begin{cases} w, \text{ unflagged} & w \\ w, \text{ flagged} & w(0) \end{cases}$		$\begin{cases} \text{unflagged} \\ \text{unflagged} \end{cases}$

Figure 3: State determination in the period corresponding to state  $w(n, 0)$  when  $n \geq 4$ .

transition may maintain the flag status: a flagged current state leads to the flagged  $w(j, 0)$ , while an unflagged current state leads to the unflagged  $w(j, 0)$  in class 4 of Figure 1; similarly, a flagged current state leads to the flagged  $w(j, 0)$  and an unflagged current state leads to the unflagged  $w(j, 0)$  in class 4A of Figure 2. In the last period, the transitions in class 4, as in all other cases, always lead to an unflagged state.<sup>23</sup>

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 the current state is still flagged, in which case the new state is  $w(0)$ . At the end of the announcement phase, the state is the announced state, which is the initial state of the next normal phase.<sup>24</sup>

Since the current state is necessarily unflagged by the end of the announcement phase, every  $2n + 1$  sequence of action profiles in an announcement phase determines uniquely an announced state, independent of the current state at the beginning of the announcement phase.

If  $w$  is the initial state of the announcement phase, and players play in the announcement phase according to the prescribed strategies, then  $w$  is also the ending state of the announcement phase, i.e., the state at the beginning of the following normal phase. Indeed, if  $w = w(d)$ , all players say  $N$  in all periods of the announcement phase, except the period corresponding to  $w(d)$ , in which they all say  $Y$ . So, the state remains a flagged  $w(d)$  until the period corresponding to  $w(d)$ , becomes an unflagged  $w(d)$  in that period, and remains an unflagged  $w(d)$  till the end of the announcement phase.

If  $w = w(i, 0)$ , then players other than  $i$  say  $N$  in periods corresponding to states  $w(d)$ ,  $d = 0, \dots, n$  and states  $w(1, 0), \dots, w(i - 1, 0)$ , while player  $i$  plays some myopic best response to  $a_{-i}^N$ . The state remains a flagged  $w(i, 0)$  until the period corresponding to state  $w(i, 0)$ , in which all players say  $Y$ , and the state becomes an unflagged  $w(i, 0)$ . Then, in the periods corresponding to  $w(i + 1, 0), \dots, w(n, 0)$ , players other than  $i$  say  $N$ , player  $i$  plays some myopic best response to  $a_{-i}^N$ , and the state remains an unflagged  $w(i, 0)$ .

It remains to check incentives. Recall that given a state  $\hat{w} \in \widehat{\mathcal{W}}^A$ , we sometimes denote the current state by  $w(\hat{w})$ .

---

<sup>23</sup>The role of this specification is, as it will become clear from the later parts of the proof, to make the automaton pseudo-strict.

<sup>24</sup>Formally, when  $w$  is the ending state from Figure 3, the state in  $\widehat{\mathcal{A}}$  is  $(w, 1) \in \widehat{\mathcal{W}}^N$ .

**Claim 1 (Incentives in the announcement phase)** Suppose  $n \geq 4$  and  $T > 2n + 1 + L$  satisfies

$$\frac{2n + 1}{T} \left[ \max_a u_i(a) - \min_a u_i(a) \right] < \varepsilon/3, \quad \forall i. \quad (12)$$

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

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

where  $\widehat{V}_i(\widehat{w})$  is player  $i$ 's payoff from state  $\widehat{w} \in \widehat{\mathcal{W}}$  under  $\widehat{\mathcal{A}}$ . There exists  $\bar{\delta} \in (0, 1)$  such that for all  $\delta \in (\bar{\delta}, 1)$ , for all  $\widehat{w} \in \widehat{\mathcal{W}}^A$ ,  $j$ , and  $a_j \neq \widehat{f}_j(\widehat{w})$ , if  $w(\widehat{w}) \neq w(j, 0)$  or  $\widehat{w} = (w(j, 0), T - k, \dagger)$  for  $k = 0, \dots, n - (j + 1)$ , then,

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

If  $w(\widehat{w}) = w(j, 0)$  and  $\widehat{w} \neq (w(j, 0), T - k, \dagger)$  for  $k = 0, \dots, n - (j + 1)$ , then

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

**Proof.** Consider first the period corresponding to state  $w(d)$ . If the current state is also  $w(d)$ , all players are supposed to take action  $Y$ . If they do so, and follow the prescribed strategies in the remaining periods of the announcement phase, the ending state of the announcement phase and so the initial state of the next normal phase will be  $w(d)$ . A unilateral deviation of player  $i$  results in the new state, and the initial state of the next normal phase being unflagged  $w(i, 0)$ .

For player  $i$ , the potential benefit of such a deviation is largest when  $i = d$  (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  $v_i - v_i^i$ ). The benefit can be bounded by noting that deviating can contribute at most  $2n + 1$  periods of benefit (the maximum impact in the announcement phase), and so the benefit is no more than

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

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

$$(2n + 1)(\max_a u_j(a) - \min_a u_j(a)) + L(v_j^p - v_j^j) < -2\varepsilon$$

(from (11)).

Similarly, if the current state is  $w(d')$ ,  $d' \neq d$ , all players are supposed to take action  $N$ . If  $d' > d$ , the prescribed strategies lead to  $w(d')$  being the initial state of the next normal phase. A unilateral deviation of player  $i$  results in the initial state being  $w(i, 0)$ , and the same argument as in the case of  $d' = d$  applies. If  $d' < d$ , the prescribed strategies lead to  $w(0)$  or  $w(d')$  being the initial state of the next normal phase, depending on the flag status of the current state: if  $w(d')$  is flagged, the current state remains a flagged  $w(d')$  until the very last period of the announcement phase, in which it becomes an unflagged  $w(0)$ ; if  $w(d')$  is unflagged, the current state remains an unflagged  $w(d')$  until the end of the announcement phase. And a unilateral deviation of player  $i$  results in the ending state of the announcement phase being  $w(i, 0)$ , and the same argument as in the case of  $d' = d$  applies again.

If the current state is  $w(i, 0)$ , then players other than  $i$  are prescribed action  $N$ , and player  $i$  is prescribed some myopic best response to  $a_{-i}^N$ . Players  $j \neq i$  have incentives to play  $N$ , because the prescribed strategies lead to the initial state of the next normal phase being  $w(i, 0)$ , while a unilateral deviation results in the initial state being  $w(0)$  or  $w(j, 0)$ ,<sup>25</sup> in which case player  $j$  incurs a long-run loss, since  $v_j^i, v_j^0 > v_j^j$ . Player  $i$ 's action cannot affect the state; the new state will be  $w(i, 0)$  with the same flag status as the current state, independently of player  $i$ 's action.

Consider now the period corresponding to state  $w(i, 0)$ ,  $i < n$ . If the current state is  $w(d)$ , all players are supposed to say  $N$ . Assuming that they play the prescribed strategies in all remaining periods, this leads to the initial state of the next normal phase being  $w(0)$  or  $w(d)$ , depending on whether the current state is flagged or unflagged. In the former case, the initial state is  $w(0)$ , and in the latter, it is  $w(d)$ . A unilateral deviation by player  $j$  changes the state to  $w(j, 0)$ . This new state is unflagged if  $j \leq i$ , and has the same flag status as the current state if  $j > i$ . However, assuming that the players follow the prescribed strategies in all remaining periods, the initial state of the next normal phase will be an unflagged  $w(j, 0)$ , independently of the flag status of this new state.

If the current state is  $w(i, 0)$ , then all players are supposed to say  $Y$ . Players  $j \neq i$  have an incentive to say  $Y$ , because they prefer the next normal phase to begin with  $w(i, 0)$  to the next normal phase beginning with  $w(j, 0)$ . Player  $i$  cannot affect the new state, including its flag status, so any myopic best response, including  $Y$ , is optimal.

---

<sup>25</sup>If the action of player  $i$  differs from  $N$ , then the ending state is  $w(0)$ . Otherwise, the ending state is  $w(j, 0)$ .



If the current state is a flagged  $w(j, 0)$ ,  $j < i$ , then all players are prescribed  $N$ . This leaves the state unaltered, and ultimately leads to the initial state of the next normal phase  $w(0)$ . A unilateral deviation by player  $k$  results instead in the initial state  $w(k, 0)$ . If the current state is an unflagged  $w(j, 0)$ ,  $j < i$ , then player  $j$  is prescribed some myopic best response. This action is optimal, because player  $j$  cannot affect the new state, which will be an unflagged  $w(j, 0)$ . Other players have an incentive to choose the prescribed  $N$ , because this results in the initial state of the next normal phase being  $w(j, 0)$ , while a unilateral deviation by player  $k \neq j$  results in the initial state  $w(0)$  or  $w(k, 0)$ , and a long-run loss of player  $k$ .

If the current state is  $w(j, 0)$ ,  $j > i$ , (no matter whether the state is flagged or not), then players other than  $j$  are supposed to say  $N$ , and player  $j$  is supposed to play some myopic best response to  $a_{-i}^N$ . Since player  $j$  cannot affect the new state, including its flag status, by a unilateral deviation, any myopic best response is optimal. Players  $k \neq j$  also have an incentive to take the prescribed action, since a unilateral deviation results in the initial state of the next normal phase being  $w(0)$  or  $w(k, 0)$ .

Finally, if the current state in the last period of the announcement phase (i.e., the period corresponding to  $w(n, 0)$ ) is  $w(d)$ , all players are supposed to choose  $a_i = N$ . This results in the initial state of the next normal phase being  $w(d)$  or  $w(0)$ , depending on the flag status of the current state. A unilateral deviation by player  $i$  results in the initial state being  $w(i, 0)$ , which is unprofitable as in several previous cases.

If the current state is  $w(n, 0)$ , all players are supposed to choose  $a_i = Y$ . For player  $n$ , this action is optimal, because  $Y$  is some myopic best response to  $a_{-n}^Y$ , and player  $n$  cannot affect the the initial state of the next normal phase by a unilateral deviation. And all other players have incentives to choose  $Y$ , because any other action of player  $i \neq n$  results in the the initial state of the next normal phase being  $w(i, 0)$  instead of  $w(n, 0)$ , and player  $i$  incurs a long-run loss.

If the current state is  $w(i, 0)$ ,  $i < n$ , all players but  $i$  are supposed to choose  $N$ ; player  $i$  is supposed to choose  $a_i = N$  if the state is flagged, and some myopic best response if the state is unflagged. Players  $j \neq i$  have an incentive to play the prescribed action, since any other action results in the the initial state of the next normal phase being  $w(j, 0)$  instead of  $w(i, 0)$ . Player  $i$  cannot affect the initial state by a unilateral deviation if the state is unflagged; and since  $i$  prefers the initial state  $w(0)$  to  $w(i, 0)$ ,  $i$  prefers action  $N$  to any other action if the state is flagged. ■

The next claim asserts that players have no incentive to deviate within the normal phase, assuming the current state at the end of a normal phase is the initial state of the next normal phase (its proof is in the appendix).

**Claim 2 (Incentives in the normal phase)** *Suppose  $T > 2n+1+L$  and satisfies (12). There exists  $\bar{\delta} \in (0, 1)$  such that for all  $\delta \in (\bar{\delta}, 1)$ ,  $\hat{w} \in \widehat{\mathcal{W}}^N$ ,  $j$ , and  $a_j \neq \hat{f}_j(\hat{w})$ , if  $w(\hat{w}) \neq w(j, t)$  for any  $t$ , then*

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

If  $w(\hat{w}) = w(j, t)$  for some  $t$ , then

$$\hat{\tau}(\hat{w}, \hat{f}(\hat{w})) = \hat{\tau}(\hat{w}, (a_j, \hat{f}_{-j}(\hat{w}))), \quad \forall a_j \in A_j.$$

The proof of the lemma is completed by noting that the payoffs under  $\hat{\mathcal{A}}, \hat{V}(w(0), 1)$ , is within  $\varepsilon$  of  $\tilde{V}(w(0)) = v^0 = v$ .  $\blacksquare$

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  $w(0)$ . This may be the result of a unilateral deviation by player 2 from the current state  $w(0)$ , 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 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  $w(j, 0)$ .

A deeper problem arises in the periods corresponding to states  $w(k, 0)$ . For example, take  $k = 2$ . Suppose that the current state is  $w(i, 0)$  for  $i = 1$  or  $3$ , and consider the action profile  $YNY$ . The new state must be  $w(k, 0)$ , since otherwise player  $k$  could avoid punishment by saying  $N$  in the period corresponding to  $w(k, 0)$  when the current state were  $w(k, 0)$ . However, since 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,  $YNY$  may be the result of a unilateral deviation by player  $j \in \{1, 3\}$ ,  $j \neq i$ , in the current state  $w(i, 0)$ . This unilateral deviation may be profitable, because player  $j$ 's payoff may be higher in state  $w(k, 0)$  than in state  $w(i, 0)$ .

**Lemma 5** *Suppose  $v$  allows strong pure-action player-specific punishments. The conclusion of Theorem 4 holds when  $n = 3$ .*

**Proof.** The two problems described above are addressed by modifying the announcement phase of the automaton from the proof of Lemma 4. The first problem requires different state transitions in the periods corresponding to states  $w(d)$  from those corresponding to states  $w(k, 0)$ , while the second problem is dealt with by adding additional states and an additional announcement period at the end. The normal phase of the automaton is unchanged. The set of states in the announcement phase is now given by

$$\widehat{\mathcal{W}}^A = \left( \mathcal{W}^* \times \{T-5, \dots, T, T+1\} \right) \cup \left( \mathcal{W}^* \times \{T-6, \dots, T\} \times \{\dagger\} \right) \\ \cup \left( \mathcal{W}^* \times \{T-1, T, T+1\} \times \{*\} \right).$$

The interpretation of the states that correspond to the first 7 periods of the announcement phases are almost the same as for the  $n \geq 4$  case. As there, the  $\dagger$ -flag means that the current state cannot be determined without knowing the state at the beginning of the announcement phase. In particular, the initial state of the announcement phase is  $\dagger$ -flagged. We explain the  $*$ -flag, used in periods  $T-2$ ,  $T-1$ , and  $T$ , when we describe the state transitions in those periods.

The last period (corresponding to  $(w, T+1)$  or  $(w, T+1, *)$ ) will communicate whether the current state is unflagged. As for the first  $n+1$  periods, in the last period, 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.

••THE OUTPUT FUNCTION IN THE ANNOUNCEMENT PHASE: Apart from the last period, the output function agrees with that for the case  $n \geq 4$ . 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 (except for player  $i$  when the current state is  $w(i, 0)$ , see the  $n \geq 4$  case for details). In the last period, players are prescribed to choose  $Y$  if the current state is unflagged, and  $N$  otherwise, except for player  $i$  when the current state is the unflagged  $w(i, 0)$ , in which case that player  $i$  is prescribed a myopic best reply to the specified behavior of  $Y$  by the other players.

••STATE TRANSITIONS IN THE ANNOUNCEMENT PHASE: The state transitions in the periods corresponding to states  $w(d)$  are given in Figure 4. As intuition for the state transitions, observe that class 2A in Figure 4 also deals with the the action profile  $YNY$  in the period corresponding to the

class	action profile	current state	ending state	flag status
1	$ \{i : a_i = Y\}  = 3$	$w$	$w(d)$	unflagged
2A	$ \{i : a_i = Y\}  = 2, a_j = N$	$\begin{cases} w(k, 0), k \neq j \\ \text{otherwise} \end{cases}$	$\begin{cases} w(0) \\ w(j, 0) \end{cases}$	$\begin{cases} \text{unchanged} \\ \text{unchanged} \end{cases}$
2B	$ \{i : a_i = Y\}  = 2, a_j \notin \{Y, N\}$	$w$	$w(j, 0)$	unflagged
3	$ \{i : a_i = Y\}  \leq 1,$ $ \{i : a_i = N\}  = 1$	$w$	$w(0)$	unflagged
4	$ \{i : a_i = N\}  = 2, a_j \neq N$	$w$	$w(j, 0)$	unchanged
5	not in the above classes	$w$	$w$	unchanged

Figure 4: State determination in the period corresponding to state  $w(d)$  when  $n = 3$ . If the flag status is “unchanged,” then the ending state is unflagged if and only if the current state is unflagged (and so the ending state is †-flagged if and only if the current state is †-flagged).

state  $w(0)$ : If this is the result of a unilateral deviation by player 2 from the current state  $w(0)$ , the new state is  $w(2, 0)$ . On the other hand, if the current state is  $w(k, 0)$  for  $k = 1$  or 3, then the new state is  $w(0)$ .

The state transitions in the  $w(k, 0)$ -periods are described in Figures 5 and 6 (matching the distinction for  $n \geq 4$ ). These transitions deal with profiles in classes 2 and 4 as follows: If the current state is  $w(d)$  and player  $j$  unilaterally deviates, the new state is  $w(j, 0)$ . If the current state is  $w(k, 0)$  and player  $j \neq k$  unilaterally deviates, the new state is  $w(0)$ , and deviations by player  $k$  are ignored.<sup>26</sup> If the current state is  $w(i, 0)$ ,  $i \neq k$ , and player  $j$  unilaterally deviates, the new state is  $w(j, 0)$  or  $w(0)$ . The specification of the state transitions is designed to address the second “deeper” problem: the appropriate specification of the ending state for the profile  $YNY$  in the period corresponding to  $w(2, 0)$ . If the current state is  $w(1, 0)$  or  $w(3, 0)$ , the ending state is the \*-flagged  $w(2, 0)$  (which will lead, in the absence of future deviations, to  $w(0)$ ). If the current state is  $w(2, 0)$ , then the ending state is the unflagged  $w(2, 0)$ .

<sup>26</sup>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 some myopic best reply.

class	action profile	current state	ending state	flag status
1	$ \{i : a_i = Y\}  = 3$	$w$	$w(k, 0)$	unflagged
2a	$ \{i : a_i = Y\}  = 2, a_k \neq Y$	$\begin{cases} w(k, 0), w(d) \\ w(i, 0), i \neq k \end{cases}$	$w(k, 0)$ $w(k, 0)$	unflagged *-flagged
2b	$ \{i : a_i = Y\}  = 2, a_k = Y$	$w$	$w(0)$	unflagged
3	$ \{i : a_i = Y\}  \leq 1,$ $ \{i : a_i = N\}  = 1$	$w$	$w(0)$	unflagged
4A	$ \{i : a_i = N\}  = 2, a_j \neq N, j > k$	$w$	$w(j, 0)$	unchanged
4B	$ \{i : a_i = N\}  = 2, a_j \neq N, j \leq k$	$w$	$w(j, 0)$	unflagged
5	not in the above classes	$w$	$w$	unchanged

Figure 5: State determination in the period corresponding to state  $w(k, 0)$  for  $k = 1, 2$  when  $n = 3$ . If the flag status is “unchanged,” then the ending state is †-flagged (\*-flagged, unflagged, resp.) if and only if the current state is †-flagged (\*-flagged, unflagged, resp.).

class	action profile	current state	ending state	flag status
1	$ \{i : a_i = Y\}  = 3$	$w$	$w(3, 0)$	unflagged
2a	$ \{i : a_i = Y\}  = 2, a_3 \neq Y$	$\begin{cases} w(3, 0), w(d) \\ w(i, 0), i \neq 3 \end{cases}$	$w(3, 0)$ $w(3, 0)$	unflagged *-flagged
2b	$ \{i : a_i = Y\}  = 2, a_k = Y$	$w$	$w(0)$	unflagged
3	$ \{i : a_i = Y\}  \leq 1,$ $ \{i : a_i = N\}  = 1$	$w$	$w(0)$	unflagged
4	$ \{i : a_i = N\}  = 2, a_j \neq N$	$w$	$w(j, 0)$	unflagged
5	not in the above classes	$\begin{cases} w, \text{ not } \dagger\text{-flagged} \\ w, \dagger\text{-flagged} \end{cases}$	$w$ $w(0)$	unchanged unflagged

Figure 6: State determination in the period corresponding to state  $w(3, 0)$  when  $n = 3$ . Note that the flag status is only “unchanged” when the current state is either unflagged or \*-flagged. The ending state is never †-flagged.

class	action profile	current state	ending state	flag status
1	$ \{i : a_i = Y\}  = 3$	$w$	$w$	unflagged
2	$ \{i : a_i = Y\}  = 2, a_j \neq Y$	$w$	$w(j, 0)$	unflagged
3	$ \{i : a_i = Y\}  \leq 1,  \{i : a_i = N\}  \leq 1$	$w$	$w(0)$	unflagged
4	$ \{i : a_i = N\}  = 2, a_j \neq N$	$w$	$w(j, 0)$	unflagged
5	not in the above classes	$w$	$w(0)$	unflagged

Figure 7: State determination in the last period of the announcement phase when  $n = 3$ . Period answers question: “Is current state unflagged?”

Note that the ending state in the  $w(3, 0)$ -period is never †-flagged. The ending state depends on the initial state of the announcement phase only if in the first four periods, the action profiles were in classes 2A or 5, in class 2a in at least one of the next three periods, and in class 5 in the remaining periods. In this case, the ending state will be of the form  $w(k, 0)$ , with  $k$  determined by the history of action profiles. The initial state can only affect the ending state’s flag status, but not the state itself.

Figure 7 describes the state transition in the last period. If the initial state affected the current state’s flag status (as explained above, it cannot affect the state itself), the current state is  $w(k, 0)$ . Truthful reporting then leads to an ending state of  $w(k, 0)$  if it was unflagged and to an ending state of  $w(0)$  if it was \*-flagged, independent of the initial state of the announcement phase. This specification of state transitions is designed to address the second “deeper” problem: the profile  $YNY$  in the period corresponding to  $w(2, 0)$  when the current state is  $w(1, 0)$  or  $w(3, 0)$  leads to the \*-flagged  $w(2, 0)$ , and so truthful reporting leads to the unflagged  $w(0)$ , making the original deviation by player 3 or 1 unprofitable.

With this specification, the current state at the end of the announcement phase is a function only of the  $2n + 2$  action profiles chosen during the announcement phase. The presence of an extra period in the announcement leads to an obvious modification in the bound on  $T$ .

**Claim 3 (Incentives in the announcement phase)** *Suppose  $T > 2n + 2 + L$  satisfies*

$$\frac{2n + 2}{T} \left[ \max_a u_i(a) - \min_a u_i(a) \right] < \varepsilon/3, \quad \forall i. \quad (13)$$

There exists  $\bar{\delta} \in (0, 1)$  such that for all  $\delta \in (\bar{\delta}, 1)$ , for all  $\hat{w} \in \widehat{\mathcal{W}}^A$ ,  $j$ , and  $a_j \neq \hat{f}_j(\hat{w})$ , if either  $w(\hat{w}) \neq w(j, 0)$  or  $\hat{w} = (w(j, 0), T + 1, *)$  or  $(w(j, 0), T - k, \dagger)$  or  $(w(j, 0), T - k, *)$  for  $k = 0, \dots, 2 - j$ , then

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

If  $w(\hat{w}) = w(j, 0)$  and  $\hat{w} \neq (w(j, 0), T + 1, *)$ ,  $(w(j, 0), T - k, \dagger)$  or  $(w(j, 0), T - k, *)$  for  $k = 0, \dots, 2 - j$ , then

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

**Proof.** Most of the proof is analogous to the proof of Claim 1; it requires only some straightforward modifications. For example, checking the incentives in \*-flagged current states is analogous to checking the incentives in †-flagged current states. Two cases are new. First, in the period corresponding to state  $w(k, 0)$ , when the current state is  $w(i, 0)$ ,  $i \neq k$ , player  $k$  has an incentive to say  $N$ , even if player  $i$  is prescribed to play some myopic best response, and this best response is  $Y$ . Since the other player (different from  $i$ ) also plays  $N$ , playing  $N$  in that case leads to the unflagged  $w(i, 0)$ . On the other hand, playing  $a_k \neq Y, N$  leads to state  $w(0)$ ; playing  $a_i = Y$  leads to the new state being a \*-flagged  $w(k, 0)$ , and in turn also to the initial state of the next normal phase being  $w(0)$ , since all players say  $N$  in period  $T + 1$ . This is an argument already familiar from the discussion of the second “deeper” problem.

Second, we must check incentive in period  $T + 1$ , which appears in the announcement phase only for  $n = 3$ . If all players are supposed to say  $N$  (because the state is a flagged  $w(i, 0)$  for some  $i$ ), a unilateral deviation by  $j$  leads to  $w(j, 0)$  instead of  $w(0)$ , and a long-run loss to player  $j$ . If the current state is  $w(d)$ , a unilateral deviation by  $j$  leads to  $w(j, 0)$  instead of  $w(d)$ , and so player  $j$  has an incentive to say  $Y$ .

If finally the current state is the unflagged  $w(i, 0)$ , then players  $j \neq i$  have an incentive to say  $Y$ , which leads to the initial state of the next normal phase being  $w(i, 0)$ , while a unilateral deviation leads to the initial state being  $w(j, 0)$ . Player  $i$  cannot prevent the initial state from being  $w(i, 0)$  by a unilateral deviation, and we have defined  $Y$  as some best response of player  $i$  to  $a_{-i}^Y$  in the case the current state is  $w(i, 0)$ . ■

Since the statement and proof of Claim 2 also hold for the case  $n = 3$ , subject to the same obvious modification to the bound on  $T$  as in Claim 3 (i.e.,  $T > 2n + 2 + L$  and (13)), the lemma is proved. ■

**Proof of Theorem 4.** It is immediate that every  $v^0 \in \text{int}\mathcal{F}^{\dagger p}$  allows strong player specific punishments  $\{v^i\}_{i=1}^n$ . Choose  $\eta > 0$  sufficiently small that (10) holds for all payoffs in  $\eta$ -neighborhoods of  $v^0$  and  $\{v^i\}_{i=1}^n$ . For  $T'$  sufficiently large, there exist  $T'$ -length histories  $h$  and  $h^i$  for  $i = 1, \dots, n$  whose average payoffs are within  $\eta$  of  $v^0$  and  $v^i$  for  $i = 1, \dots, n$ , respectively. We modify the automaton  $\tilde{\mathcal{A}}$  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 proofs of Lemmas 4 and 5 now completes the argument with obvious minor modifications. ■

**Remark 2** Except for a non-generic set of stage games, every player has a unique myopic best reply to any action profile of his opponents. Thus, the equilibria described in the proof of Theorem 4 are patiently strict for all but a non-generic collection of stage games. ■

## 6 Private Monitoring Games

In this section we show that the perfect monitoring folk theorem is robust to the introduction of private monitoring, as long as it is highly correlated. In contrast to Hörner and Olszewski (2006, 2009) 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.<sup>27</sup>

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. Matsushima (1991) shows that if private monitoring is conditionally independent, then the only pure-strategy equilibria, satisfying a condition he called independence of irrelevant information, are repetitions of stage-game Nash equilibria.

---

<sup>27</sup>A referee pointed out that this does not imply that asymptotically, play is similar under perfect and imperfect monitoring. In particular, under perfect monitoring,  $a(0)$  is played in every period, while under imperfect monitoring, eventually, with high probability, one of  $a(i)$ , but not  $a(0)$ , will be played. Our results, however, do imply that it is possible to predict behavior (with high accuracy) after any finite time, for sufficiently small monitoring imperfections.



A *game with public monitoring* has a public signal  $y$  drawn from a finite set  $Y$ , with probability  $\rho(y | a)$ . Ex ante payoffs are given by  $u_i : \prod_j A_j \rightarrow \mathbb{R}$ . (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  $f : \mathcal{W} \rightarrow A$  and  $\tau : \mathcal{W} \times Y \rightarrow \mathcal{W}$ . Note that a game with perfect monitoring is a game with public monitoring, where we take  $Y = A$  and  $\rho(y | a) = 1$  when  $y = a$ . 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 *public 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  $g^w : A \rightarrow \mathbb{R}^n$ , where

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

The PPE is *strict* if  $f(w)$  is a strict equilibrium of  $g^w$  for all  $w$ .<sup>28</sup>

A *game with private monitoring* has a private signal  $z_i \in Z_i$  for each player, with the vector  $z \equiv (z_1, \dots, z_n) \in Z \equiv Z_1 \times \dots \times Z_n$  drawn according to a joint probability distribution  $\pi(z | a)$ . Ex ante payoffs are given as before by  $u_i : \prod_j A_j \rightarrow \mathbb{R}$ . (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*  $(Z, \pi)$  is  $\beta$ -close to a *full support public monitoring distribution*  $(Y, \rho)$  if

1.  $Z_i = Y$  for all  $i$ , and
2. for all  $y \in Y$  and all  $a \in A$

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

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)

<sup>28</sup>If  $\rho$  has *full support* (i.e.,  $\rho(y | a) > 0 \forall y \in Y, a \in A$ ), 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).

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 a private monitoring distribution being *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* private monitoring sufficiently close to public monitoring that also satisfies a richness condition,<sup>29</sup> 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 sequential equilibrium in *all* close-by games with private monitoring.

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 4 implies that bounded recall is not a substantive restriction.

A game with full support public monitoring is  $\eta$ -perfect if  $Y = A$  and

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

Clearly, any patiently strict subgame perfect equilibrium of the perfect monitoring game induces a patiently strict PPE of  $\eta$ -perfect public monitoring games, for  $\eta$  sufficiently small. Recall that if every player has a unique myopic best reply to every action profile of the other players (a generic property, see Remark 2), then the profile of Theorem 4 is patiently strict. We then have as an implication of Theorems 3 and 4, and Mailath and Samuelson (2006, Proposition 13.5.1):

**Theorem 5** *When  $n \geq 2$ , suppose each player  $i$  has a unique stage-game best response to every action profile  $a_{-i} \in A_{-i}$ . For all  $v \in \text{int}\mathcal{F}^{\dagger p}$  and*

---

<sup>29</sup>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.

$\varepsilon > 0$ , there exists a bounded recall strategy profile  $\sigma$ ,  $\bar{\delta} < 1$ , and  $\eta > 0$  such that for all  $\eta$ -perfect full support public monitoring distributions  $(Y, \rho)$ , there exists  $\beta > 0$  such that for all private monitoring distributions  $\beta$ -close to  $(Y, \rho)$ , for all  $\delta \in (\bar{\delta}, 1)$ ,  $\sigma$  describes a sequential equilibrium of the private monitoring repeated game, and has payoffs within  $\varepsilon$  of  $v$ .

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

While the behavioral robustness obtained in Theorem 5 does not hold for nongeneric games, a slightly weaker form does. The difficulty is that while player  $i$  is indifferent over different myopic best replies to  $\hat{a}_{-i}$  in the game with perfect monitoring, in the  $\eta$ -perfect public monitoring game, different myopic best replies may generate different continuations (with probability  $\eta$ ).

Suppose  $\mathcal{A}$  induces a pseudo-strict equilibrium. Denote by  $[\mathcal{A}]$  a new kind of automaton: it is identical to  $\mathcal{A}$  except at states where a player has multiple best replies. At these states, if  $i$  is a player with multiple best replies,  $f_i(w)$  is the set of best replies. Since the original automaton is pseudo-strict, this does not affect any continuations. Note that  $[\mathcal{A}]$  corresponds to a collection of strategy profiles, whose members only differ in the specification of action choice for a player indifferent over different myopic best replies. Let  $[\sigma]$  denote this collection. We emphasize that this collection effectively describes play uniquely: behavior only differs at histories where a player has multiple myopic best replies, and continuations are independent of the myopic best reply made. We say  $[\sigma]$  induces a sequential equilibrium if the collection contains at least one strategy profile that is a sequential equilibrium.

**Theorem 6** *For all  $v \in \text{int}\mathcal{F}^{\dagger p}$  and  $\varepsilon > 0$ , there exists a bounded recall strategy profile  $[\sigma]$ ,  $\bar{\delta} < 1$ , and  $\eta > 0$  such that for all  $\eta$ -perfect full support public monitoring distributions  $(Y, \rho)$ , there exists  $\beta > 0$  such that for all private monitoring distributions  $\beta$ -close to  $(Y, \rho)$ , for all  $\delta \in (\bar{\delta}, 1)$ ,  $[\sigma]$  induces a sequential equilibrium of the private monitoring repeated game, and has payoffs within  $\varepsilon$  of  $v$ .*

**Proof.** Observe that in our construction of the patiently pseudo-strict equilibrium  $\hat{\mathcal{A}}$  in Section 5, when a player had multiple myopic best replies, the choice was arbitrary. Consequently, every profile in  $[\hat{\mathcal{A}}]$  is patiently pseudo-strict. Since the stage game is finite,  $\bar{\delta}$  and  $\varepsilon$  can be chosen so that (6) is

satisfied for all automata in  $[\widehat{\mathcal{A}}]$ . This uniformity is sufficient for the conclusion.  $\blacksquare$

## A Appendix: Omitted Proofs

### A.1 Proof of Theorem 3

Recall that  $\widehat{a}_{-i}^i$  denotes the action of player  $-i$  that minmaxes player  $i = 1, 2$ ; to simplify notation, we write  $\widehat{a}$  for mutual minmax  $(\widehat{a}_1^2, \widehat{a}_2^1)$ . Let  $\widehat{b}_{-i}$  denote an action of player  $-i$  distinct from  $\widehat{a}_{-i}^i$ , and set  $\widehat{b} \equiv (\widehat{b}_1, \widehat{b}_2)$ . There is a cycle of action profiles  $h^n \equiv (a^1, \dots, a^n)$  whose average payoff is within  $\varepsilon/2$  of  $v$ . Without loss of generality assume that:

1. the first  $k \in \{2, \dots, n-1\}$  action profiles of the cycle are  $\widehat{b}$ ,
2. none of the remaining  $n-k$  action profiles is  $\widehat{b}$ , and
3. the  $(k+1)$ -st action profile of the cycle is  $\widehat{a}$ .

Suppose that  $T-n$  periods of playing the mutual minmax suffice to make any sequence of  $n$  unilateral deviations unprofitable; more precisely, suppose that

$$n \cdot [\max_{a \in A} u_i(a) - \min_{a \in A} u_i(a)] < (T-n) \cdot [v_i - u_i(\widehat{a})], \quad \forall i = 1, 2. \quad (14)$$

The profile is described by an automaton with states  $\{w(0, \ell) : \ell = 1, \dots, n\} \cup \{w(1, \ell) : \ell = 1, \dots, T\} \cup \{w(2, \ell) : \ell = 2, \dots, k\}$ , initial period  $w^0 = w(0, 1)$ , and output function  $f(w(0, \ell)) = a^\ell$  and  $f(w(1, \ell)) = f(w(2, \ell)) = \widehat{a}$  for all  $\ell$ . For states  $w(0, \ell)$ , transitions are given by

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

$$\tau(w(0, \ell), \widehat{a}) = \begin{cases} w(1, m+1), & \text{if } a^\ell \neq \widehat{a} \text{ and } m < T, \\ w(0, 1), & \text{if } a^\ell \neq \widehat{a}, \text{ and } m \geq T, \end{cases}$$

where  $m - 1$  is the number of consecutive action profiles  $\widehat{a}$  that precede the current period (in which  $\widehat{a}$  was played again);

$$\tau(w(0, \ell), \widehat{b}) = \begin{cases} w(0, k + 1), & \text{if } \ell = k + 1, \\ w(2, 2), & \text{if } \ell > k + 1; \end{cases}$$

recall that  $a^\ell \neq \widehat{b}$  for  $\ell \geq k + 1$ , so the first two and the last two cases in the definition of  $\tau(w(0, \ell), a)$  are mutually exclusive. Moreover, for  $\ell \leq k$ ,  $a^\ell = \widehat{b}$ , and so all transitions from  $w(0, \ell)$  are described. For states  $w(1, \ell)$  and  $w(2, \ell)$ , the transitions are given by

$$\tau(w(1, \ell), a) = \begin{cases} w(1, \ell + 1), & \text{if } \ell \leq T - 1 \text{ and } a = \widehat{a}, \\ w(0, 1), & \text{if } \ell = T \text{ and } a = \widehat{a}, \\ w(2, 2), & \text{if } \ell \leq T \text{ and } a = \widehat{b}, \text{ and} \\ w(1, 1), & \text{if } \ell \leq T \text{ and } a \neq \widehat{a}, \widehat{b}, \end{cases}$$

and

$$\tau(w(2, \ell), a) = \begin{cases} w(2, \ell + 1), & \text{if } \ell \leq k - 1 \text{ and } a = \widehat{b}, \\ w(0, k + 1), & \text{if } \ell = k \text{ and } a = \widehat{b}, \\ w(1, 2), & \text{if } \ell \leq k \text{ and } a = \widehat{a}, \text{ and} \\ w(1, 1), & \text{if } \ell \leq k \text{ and } a \neq \widehat{a}, \widehat{b}. \end{cases}$$

The “bad” states  $w(1, \ell)$  and  $w(2, \ell)$  encode the number of consecutive action profiles  $\widehat{a}$  and  $\widehat{b}$ , respectively, that precede the current period. That is,  $w(1, \ell)$  means that  $\widehat{a}$  has been played in the  $\ell - 1$  most recent periods, and  $w(2, \ell)$  means that  $\widehat{b}$  has been played in the  $\ell - 1$  most recent periods. If action profile  $\widehat{b}$  happens to be played in  $k$  consecutive periods, then the profile *reinitializes*: players are prescribed to continue playing cycle  $h^n$ , independent of the history more than  $k$  periods ago. It is important that  $k \geq 2$ : If  $k$  were equal to 1, a unilateral deviation in states  $w(0, \ell)$ ,  $\ell > k$ , could cause a (possibly profitable) transition to state  $w(0, k + 1)$ .

According to these strategies, players are prescribed to play the next action profile in the cycle  $h^n$  if the last  $\ell$  periods of the history is consistent with the first  $\ell$  periods of the cycle, for  $\ell = k, k + 1, \dots, n - 1$ . Similarly, they are prescribed to play  $\widehat{b}$  if  $\widehat{b}$  had been played in the  $\ell = 0, \dots, k - 1$  most recent periods (i.e., the first  $\ell$  periods of the cycle has been played), and in the immediately preceding periods, either a full cycle  $h^n$  or  $T$  consecutive periods of  $\widehat{a}$  had been played. In all other cases, players are prescribed to play  $\widehat{a}$ . By construction, the prescribed repeated-game strategies have bounded recall of length  $\max\{T, n + k - 1\}$ .

We now verify that the profile is a strict subgame perfect equilibrium. By a unilateral deviation in state  $w(0, \ell)$ , a player gains at most  $n \cdot [\max_{a \in A} u_i(a) - \min_{a \in A} u_i(a)]$ . Indeed, the player may gain not only in terms of the current flow payoff, but also in terms of the flow payoffs in other periods remaining to the end of the cycle, because for some periods of the cycle, the payoff to playing  $\hat{a}$  may exceed the payoff to playing the prescribed strategy. However, the gain can be estimated by  $[\max_{a \in A} u_i(a) - \min_{a \in A} u_i(a)]$  times the number of remaining periods.

The player loses (approximately)  $v_i - u_i(\hat{a})$  in each of the next periods of mutual minmax, after the cycle ends. And there are at least  $T - n$  such periods. Indeed, any sequence of consecutive action profiles  $\hat{a}$  at the end of the cycle counts to the  $T$  periods of mutual minmax following the deviation, but such a sequence cannot be longer than  $n$  (or more precisely,  $n - 1$ ), because a unilateral deviation in the first  $k$  periods of the cycle cannot yield action profile  $\hat{a}$ . Thus, any unilateral deviation is unprofitable by virtue of (14).

Consider now any state  $w(1, \ell)$ . Players are prescribed to play  $\hat{a}$ . Note that no unilateral deviation can result in  $a = \hat{b}$ . If players play the prescribed action profile  $\hat{a}$ , the next state will be  $w(1, \ell + 1)$  (in the case when  $\ell \leq T - 1$ ) or  $w(0, 1)$  (when  $\ell = T$ ). After a unilateral deviation, the next state will be  $w(1, 1)$ . Players prefer to play the prescribed strategies, since this results in  $T - \ell + 1$  periods of mutual minmax, followed by playing the cycle. Any unilateral deviation results instead in (at best) the minmax payoff in the current period, followed by  $T$  periods of mutual minmax and then playing the cycle. The former option dominates the latter for sufficiently large  $\delta$ .

Similarly in every state  $w(2, \ell)$ , players are prescribed to play  $\hat{a}$ , and the next state will be  $w(1, 2)$  or  $w(1, 1)$ , depending on whether the prescribed action profile  $a = \hat{a}$  or a unilateral deviation occurs. Players prefer to play the prescribed strategies, since this results in  $T$  periods of mutual minmax, followed by playing the cycle. Any unilateral deviation results instead in (at best) the minmax payoff in the current period, followed by  $T$  periods of mutual minmax and then playing the cycle.

Proving patient strictness is similar to the arguments in the proof of Lemma 2.

## A.2 Proof of Lemma 3

Since there are three or more players, (9) implies that every unilateral deviation from an action profile in  $\{a(d) : d = 0, \dots, n\} \cup \{\hat{a}^i : i = 1, \dots, n\}$  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  $\tilde{\mathcal{A}}$ , choose  $L$  sufficiently large that (8) is satisfied. The new automaton has set of states

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

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

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

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

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

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

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  $a_{-i} \neq \widehat{a}_{-i}^i$  for all  $i$ , the automaton immediately transits to the state  $w(0)$ . Finally, after an action profile satisfying  $a_{-i} = \widehat{a}_{-i}^i$  for some  $i$ , the automaton transits to a state  $w(i, t)$ , with the value of  $t$  determined by the previous state. Subsequent  $a_{-i} = \widehat{a}_{-i}^i$  increment the counter  $t$ , till  $t = L$ .

Consider now a  $T$ -length history, with  $a^T$  being the last period action profile. If  $a_{-i}^T \neq \widehat{a}_{-i}^i$  for all  $i$ , then the current state is determined from the previous paragraph. Suppose now that there is some  $i$  for which  $a_{-i}^T = \widehat{a}_{-i}^i$ , and let  $\ell = \max\{t : a_{-i}^t \neq \widehat{a}_{-i}^i\}$ ; note that  $\ell < T$ . Then, the current state of the automaton is given by  $w(i, t')$ , where  $t' = \min\{T - \ell + 1, L\}$ . 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 strict.

### A.3 Proof of Claim 2

Observe first that, due to (12), for a sufficiently patient player  $i$ ,  $\widehat{V}_i(w, r)$ , is within  $\varepsilon/3$  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 (7) that for large  $\delta$ ,<sup>30</sup> for all  $a_j \neq \widehat{f}_j(w, r)$ ,

$$\widehat{g}_j^{(w,r)}(\widehat{f}(w, r)) - \widehat{g}_j^{(w,r)}(a_j, \widehat{f}_{-j}(w, r)) > \varepsilon/3$$

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

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

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

<sup>30</sup>The bound on  $\delta$  is tighter than that yielding (7) 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.



Taking limits,

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

which exceeds  $2\varepsilon$  from (11). This yields the desired inequality for large  $\delta$ .

Finally, independence of state transitions in state  $w(j, t)$  to player  $j$ 's behavior is by construction.

## References

- AUMANN, R. J. (1981): "Survey of Repeated Games," in *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*, ed. by R. A. et al., pp. 11–42. Bibliographisches Institut, Zurich.
- BARLO, M., AND G. CARMONA (2004): "Time Dependent Bounded Recall Strategies Are Enough to Play the Discounted Repeated Prisoners' Dilemma," Sabanci University and Universidade Nova de Lisboa.
- BARLO, M., G. CARMONA, AND H. SABOURIAN (2008): "Bounded Memory with Finite Action Spaces," Sabanci University, Universidade Nova de Lisboa and University of Cambridge.
- (2009): "Repeated Games with One-Period Memory," *Journal of Economic Theory*, 144(1), 312–336.
- 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.
- BHASKAR, V., G. J. MAILATH, AND S. MORRIS (2008): "Purification in the Infinitely-Repeated Prisoners' Dilemma," *Review of Economic Dynamics*, 11(3), 515–528.
- COLE, H. L., AND N. R. KOCHERLAKOTA (2005): "Finite Memory and Imperfect Monitoring," *Games and Economic Behavior*, 53(1), 59–72.

- COMPTE, O. (1998): “Communication in Repeated Games with Imperfect Private Monitoring,” *Econometrica*, 66(3), 597–626.
- ELY, J. C., J. HÖRNER, AND W. OLSZEWSKI (2005): “Belief-Free Equilibria in Repeated Games,” *Econometrica*, 73(2), 377–416.
- ELY, J. C., AND J. VÄLIMÄKI (2002): “A Robust Folk Theorem for the Prisoner’s Dilemma,” *Journal of Economic Theory*, 102(1), 84–105.
- 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.
- GILBOA, I., AND D. SCHMEIDLER (1994): “Infinite Histories and Steady Orbits in Repeated Games,” *Games and Economic Behavior*, 6, 370–399.
- HART, S., AND A. MAS-COLELL (2006): “Stochastic Uncoupled Dynamics and Nash Equilibrium,” *Games and Economic Behavior*, 57(2), 286–303.
- HÖRNER, J., AND W. OLSZEWSKI (2006): “The Folk Theorem for Games with Private Almost-Perfect Monitoring,” *Econometrica*, 74(6), 1499–1544.
- (2009): “How Robust is the Folk Theorem?,” *Quarterly Journal of Economics*, 124(4), 1773–1814.
- KALAI, E., AND W. STANFORD (1988): “Finite Rationality and Interpersonal Complexity in Repeated Games,” *Econometrica*, 56(2), 397–410.
- KANDORI, M., AND H. MATSUSHIMA (1998): “Private Observation, Communication and Collusion,” *Econometrica*, 66(3), 627–652.
- 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.
- MATSUSHIMA, H. (1991): “On the Theory of Repeated Games with Private Information. Part I: Anti-Folk Theorem Without Communication,” *Economic Letters*, 35, 253–256.
- PICCIONE, M. (2002): “The Repeated Prisoner’s Dilemma with Imperfect Private Monitoring,” *Journal of Economic Theory*, 102(1), 70–83.
- RENAULT, J., M. SCARSINI, AND T. TOMALA (2007): “A Minority Game with Bounded Recall,” *Mathematics of Operations Research*, 32(4), 873–889.
- SABOURIAN, H. (1998): “Repeated Games with  $M$ -Period Bounded Memory (Pure Strategies),” *Journal of Mathematical Economics*, 30, 1–35.