Private Monitoring without Conditional Independence (PRELIMINARY AND INCOMPLETE)

Olivier Gossner and Johannes Hörner

February 15, 2007

Abstract

We prove that efficiency can be asymptotically achieved in the two-player's prisoner's dilemma under private monitoring. While we impose some restrictions on the monitoring structure, we do not require that the monitoring be either almost-perfect or conditionally independent.

1 Introduction

The last decade has witnessed a significant improvement in our understanding of repeated games with private monitoring. An excellent survey is provided by [Kan02]. Starting with [Sek97], most of the literature has focused on a specific game, the two-player prisoner's dilemma (or straightforward generalizations thereof), and endeavored to show that cooperation can be approximately sustained when players are sufficiently patient. With a few notable exceptions, all papers have assumed that the monitoring was almost-perfect. That is, players' signals were assumed to be arbitrarily, if not perfectly, informative. Under such limiting structures, it was shown that the payoff from cooperation, and indeed, any feasible and individually rational payoff, could be (approximately) achieved in some sequential equilibrium of the game, provided the discount factor was high enough (see [Sek97], [EV02], [BO02], [Pic02]), a result later generalized to all games (with two players or more) by [HO06].

Several papers have considered private monitoring structures that were not close to perfect, but instead arbitrarily close to a public monitoring structure ([MM02], [MM]). It was recently shown that, in this case as well, a version of the folk theorem holds under some identifiability assumptions (see Hörner and Olszewski, 2007). In particular, under those identifiability assumptions, the folk theorem holds for the two-player prisoner's dilemma.

All these papers establish 'limit' results, in the sense that they only hold for monitoring structures within a (possibly arbitrarily small) neighborhood of either perfect, or imperfect public monitoring structures. In an important paper, Matsushima (2004) shows that the folk theorem also holds for the two-player prisoner's dilemma under monitoring structures that are bounded away from public monitoring structures, provided signals are not completely uninformative. Unfortunately, Matsushima requires the monitoring structure to satisfy conditional independence. That is, conditional on any (pure) action profile, the two players' signals are independently distributed. Furthermore, this assumption plays a central role in his proof. Roughly speaking, it ensures that a player's signals do not affect his beliefs about his own performance in some statistical test performed by his opponent on the basis of her signals. This drastically reduces the dimensionality of the deviations whose suboptimality must be verified.

The goal of this paper is to show that efficiency can be achieved in the two player's prisoner's dilemma without such a restriction. However, our result requires restrictions on the monitoring structure as well, but, as opposed to all the previous papers, the set of monitoring structures that satisfy these restrictions is not non-generic. Nevertheless, it does not include all the monitoring structures considered in Matsushima (2004), as we require that one signal be sufficiently (but not arbitrarily) informative. The precise restrictions are described in the next section.

As mentioned, this paper does not prove efficiency under the most general assumptions one could hope for. While we do not know whether our restrictions are necessary for our result, we do hope that this paper will provide an impetus for the general case similar to the one generated by [Sek97] for the case of almost-perfect monitoring.[1]

The next section describes the set-up and states the result. Section 3 provides the proof of the main lemma.

2 The Model

We consider the δ-discounted infinitely repeated prisoner's dilemma, the normalized stage-game payoffs of which are displayed in the next figure.

	C	D
C	$(1,1)$	$(-L,1+G)$
D	$(1+G,-L)$	$(0,0)$

We assume that $G>0, L>0$ and $G-L<1$, so that the efficient outcome is (C,C). Player i's action is denoted $a_i \in A_i = \{C,D\}$, and we write $a=(a_1,a_2)\in A = A_1\times A_2$ for a generic action profile. Mixed actions are denoted $\alpha_i \in \triangle A_i$, and $\alpha \in \triangle A_1 \times \triangle A_2$ denotes a mixed action profile. Mixed actions (in fact, actions) are unobservable, and there is no public randomization device.

Conditional on any action profile a, each player receives a signal ω_i from a finite set Ω_i according to some probability distribution $m(\omega|a)$, where $\omega = (\omega_1,\omega_2)\in\Omega=\Omega_1\times\Omega_2$. A monitoring structure is a collection $\{m(\cdot|a): a\in A\}$ of such distributions. Given $m(\cdot|a)$, $m_i(\cdot|a)$ denotes the marginal distribution over player i's signals ω_i. For simplicity, we assume that monitoring has full support. That is, $m(\omega|a)>0$, $\forall a,\omega$. Whenever we write $m_i(\omega_i|a)$, we order a so that player i's action is last. That is, $m_i(\omega_i|CD)$ means that player $-i$ plays C and player i plays D. Similarly, $m_{-i}(\omega_{-i}|CD)$ means that player i plays C and player $-i$ plays D.

We view our first assumption as relatively innocuous, as it is generically satisfied provided $|\Omega_1| = |\Omega_2|$.

Assumption 1: $|\Omega_1| = |\Omega_2|$. In addition, for some totally mixed action of player $-i$, the matrix with $2\times|\Omega_i|$ rows and $2\times|\Omega_{-i}|$ columns, with generic entry $\Pr\{a_{-i}\omega_{-i}|a_i\omega_i\}$, is invertible.

[1] Indeed, for the case of almost-perfect monitoring, [Sek97] imposed some restrictions on the monitoring structures that were subsequently dispensed with.

Our second assumption states that each player observes a sufficiently informative signal.

Assumption 2: *For* $i = 1, 2$, there exists ω'_{-i}, ω''_{i} with

$$\frac{m_{-i}\left(\omega'_{-i}|CD\right)}{m_{-i}\left(\omega'_{-i}|DD\right)} > 1 + L,$$

and

$$\frac{m_{-i}\left(\omega''_{-i}|CC\right)}{m_{-i}\left(\omega''_{-i}|DC\right)} > 1.$$

Obviously, the first requirement is stronger than simply requiring that some signal be not completely uninformative, which would only require that $m_{-i}\left(\omega'_{-i}|CD\right) > m_{-i}\left(\omega'_{-i}|DD\right)$. The signal ω''_{-i} satisfying this assumption is referred to as c, or a 'good' signal. Furthermore,

Assumption 3: *For* $i = 1, 2$,

$$\sum_{\omega_{-i}} m_{-i}\left(\omega_{-i}|DD\right) \frac{m\left(c\omega_{-i}|CD\right)}{m_{-i}\left(\omega_{-i}|CD\right)} < \sum_{\omega_{-i}} m_{-i}\left(\omega_{-i}|CC\right) \frac{m\left(\omega_{-i}c|CC\right)}{m_{-i}\left(\omega_{-i}|CC\right)}.$$

Observe that the right-hand side is simply $m_i\left(c|CC\right)$, the probability that player i observes a good signal when both players play C. The left-hand side is the expected probability that, given his signal ω_{-i}, player $-i$ assigns to his opponent receiving signal c, if player $-i$ plays D and believes that i plays C. The expectation, however, is computed with respect to the action profile (D, D). That is, the right-hand side is i's belief about i's signal when $-i$ plays D and erroneously believes that i plays C. Therefore, Assumption 3 rules cases in which player $-i$'s feedback about the signal i has observed about him is better when $-i$ plays D, but erroneously believes that i plays C, than when $-i$ plays C and correctly believes that his opponent plays C as well. Loosely speaking, Assumption 3 requires $-i$'s action to be more important than his opponent's action in determining the probability that i's signal about $-i$'s action is good or bad.

While restrictive, assumptions 1-3 are not non-generic. For instance, in the standard case in which Ω_i has two elements, $L = 1$, and signals are either both correct or both incorrect with respective probabilities $(1-\varepsilon)^2 + \rho\varepsilon(1-\varepsilon)$ and $\varepsilon^2 + \rho\varepsilon(1-\varepsilon)$ (with the remaining two signal profiles each occurring with probability $(1-\rho)\varepsilon(1-\varepsilon)$), the resulting set of parameters is:

$$\varepsilon < 1/3,$$

as the first assumption is trivially satisfied, the second requires $\varepsilon < 1/3$, and the third requires $\varepsilon < 1/2$.

Our main result is the following:

Proposition 1 *Under assumptions 1-3, for any $\epsilon > 0$, there exists $\bar{\delta} < 1$ such that, $\forall \delta > \bar{\delta}$, (x, x), for some $x > 1 - \epsilon$, is an equilibrium payoff vector of the infinitely repeated game with discount factor δ.*

The main step in the proof is the following Lemma, proved in the next section. Before stating this Lemma, additional notation must be introduced. Given $T < \infty$, we consider the δ-discounted game with finite horizon T, augmented by the terminal payoff $\pi_i : H_{-i}^T \to \mathbb{R}$. That is, i's payoff in the finitely repeated game is:

$$U_i^A = \frac{1-\delta}{1-\delta^{T+1}} \left(\sum_{t=1}^{T} \delta^{t-1} u_i \left(a^t\right) + \delta^T \pi_i \left(h_{-i}^T\right) \right),$$

where a^t is the realized action profile in period t, h_{-i}^T denotes the T-history observed by player $-i$, and u_i is player i's payoff in the prisoner's dilemma. Let

$$U_i^T \left(a_1, \ldots, a^T\right) = \sum_{t=1}^{T} \delta^{t-1} u_i \left(a^t\right)$$

denote the total flow payoffs in the finitely repeated game. We extend the definition to mixed actions and to strategy profiles $s = (s_1, s_2)$, where $s_i : \cup_{t \le T} H_i^t \to \triangle A_i$. Further, we let $B\left(s_{-i}, \pi_i\right)$ denote player i's optimal strategies in the game in which player $-i$ uses strategy $s_{-i} \in S_i^T$ (the set of strategies of player i in the T-finitely repeated game) and i's terminal payoff function is given by π_i. We may now state:

Lemma 2 *For $i = 1, 2$, there exists $s_i^C, s_i^D \in S_i^T$ and $\pi_i^C : H_{-i}^T \to \mathbb{R}$ such that:*

1. $\left\{s_i^C, s_i^D\right\} \subseteq B\left(s_{-i}^C, \pi_i^C\right)$;
2. $\lim_T U_i^T \left(s^C\right) / T = 1$;
3. $\lim_T U_i^T \left(s^D\right) / T = 0$, *and:*
4. $\lim_T \left(E\left(\pi_i^C \left(h_{-i}^T\right) | s^C\right) - \max_{h_{-i}^T} \pi_i^C \left(h_{-i}^T\right) \right) / T = 0$,

where $s^C = \left(s_1^C, s_2^C\right)$ *and* $s^D = \left(s_1^D, s_2^D\right)$.

Proof of Proposition 1:

The strategies we specify can be described by automata, which revise states and actions at the beginning of each round. Before defining these automata, we shall define a sequence of payments θ_i^t. Given s_{-i}^D and h_{-i}^T, define θ_i^T such that (i) given $s_{-i}^D\left(h_{-i}^T\right)$, player i is indifferent between C and D. This is possible because of Assumption 1. Adding a constant if necessary, we may assume that $\theta_i^T \geq 0$ for all h_{-i}^T, with the lowest being exactly 0. We view θ_i^T as additional payments to player i, that will be achieved through appropriate transition probabilities.

More generally, given h_{-i}^t, a_{-i}^t and ω_{-i}^t, define $\theta_i^t \geq 0$ such that i is indifferent between playing C and D in period t. That is, given i's expected continuation payoff from $s_{-i}|_{h_{-i}^{t+1}}$, including the expected values of $\theta_i^{t'}$, $t' \geq t+1$, we choose θ_i^t so that the current payoff in period t, plus the continuation payoff be independent of i's action, which again is possible by Assumption 1. Again, for each h_{-i}^t, a_{-i}^t and ω_{-i}^t, pick the lowest θ_i^t equal to 0.

Let $\underline{v}_i$ be equal to

$$\frac{1-\delta}{1-\delta^{T+1}} \mathbb{E}_{s^D} \left\{ \sum_{t=1}^{T} \delta^{t-1} \left(u_i\left(a^t\right) + \theta_i^t\right) \right\}.$$

We claim that, for δ close enough to one, $\underline{v}_i < 1$, by Assumption 2. Let

$$\pi_i^D\left(h_{-i}^T\right) = \delta^{-T}\sum_{t=1}^{T}\delta^{t-1}\theta_i^t.$$

Finally, define

$$\bar{v}_i = \frac{1-\delta}{1-\delta^{T+1}}\mathbb{E}_{s^C}\left\{\sum_{t=1}^{T}\delta^{t-1}u_i\left(a^t\right) + \delta^T\pi_i^C\left(h_{-i}^T\right)\right\},$$

which tends to 1 if $T\to\infty, \delta^T\to 1$.

The state space of $-i$'s automaton will be $[\underline{v}_i, \bar{v}_i]$.

Initial state: player $-i$ starts in state $u = v_i$, the payoff to be achieved.

Actions: in state u, player $-i$ picks strategy s_{-i}^C with probability q and picks s_{-i}^D with probability $1-q$, where q solves $u = q\bar{v}_i + (1-q)\,\underline{v}_i$.

Transitions: If player $-i$ chose s_{-i}^D, and his history is h_{-i}^T, then at the end of the round he will transit to state $\underline{v}_i + (1-\delta)\,\pi_i^D\left(h_{-i}^T\right)$, which is in $[\underline{v}_i, \bar{v}_i]$ for δ close enough to one. If player $-i$ chose s_{-i}^C, and his history is h_{-i}^T, then at the end of the round he will transit to state $\bar{v}_i + (1-\delta)\,\pi_i^C\left(h_{-i}^T\right)$, which is in $[\underline{v}_i, \bar{v}_i]$ for δ close enough to one.

We claim that these strategies constitute an equilibrium. First, observe that player $-i$ is indifferent across all strategies when i plays s_i^D, and in particular both s_{-i}^D and s_{-i}^C are optimal. On the other hand, if i plays s_i^C, both strategies are optimal by the result of the lemma. The overall optimality follows from the one-shot deviation principle (from one round to the next). By construction, the desired payoff is achieved.

3 Proof of Lemma 2

We first show that, given τ, T, any termination payoff $\bar{\pi}_i : H_{-i}^T \to \mathbb{R}$, and any strategy $s_{-i}^C \in S_{-i}^T$ such that $s_{-i}^C\left[h_{-i}^t\right](C) = 1-\varepsilon$ for any h_{-i}^t, $t\le\tau$, for some small $\varepsilon > 0$, we can define a modified termination payoff $\pi_i : H_{-i}^T \to \mathbb{R}$ such that i is indifferent between both actions for any h_i^t, $t\le\tau$. Given h_i^t and s_{-i}^C, let $V_i\left(h_i^t\right)$ denote i's (expected) continuation payoff, after history h_i^t, $t\le T$, including the termination payoff $\bar{\pi}_i$, and let $V_i\left(h_i^t, a_i^t\right)$ denote this continuation payoff, conditional on action a_i^t in period t. We claim that we can recursively define payments $\theta\left(h_{-i}^t, \omega_{-i}\right)$, starting from period $t=\tau$, and going backward to $t=1$, satisfying the desired property. In fact, we can take $\theta\left(h_{-i}^t, \omega_{-i}\right) = \theta\left(h_{-i}^t, \tilde{\omega}_{-i}\right) =: \theta\left(h_{-i}^t, \not\omega'_{-i}\right)$, for $\omega_{-i}\neq\omega'_{-i}$, $\tilde{\omega}_{-i}\neq\omega'_{-i}$. In period $t=\tau$, we solve:

$$\begin{aligned} & (1-\delta)\,u_C + \delta V\left(h_i^t, C\right) + \sum_{h_{-i}^t} m\left(h_{-i}^t|h_i^t\right)\left(p\theta\left(h_{-i}^t,\omega'_{-i}\right) + (1-p)\,\theta\left(h_{-i}^t,\not\omega'_{-i}\right)\right) \\ = \; & (1-\delta)\,u_D + \delta V\left(h_i^t, D\right) + \sum_{h_{-i}^t} m\left(h_{-i}^t|h_i^t\right)\left(q\theta\left(h_{-i}^t,\omega'_{-i}\right) + (1-q)\,\theta\left(h_{-i}^t,\not\omega'_{-i}\right)\right), \end{aligned}$$

where u_C (u_D) is the expected reward in period t from playing C (resp. D) (observe that $s_{-i}^C\left[h_{-i}^\tau\right]$ is independent of h_{-i}^τ), p (resp. q) is the probability of signal ω'_{-i} given that i uses C (resp. D) and $-i$ uses $s_{-i}^C\left[h_{-i}^\tau\right]$. This can be done as long as the matrix

$$M^t = \left(m\left(h_{-i}^t|h_i^t\right)\right)_{h_i^t, h_{-i}^t}$$

with rows corresponding to h_i^t and columns corresponding to h_{-i}^t, is invertible for $t = \tau$. We show that this is true for all $t \leq \tau$.

Given some history of length $t < \tau$, observe that

$$\Pr\left\{a_{-i}, \omega_{-i}, h_{-i}^t | a_i, \omega_i, h_i^t\right\} = \Pr\left\{a_{-i}, \omega_{-i} | a_i, \omega_i\right\} \Pr\left\{h_{-i}^t | h_i^t\right\}.$$

Therefore, M^{t+1} is invertible if M^t is, since we can partition M^{t+1} as

$$\begin{array}{cccccccc}
 & \overbrace{\qquad}^{C\omega_{-i}^1 h_{-i}^t} & \overbrace{\qquad}^{C\omega_{-i}^2 h_{-i}^t} & \cdots & \overbrace{\qquad}^{D\omega_{-i}^1 h_{-i}^t} & \overbrace{\qquad}^{D\omega_{-i}^2 h_{-i}^t} & \cdots \\
C\omega_i^1 h_i^t\{ & \Pr\left\{C\omega_{-i}^1 | C\omega_i^1\right\} M^t & \Pr\left\{C\omega_{-i}^2 | C\omega_i^1\right\} M^t & \cdots & \vdots & \vdots & \cdots \\
C\omega_i^2 h_i^t\{ & \Pr\left\{C\omega_{-i}^1 | C\omega_i^2\right\} M^t & \Pr\left\{C\omega_{-i}^2 | C\omega_i^2\right\} M^t & \cdots & \vdots & \vdots & \cdots \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots \\
D\omega_i^1 h_i^t\{ & \Pr\left\{C\omega_{-i}^1 | D\omega_i^1\right\} M^t & \Pr\left\{C\omega_{-i}^2 | D\omega_i^1\right\} M^t & \cdots & \vdots & \vdots & \cdots \\
D\omega_i^2 h_i^t\{ & \Pr\left\{C\omega_{-i}^1 | D\omega_i^2\right\} M^t & \Pr\left\{C\omega_{-i}^2 | D\omega_i^2\right\} M^t & \cdots & \vdots & \vdots & \cdots \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots
\end{array}.$$

That is, $M^{t+1} = M^1 \otimes M^t$, where $\otimes$ is the tensor product (where $M^0 = 1$), so that M^{t+1} is invertible and $\left(M^{t+1}\right)^{-1} = \left(M^1\right)^{-1} \otimes \left(M^t\right)^{-1}$ so the result follows from a recursion, along with Assumption 1 (Since Assumption 1 holds for some totally mixed action, it holds for all but finitely many totally mixed actions, so we pick ε so that M^1 is invertible).

Without loss of generality, we pick $\theta\left(h_{-i}^\tau, \omega'_{-i}\right) \leq 0$, $\theta\left(h_{-i}^\tau, \not\omega'_{-i}\right) \leq 0$, with one inequality holding with equality.

We may now define recursively $\theta\left(h_{-i}^t, \omega_{-i}\right)$ for all $t \leq \tau$. For $t < \tau$, define $\tilde{V}_i\left(h_i^t\right)$ denote i's (expected) continuation payoff, after history h_i^t, $t \leq T$, including the termination payoff $\bar{\pi}_i$ *and* including additional payments $\theta\left(h_{-i}^{t'}, \omega_{-i}\right)$ (received in period t') for all $t < t' \leq \tau$; and let $\tilde{V}_i\left(h_i^t, a_i^t\right)$ denote this continuation payoff, conditional on action a_i^t in period t. Replacing V_i by $\tilde{V}_i$, we can solve the equation in period $t < \tau$ analogous to the one for $t = \tau$, to otbain a solution $\theta\left(h_{-i}^t, \omega'_{-i}\right) \leq 0$, $\theta\left(h_{-i}^t, \not\omega'_{-i}\right) \leq 0$, with one inequality holding with equality.

Finally, we can replace $\bar{\pi}_i$ by π_i so that $\pi_i\left(h_{-i}^T\right) = \bar{\pi}_i\left(h_{-i}^T\right) + \delta^{-T}\sum_{t=1}^{\tau}\delta^t\theta\left(h_{-i}^t, \omega_{-i}\right)$, with $\left(h_{-i}^t, \omega_{-i}\right)$ corresponding to h_{-i}^T. In this fashion, we can include these additional payoffs into the termination payoff. Observe also that, for fixed τ, if $\left|V_i\left(h_i^{\tau+1}\right) - V_i\left(h_i'^{\tau+1}\right)\right| < K$ for all $h_i^{\tau+1}$ and $h_i'^{\tau+1}$, for some constant K independent of T, then $\sum_{t=1}^{\tau}\delta^t\theta\left(h_{-i}^t, \omega_{-i}\right)$ is also bounded above, independently of T. End of first step.

TO BE COMPLETED

References

[BO02] V. Bhaskar and I. Obara. Belief-based equilibria in the repeated prisoners' dilemma with private monitoring. *Journal of Economic Theory*, 102:40–70, 2002.

[EV02] J.C. Ely and J. Välimaäki. A robust folk theorem for the prisoner's dilemma. *Journal of Economic Theory*, 102:84–106, 2002.

[HO06] J. Hörner and W. Olszewski. The folk theorem with private almost-perfect monitoring. *Econometrica*, 2006.

[Kan02] M. Kandori. Introduction to repeated games with private monitoring. *Journal of Economic Theory*, 102:1–16, 2002.

[MM] G. Mailath and S. Morris. Coordination failure in repeated games with almost public monitoring. *Theoretical Economics*.

[MM02] G. Mailath and S. Morris. Repeated games with almost-public monitoring. *Journal of Economic Theory*, 102:189–229, 2002.

[Pic02] M. Piccione. The repeated prisoner's dilemma with imperfect private monitoring. *Journal of Economic Theory*, 102:70–84, 2002.

[Sek97] T. Sekiguchi. Efficiency in the repeated prisoner's dilemma with private monitoring. *Journal of Economic Theory*, 76:345–361, 1997.