

Efficiency in a Repeated Prisoners' Dilemma with Imperfect Private Monitoring.

Kyna Fong* Olivier Gossner† Johannes Hörner‡ Yuliy Sannikov§

March 1, 2007

Abstract

We study the repeated two-player Prisoners' Dilemma with imperfect private monitoring and no communication. Letting the discount factor go to one and holding the monitoring structure fixed, we achieve asymptotic efficiency. Unlike previous works on private monitoring, which have confined attention to signals that are either almost perfect or conditionally independent, we allow for both imperfect and correlated signals but assume that they are *sufficiently private*, i.e. private actions are more informative than private signals about the opponent's signals. Interestingly, for the game we study, even the existing literature that allows communication has not yet yielded efficiency.

1 Introduction.

A central motivation in the study of infinitely repeated games is to explore the possibility of cooperation in long-term relationships. Cooperation is difficult to achieve when each player can only observe other players' actions with noise. For example, the classical work of Stigler (1964) points out this difficulty in an oligopoly with "secret price cutting." After announcing a posted price, each firm may offer secret discounts to clients. Although those

*Stanford Graduate School of Business, contact: fong.kyna at gsb.stanford.edu.

†Northwestern University, contact: o-gossner at kellogg.northwestern.edu.

‡Northwestern University, contact: j-horner at kellogg.northwestern.edu.

§UC Berkeley, contact: sannikov at econ.berkeley.edu.

actual prices are not observable, each firm may use its own sales level as a private (but imperfect) signal of other firms' pricing behavior, because sales depend on both actual prices and unobservable random shocks to demand. For such games, in which players observe imperfect private signals about the actions of other players, the possibility of cooperation has remained an open question.

An extensive literature has established Folk Theorems under special assumptions. Fudenberg, Levine and Maskin (1994) provide an extremely clear answer for games with public monitoring, in which players' signals are common knowledge. Hörner and Olszewski (2006*a*) tackle the case of private monitoring where observations are nearly perfect. Matushima (2004) establishes the result for imperfect private monitoring assuming that players' signals are conditionally independent.¹

This paper is the first to establish the possibility of cooperation without narrowing attention to public, almost perfect, or conditionally independent signals. We prove that efficiency is attainable in the repeated two-player Prisoners' Dilemma with private monitoring, where we allow for truly imperfect, truly correlated private signals and assume that there is no communication. Our construction relies on two fairly mild assumptions. First, we assume that when a player is cooperating, his signals about the actions of his opponent are sufficiently private. That is, the arrival of each signal depends more strongly on the opponent's action than on the opponent's private signal. Second, we assume that when a player is defecting, at least one signal has a sufficiently high likelihood ratio for a test of his opponent's cooperation.

That such a result has remained inaccessible so far is not surprising. Unlike public monitoring, private monitoring eliminates common information that allows players to coordinate actions; as a result, cooperation becomes significantly harder. Players have to draw complex statistical inferences about other players' private histories to anticipate behavior. See Kandori (2002) or the excellent book of Mailath and Samuelson (2006) for a survey of the issues surrounding private monitoring.

In dealing with private monitoring that is *almost perfect*, the belief-free approach has been effective (Piccione (2002), Ely and Valimaki (2002), and Ely, Hörner and Olszewski (2005)). By setting continuation payoffs so that players are indifferent between their actions in every period, belief-free equilibria make inference about the past history of play completely irrelevant. Such a construction establishes the Folk Theorem for the repeated Prisoners' Dilemma with almost-perfect private monitoring (Ely and Valimaki (2002)). Hörner and Olszewski (2006*a*) extend the result to general games with almost perfect monitoring using equilibria that are belief-free over multi-period review phases.

¹Recent work by Hörner and Olszewski (2006*b*) shows that the Folk Theorem is robust also to imperfect private monitoring that is almost public.

When private monitoring is imperfect, efficient equilibria require aggregation of information across periods to minimize the probability of punishments on the equilibrium path. Matsushima (2004) overcomes this challenge for games in which signals are independent across players. In his construction, players aggregate information about the behavior of their opponents over T -period review phases. The equilibrium is belief-free across review phases. At the end of every phase, each player performs a statistical test of the signals observed during that review phase to see whether his opponent has cooperated or defected. Information indicating cooperative behavior leads to expected rewards, while information indicating defective behavior leads to expected punishments. Matsushima’s technique is in the spirit of Abreu, Milgrom and Pearce (1991), who use statistical discrimination through delayed information release to achieve efficiency in repeated interactions.

Matsushima’s construction, however, falls prey to a significant difficulty, which he averts by requiring signals to be conditionally independent. When signals are independent across players, a player receives no feedback during a review phase about how well he is doing (besides knowing his own history of actions); in other words, his incentives are not altered by the information he receives during the review phase.

In our game, however, we allow signals to be correlated. So each player receives private feedback during a review phase not only about what actions his opponent has taken but also about what signals his opponent has received. The player uses that feedback to draw private inferences about the recent history. What becomes important then is not only creating incentives for a player to choose a specific strategy at the beginning of a review phase but also creating incentives for him to continue following that strategy during a review phase.

Here is precisely where our paper takes a significant step forward from the existing literature: our equilibrium sustains incentives for players to cooperate even when they are receiving private, imperfect and correlated signals. A key novelty lies in the specification of “reward functions” or promised future payoffs. Those are constructed so that, in equilibrium, when review phases are long enough, all the way until a player’s own incentives to cooperate break down (which itself is unlikely to happen), he places a very low probability on the breakdown of the other player’s incentives to cooperate.

In each T -period review phase of our equilibrium, players maintain private counts of the number of “good” signals received, i.e. signals indicating cooperative behavior. Let these private counts have positive correlation $\rho \in (0, 1)$. When both players are cooperating, a player’s reward is strictly increasing in his opponent’s private count until that count reaches some unexpectedly high level; at that point, the player obtains his maximum reward. So if a player observes a high enough private count to suggest that his maximum reward has been reached (an event that occurs *ex ante* with low probability), his incentives to cooperate break down.

A critical insight in this paper is that when one player observes an exceptionally high private count of good signals, say n standard deviations above the mean, he expects that his opponent's private count is only ρn standard deviations above the mean. Therefore, the fact that the opponent may defect when he sees enough good signals does not cause a player to stop cooperating, even with imperfect signals. So long as a player's own private count tells him to continue cooperating, he places a very low probability on his opponent's private count being so high as to induce defection; therefore, the player's incentives to cooperate are maintained.

We finish the introduction by reviewing other approaches to attaining efficiency in games with private monitoring. Several papers have focused on belief-based, rather than belief-free, techniques. In such equilibria, players' strategies involve statistical inference about the past history of play (see Sekiguchi (1997), Bhaskar and Obara (2002), and Mailath and Morris (2002)). Results in those papers are limited to almost perfect or almost public monitoring.

Another closely related literature has been fairly successful in analyzing repeated games with private imperfect monitoring by introducing communication. Versions of the Folk Theorem have been proven (see Compte (1998), Kandori and Matsushima (1998), Aoyagi (2002), Fudenberg and Levine (2002), and Obara (2007)). Introducing a public element allows these papers to sidestep the inherent issues unresolved in games with private monitoring. However, the analyses do not apply to some practical economic settings in which communication is not possible - for example, in Stigler's oligopoly example above, anti-trust laws make communication illegal.

Interestingly, in the game we study, the two-player repeated Prisoners' Dilemma with imperfect, correlated, private signals, efficiency has not yet been established even when communication is allowed. So ours is the first work to establish efficiency, with or without communication.

We organize the paper as follows. Section 2 introduces the model and states the main result. Section 3 presents a brief overview of our argument, and readers are encouraged to read this section to understand the basic theoretical ideas behind the equilibrium construction. Section 4 presents the formal proofs. Section 5 discusses how the methods of this paper extend to general games.

2 The Model.

We consider an infinitely repeated Prisoners' Dilemma with private monitoring and no communication. Each player $i = 1, 2$ chooses an action $a_t^i \in A^i = \{D^i, C^i\}$ in every period

$t \geq 0$. Players do not observe each other's actions directly. Instead, at the end of each period, player i observes a private signal $y_t^i \in Y^i$.

For each action profile, every profile of private signals realizes with a strictly positive joint probability, which we denote by $\pi(y^i, y^j | a^i a^j)$.² We use $\pi(y^i | a^i a^j)$ to denote the marginal probability that player i receives signal y^i , and $\pi(y^i | a^i a^j, y^j)$ to denote the conditional probability that player i receives signal y^i given that player j receives signal y^j .

For simplicity, we assume that there are two signals $Y^i = \{0^i, 1^i\}$, where 1^i represents good news. Furthermore, we assume that the function

$$\pi(1^i | a^i a^j, y^j)$$

is strictly increasing in all three of its arguments (where $C^i > D^i$ for $i \in \{1, 2\}$). Therefore, good signals are positively correlated in our game. In Section 5 we discuss how our analysis can be extended to other signal structures and general games.

Denote by $g^i(a^i a^j)$ the expected stage-game payoff of player i .³ The stage-game payoffs are those of a Prisoners' Dilemma, i.e.

$$g^i(D^i C^j) > g^i(C^i C^j) > g^i(D^i D^j) > g^i(C^i D^j).$$

Players discount future payoffs at a common rate δ . So, player i 's expected payoff in the repeated game is

$$E \left[\sum_{t=0}^{\infty} \delta^t g^i(a_t^i a_t^j) \right]. \quad (1)$$

A t -period private history of player i is a sequence of private actions and private signals, denoted by $h_t^i = (a_0^i, y_0^i, a_1^i, y_1^i \dots a_{t-1}^i, y_{t-1}^i)$. Let H_t be the set of t -period private histories. A private strategy of player i is a function $\alpha^i : \bigcup_{t=0}^{\infty} H_t \rightarrow [0, 1]$ that gives the probability with which player i plays C^i after each private history $h_t^i \in H_t$ for all $t \geq 0$.

A profile of private strategies (α^1, α^2) forms a *Nash equilibrium* of the repeated game if each player's strategy maximizes his expected payoff (1) given the strategy of his opponent. A strategy profile (α^1, α^2) is a *Perfect Bayesian Equilibrium* of the repeated game if each player's strategy maximizes the conditional expectation of his payoff

$$E \left[\sum_{t=0}^{\infty} \delta^t g^i(a_t^i a_t^j) \mid h_t^i \right],$$

²In this paper, whenever we refer to players i and j , we assume $i \in \{1, 2\}$ and $i \neq j$.

³As usual, the actual payoff realization of player i in a given period depends on his action and his private signal.

for any private history h_t^i . Since our game satisfies the full support assumption (i.e. each profile of signals realizes with a positive probability for each profile of actions), any Nash equilibrium can be converted to a Perfect Bayesian Equilibrium by modifying the players' actions after zero-probability private histories.

We construct equilibria that converge to full cooperation as $\delta \rightarrow 1$. Our proof relies on the following two assumptions on the information structure of the game.

Assumption 1. *When player j is cooperating, signals are sufficiently private, so that*

$$\pi(1^j | C^j C^i, 0^i) > \pi(1^j | C^j D^i, 1^i).$$

That is, player i 's action matters more than his signal for player j 's signal.

Assumption 2. *When player j is defecting, given payoffs $g^i(\cdot)$, signal 1^j has a sufficiently high likelihood ratio to test for player i 's cooperation, so that*

$$\frac{\pi(1^j | D^j C^i)}{\pi(1^j | D^j D^i)} > \frac{g^i(C^i C^j) - g^i(C^i D^j)}{g^i(C^i C^j) - g^i(D^i D^j)}.$$

We now state our main result.

Theorem 1. *Under Assumptions 1 and 2, there exists a Perfect Bayesian Equilibrium in which players' expected payoffs become arbitrarily close to efficient as δ approaches 1.*

3 An Overview of the Argument.

This section intuitively explains our construction of equilibria that approach full cooperation as the discount factor goes to 1. While our formal proofs in Section 4 consider an asymmetric game, it is easier to think of a symmetric game to understand the main logic.

For each δ , the equilibrium is based on review phases of length T . To be specific, we let $T = O((1 - \delta)^{-1/2})$ so that

$$T \rightarrow \infty \quad \text{and} \quad \delta^T \rightarrow 1 \quad \text{as} \quad \delta \rightarrow 1.$$

Longer review phases allow for better aggregation of information. Unlike games where monitoring is almost perfect,⁴ more general games of imperfect monitoring require information aggregation to reduce inefficient punishments that occur on the equilibrium path.

⁴See, for example, Sekiguchi (1997), Bhaskar and Obara (2002), Ely and Valimaki (2002), Mailath and Morris (2002), or Hörner and Olszewski (2006a).

At the same time, it is important that δ^T converge to 1 as $\delta \rightarrow 1$ to allow for a wide range of rewards and punishments at the end of a review phase.

Incentives arise from an equilibrium structure, in which at the beginning of every review phase each player i is indifferent between two payoff-maximizing T -period strategies: \hat{C}^i and \hat{D}^i . Strategy \hat{C}^i involves cooperation in all T periods with probability near 1, and strategy \hat{D}^i consists of defection in all T periods. Player i creates incentives for his opponent through a transition rule that determines which strategy, \hat{D}^i or \hat{C}^i , is chosen in the next review phase. The transition rule depends on (1) player i 's strategy during the last review phase, and (2) his private history during the last review phase. Effectively, the transition rule implements a *reward function* for the review phase, provided by player i at the end of each review phase to reward or punish the perceived behavior of his opponent. This reward function thus creates incentives across review phases.⁵ Denote by $[G_D^j, G_C^j]$ the range of rewards and punishments that can be assigned to player j by player i 's mixing between \hat{D}^i and \hat{C}^i .

To ensure that both \hat{C}^i and \hat{D}^i are optimal to choose at the beginning of every review phase, we construct strategies and transition rules with the following two properties:

1. When the opponent is playing \hat{D}^j for sure, player i is indifferent between *any* sequence of actions.
2. When the opponent is playing \hat{C}^j for sure, both strategies \hat{C}^i and \hat{D}^i are optimal.

Therefore, at the beginning of each review phase, for any belief of player i about the strategy player j will follow, strategies \hat{C}^i and \hat{D}^i are both payoff-maximizing strategies of player i , between which he is indifferent.⁶

Players attain payoffs arbitrarily close to full cooperation as $\delta \rightarrow 1$ because with probability near 1 the players again play \hat{C}^1 and \hat{C}^2 at the end of a review phase in which they have just played \hat{C}^1 and \hat{C}^2 . Information aggregation by both players during the review phase makes this feature possible.

Let us discuss the strategies and reward functions in greater detail, beginning with the easier one, \hat{D}^j . This strategy consists of defection in all T periods. When player j follows \hat{D}^j , he rewards player i with an amount K_D^i for each good signal, which makes i just indifferent between cooperating and defecting in each period. The value of K_D^i is defined

⁵Note that the construction of reward functions based on aggregated information in a review phase is similar to the technique used in Matsushima (2004).

⁶So our construction is belief-free across review phases, as in Hörner and Olszewski (2006a) and Matsushima (2004).

by

$$g^i(C^i D^j) + \pi(1^j | D^j C^i) K_D^i = g^i(D^i D^j) + \pi(1^j | D^j D^i) K_D^i. \quad (2)$$

Formally, to take discounting into account, the reward function associated with strategy \hat{D}^j that is awarded at time T is

$$W_D^j(h_T^j) = G_D + K_D^i \sum_{t=0}^{T-1} \delta^{t-T} y_t^j, \quad (3)$$

where h_T^j is player j 's private history during the review phase.

Proposition 1. *Suppose that player j is playing the strategy \hat{D}^j of defecting in every period and assigns to player i the reward function W_D^j . Then player i is indifferent between all T -period strategies. Furthermore, we have*

$$(1 - \delta)G_D^i = g^i(D^i D^j) + K_D^i \pi(1^j | D^j D^i) < g^i(C^i C^j)$$

when Assumption 2 holds.

Proof. See Appendix. □

Let us turn our attention to strategy \hat{C}^j and its reward function W_C^j . First, define the discounted sum of good signals that player j has observed at time t by

$$X(h_t^j) = \sum_{s=0}^{t-1} \delta^s y_s^j. \quad (4)$$

By the central limit theorem, when both players are cooperating, $X(h_t^j)$ is distributed approximately normally with mean and variance

$$\sum_{s=0}^{t-1} \delta^s \pi(1^j | C^j C^i) \quad \text{and} \quad \sum_{s=0}^{t-1} \delta^{2s} \pi(1^j | C^j C^i) (1 - \pi(1^j | C^j C^i)) \quad (5)$$

respectively when t is large. Define by Φ_t^j the event that

$$X(h_s^j) \leq \sum_{u=0}^{s-1} \delta^u \pi(1^j | C^j C^i) + T^{1/6} \sigma_C^j \quad \forall s \leq t, \quad (6)$$

where $\sigma_C^j = O(T^{1/2})$ is the standard deviation of $X(h_T^j)$ under cooperation by both players (see (5)). Player j observes a private history $h_t^j \notin \Phi_t^j$ only when $X(h_t^j)$ is $O(T^{1/6})$ standard deviations above its mean. Therefore, when T is large the probability of Φ_t^j is close to 1.

We can now describe the strategy \hat{C}^j together with its reward function W_C^j . Strategy \hat{C}^j prescribes cooperation in period t when $h_t^j \in \Phi_t^j$ and is determined by a fixed-point argument when $h_t^j \notin \Phi_t^j$. When $h_T^j \in \Phi_T^j$, the reward function W_C^j is convex and quadratic in $X(h_T^j)$, satisfying

$$W_C^j(h_T^j) = G_C^i \quad \text{when} \quad X(h_T^j) = \sum_{t=0}^{T-1} \delta^t \pi(1^j | C^j C^i) + T^{1/6} \sigma_C^j. \quad (7)$$

When $h_T^j \notin \Phi_T^j$, W_C^j is of the form that makes player i indifferent between all actions, independently of player j 's actions. Section 4 provides that formula for W_C^j when $h_T^j \notin \Phi_T^j$.

Let us argue informally that such strategies \hat{C}^j and reward functions W_C^j

1. attain efficiency as $\delta \rightarrow 1$ and
2. fit our equilibrium construction for appropriate parameters of the quadratic portion of W_C^j .

To streamline our argument, we introduce the following definition, which we also use in Section 4.

Definition 1. *A T -period strategy of player j is from class Z^j if cooperation is always prescribed in period t in the event Φ_t^j .*

If we ignore the concern of incentives, strategies $\hat{C}^j \in Z^j$ together with reward functions W_C^j described above attain efficiency as $\delta \rightarrow 1$ and $T \rightarrow \infty$ for two reasons. First, the loss of efficiency due to the possibility of defection within a review phase becomes small, since the probability of Φ_t^j is close to 1 when T is large. Second, the loss of efficiency due to transitions to strategy \hat{D}^j at the end of a review phase also becomes very small as T gets large. Indeed, (7) implies that $W_C^j(h_T^j) = G_C^i - O(T^{2/3})$ on average, since (a) $\sigma_C^j = O(T^{1/2})$, (b) $\sum_{t=0}^{T-1} \delta^t \pi(1^j | C^j C^i)$ is the mean of $X(h_T^j)$ under cooperation, and (c) the slope of W_C^j is on the order of a constant (see Section 4). Thus, the loss of efficiency is only $O(T^{2/3})$ per T periods.

Now what about incentives? Let us argue informally that it is optimal for player i to follow either \hat{D}^i or a strategy from class Z^i in response to any strategy $\hat{C}^j \in Z^j$ of player j together with a reward function W_C^j as outlined above. Our argument proceeds in three steps.

First, suppose that player j *always* cooperates and that W_C^j is convex and quadratic in $X(h_T^j)$ for *all* histories h_T^j . Let us explore player i 's incentives. With a quadratic reward function, player i 's marginal expected reward from cooperating in period t is linearly

increasing in

$$E[X(h_t^j) | h_t^i] = \sum_{s=0}^{t-1} \delta^s \pi(1^j | C^j C^i, y_s^i).$$

Since Assumption 1 implies that

$$\pi(1^j | C^j C^i, y^i) > \pi(1^j | C^j D^i, z^i) \quad \forall y^i, z^i \in \{0^i, 1^i\},$$

past cooperation always makes future cooperation more attractive to player i for any sequence of private signals. Similarly, past defection makes future defection more attractive. Therefore, faced with a quadratic reward function W_C^j , player i optimally either always cooperates or always defects, but nothing in between.

Now, second, suppose that player j still *always* cooperates but that W_C^j takes the form originally specified. So it is only convex and quadratic for histories $h_T^j \in \Phi_T^j$. Let us argue that the amendment of the reward function W_C^j for $h_T^j \notin \Phi_T^j$ has only negligible effects on player i 's incentives when $h_t^i \in \Phi_t^i$. This conclusion follows from the following lemma.

Lemma 1. *If player j cooperates in every period of the review phase and $h_t^i \in \Phi_t^i$, then*

$$Pr[-\Phi_T^j | h_t^i] \leq T \exp(-K_j T^{1/3}), \quad (8)$$

where

$$K_j = -2\pi(1^j | C^j C^i)[1 - \pi(1^j | C^j C^i)][1 - \pi(1^j | C^j C^i, 1^i) + \pi(1^j | C^j C^i, 0^i)]^2. \quad (9)$$

Proof. See Appendix. □

Note that $T \exp(-K_j T^{1/3})$ decays exponentially as T goes to infinity. Thus, the probability that $h_T^j \notin \Phi_T^j$ is negligible when $h_t^i \in \Phi_t^i$.

To understand intuitively why Lemma 1 is true, consider player i 's belief about $X(h_t^j)$ conditional on $h_t^i \in \Phi_t^i$. For a given time- t history of player i 's private signals, $X(h_t^j)$ is expected to take the largest values when player i cooperates in every period. In that case, the joint distribution of $X(h_t^i)$ and $X(h_t^j)$ is approximately normal with some correlation $\rho < 1$. When

$$X(h_t^i) = \sum_{s=0}^{t-1} \delta^s \pi(1^i | C^i C^j) + T^{1/6} \sigma_C^i,$$

i 's conditional belief about the mean of $X(h_t^j)$ given h_t^i is

$$\sum_{s=0}^{t-1} \delta^s \pi(1^j | C^j C^i) + \rho T^{1/6} \sigma_C^j.$$

Since the conditional standard deviation of $X(h_t^j)$ is less than σ_C^j , it is extremely unlikely that

$$X(h_t^j) > \sum_{s=0}^{t-1} \delta^s \pi(1^j | C^j C^i) + T^{1/6} \sigma_C^j,$$

which is $O(T^{1/6})$ standard deviations above the conditional mean of $X(h_t^j)$. Thus the conditional probability of the event $\neg\Phi_T^j$ is very small.

Finally, as the third step in our argument, suppose that W_C^j continues to take the form originally specified but that player j may now defect after histories $h_t^j \notin \Phi_t^j$ (i.e. he plays a strategy from class Z^j). There arises the concern that player i 's inference about the relative likelihoods of Φ_t^j and $\neg\Phi_t^j$ may change. Therefore, Lemma 1 does not have to hold for a general strategy of player j from class Z^j . However, if the reward function that player j offers when $h_T^j \notin \Phi_T^j$ makes player i just indifferent between all actions, then the possibility of the event $\neg\Phi_t^j$ is irrelevant for player i 's incentives.⁷

We conclude that player i has incentives to always defect or to follow a strategy from class Z^i when player j follows a strategy from class Z^j and assigns a reward function W_C^j , where W_C^j is quadratic and convex if $h_T^j \in \Phi_T^j$ and makes player i indifferent between all strategies otherwise.

Section 4 formalizes the logic we have used in this section to motivate the construction of our equilibria. To complete our proof that efficiency is attainable as $\delta \rightarrow 1$, we provide a fixed-point argument to show that there exist strategies $\hat{C}^1 \in Z^1$, $\hat{C}^2 \in Z^2$ and parameters of the quadratic portions of W_C^1, W_C^2 such that for $i = 1, 2$ both strategies \hat{D}^i and \hat{C}^i maximize player i 's expected payoff in response to \hat{C}^j with the reward function W_C^j .

4 Formal Proofs.

In this section, we formally construct Perfect Bayesian Equilibria of the repeated Prisoners' Dilemma that approach efficiency as $\delta \rightarrow 1$. As discussed in Section 3, our construction is based on T -period review phases, where $T = O((1 - \delta)^{-1/2})$. In every review phase each player $i = 1, 2$ follows a T -period strategy \hat{C}^i or \hat{D}^i . Both of these strategies are optimal independently of the beliefs about the strategy of the opponent. While \hat{D}^i involves defection in every period, $\hat{C}^i \in Z^i$ involves cooperation in every period with probability near 1. A player's strategy choice creates equilibrium incentives for his opponent.

Proposition 2 shows that, to construct efficient equilibria along these lines, we just need to

⁷Note that this argument is similar to one made in Hörner and Olszewski (2006a).

find appropriate T -period strategies and reward functions.

Proposition 2. *Suppose that for $i = 1, 2$ and some $T \geq 0$, there are T -period strategies \hat{C}^i and \hat{D}^i and reward functions $W_C^i : H_T \rightarrow [G_D^j, G_C^j]$ and $W_D^i : H_T \rightarrow [G_D^j, G_C^j]$ that satisfy the following conditions.*

First, when player $j = 1, 2$ is following strategy \hat{C}^j , then

$$G_C^i = \max E \left[\sum_{s=0}^{T-1} \delta^s g^i(a_s^i a_s^j) + \delta^T W_C^j(h_T^j) \right], \quad (10)$$

where the maximum, taken over all T -period strategies of player i , is achieved by both \hat{C}^i and \hat{D}^i .

Second, when player j is following strategy \hat{D}^j , then

$$G_D^i = \max E \left[\sum_{s=0}^{T-1} \delta^s g^i(a_s^i a_s^j) + \delta^T W_D^j(h_T^j) \right], \quad (11)$$

where the maximum, taken over all T -period strategies of player i , is again achieved by both \hat{C}^i and \hat{D}^i .

Then any pair of payoffs $(w_1, w_2) \in [G_D^1, G_C^1] \times [G_D^2, G_C^2]$ is achievable by a Perfect Bayesian Equilibrium of an infinitely repeated game with discount factor δ .⁸

The proof in the Appendix provides a careful verification of these intuitive claims.

To construct equilibria using Proposition 2, for $i = 1, 2$, strategies \hat{C}^i and \hat{D}^i together with reward functions W_C^i and W_D^i have to satisfy two sets of constraints

1. *incentive constraints*, i.e. both \hat{C}^i and \hat{D}^i have to maximize (10) and (11) and
2. *feasibility constraints*, i.e. the reward functions must take values in the ranges $[G_D^j, G_C^j]$, which are defined by (10) and (11).

Patient players can attain efficiency in equilibrium if for $i = 1, 2$,

$$(1 - \delta)G_C^i \rightarrow g^i(C^i C^j) \quad \text{as } \delta \rightarrow 1. \quad (12)$$

⁸When we let $\delta \rightarrow 1$, for simplicity our notation suppresses the dependence of the strategies, reward functions and bounds G_D^i and G_C^i on δ .

Strategies \hat{D}^j and reward functions W_D^j . In Section 3 we show that given the strategy \hat{D}^j of player j together with a linear reward function

$$W_D^j(h_T^j) = G_D^i + \delta^{-T} K_D^i X(h_T^j), \quad \text{where} \quad (1 - \delta)G_D^i = g^i(D^i D^j) + K_D^i \pi(1^j | D^j D^i),$$

any strategy of player i maximizes (11). This reward function W_D^j makes player i indifferent between *all* T -period strategies. Moreover, when δ is sufficiently close to 1, so that $(1 - \delta)T$ is sufficiently small, Proposition 1 implies that

$$\max (1 - \delta)W_D^j(h_T^j) < (1 - \delta)(G_D^i + \delta^{-T} K_D^i T) < g^i(C^i C^j) - k^i,$$

for some constant $k^i > 0$. Therefore, if (12) holds, the reward function W_D^j also satisfies the feasibility constraints when T is sufficiently large.

Strategies \hat{C}^j and reward functions W_C^j . To complete the proof that equilibrium efficiency is attainable in the limit as $\delta \rightarrow 1$, for each δ and $i = 1, 2$ we need to construct strategies $\hat{C}^j \in Z^j$ and reward functions W_C^j that satisfy (10), the feasibility constraints and (12). This task is significantly more challenging than the construction of \hat{D}^j . Since our goal is to attain efficiency, the reward function W_C^j cannot make player i indifferent between all strategies, or else function W_C^j will destroy too much value to create those incentives.⁹ Also, because player j observes signals that are correlated with player i 's signals, strategy \hat{C}^j cannot involve cooperation in all periods (unlike in Matsushima (2004)).

Thus, we look for a suitable strategy \hat{C}^j from a whole class of strategies Z^j , and consider reward functions from the following class.

Definition 2. Denote

$$X_C^j = \sum_{t=0}^{T-1} \delta^t \pi(1^j | C^j C^i) + T^{1/6} \sigma_C^j.$$

For a pair of positive constants α^j and β^j that satisfy $\alpha^j X_C^j < \beta^j$, define¹⁰

$$W_C^j(h_T^j) = V^j(X(h_T^j)), \quad \text{where} \quad V^j(x) = G_C^i - \beta^j (X_C^j - x) + \alpha^j (X_C^j - x)^2 \quad (13)$$

when $h_T^j \in \Phi_T^j$ and

$$W_C^j(h_T^j) = U^j(h_T^j), \quad \text{where} \quad U^j(h_T^j) = G_C^i - \delta^{-T} \sum_{t=0}^{T-1} \delta^t K(a_t^j, y_t^j) \quad (14)$$

otherwise, where $K(C^j, 1^j) = 0$ and $K(a^j, y^j) \geq 0$ are such that

$$g^i(a^i a^j) - \pi(1^j | a^j a^i) K(a^j, 1^j) - \pi(0^j | a^j a^i) K(a^j, 0^j)$$

takes the same value for all $a \in \{D, C\}$.

⁹Such an equilibrium would be belief-free. As shown in Ely, Hörner and Olszewski (2005), the Folk Theorem does not hold in belief-free equilibria when monitoring is imperfect.

¹⁰When $\alpha^j X_C^j < \beta^j$, we have $V^j(x) \leq G_C^i$ for all $x \in [0, X_C^j]$.

Note that the range of W_C^j is $O(T)$. At the same time, Proposition 1 implies that $G_C^i - G_D^i = O(1/(1 - \delta)) = O(T^2)$ when the efficiency condition (12) holds. Therefore, function W_C^j satisfies the feasibility constraints for sufficiently large T under condition (12). We verify (12) in Proposition 5.

Proposition 3 shows that, under an additional assumption (15), when player j follows a strategy $\hat{C}^j \in Z^j$ and assigns a reward function W_C^j , player i has incentives to follow \hat{D}^i or a strategy from class Z^i , depending on α^j and β^j . Assumption (15) is not essential to our result. A remark at the end of this section argues that Proposition 3 remains valid without (15) if we implement a slightly more complicated form of W_C^j .

Proposition 3. *Assume that*

$$\pi(1^j|C^j C^i, 0^i) > \frac{\pi(1^j|C^j C^i) + \pi(1^j|C^j D^i)}{2} > \pi(1^j|C^j D^i, 1^i). \quad (15)$$

Suppose that player j is following a strategy $\hat{C}^j \in Z^j$ and assigns to player i a reward function $W_C^j(h_T^j)$ given by Definition 2, with $\alpha^j = O(T^{-2})$ and $\beta^j = O(\text{const})$. Let

$$\begin{aligned} M^j(t) = & X_C^j - \frac{\delta^t + \delta^T}{2(1 + \delta)} - \frac{\beta^j}{2\alpha^j} + \delta^{-T} \frac{g^i(D^i C^j) - g^i(C^i C^j)}{2\alpha^j(\pi(1^j|C^j C^i) - \pi(1^j|C^j D^i))} \\ & - \frac{\delta^{t+1} - \delta^T}{(1 - \delta)(1 + \delta)} (\pi(1^j|C^j C^i) + \pi(1^j|C^j D^i)) \end{aligned} \quad (16)$$

Then, when T is sufficiently large, it is optimal for player i to follow strategy \hat{D}^i or a strategy from class Z^i . When $M^j(0) > O(T^4 \exp(-K_j T^{1/3}))$ then the optimal strategy is from class Z^i ; when $M^j(0) < -O(T^4 \exp(-K_j T^{1/3}))$ then the optimal strategy is \hat{D}^i .

Let us summarize the three-step intuition behind our proof of Proposition 3, which appears in subsection 4.1. If player j always cooperate and assigns to player i the reward function V^j , then $M^j(t)$ is defined in such a way that at time t player i prefers to cooperate in all future periods if $E[X(h_t^j)|h_t^i] \geq M^j(t)$, and to defect in all future periods otherwise. If player j always cooperates but assigns a different reward function U^j in the event $-\Phi_T^j$, player i 's incentives remain roughly the same when $h_t^i \in \Phi_t^i$, because player i assigns a negligible probability to the event $-\Phi_T^j$. Finally, if player j defects in the event $-\Phi_t^j$, this may affect player i 's inferences about the relative probabilities of Φ_t^j and $-\Phi_t^j$, but does not affect player i 's incentives. The reason is that player i is indifferent between all strategies conditional on $-\Phi_t^j$. In subsection 4.1 we reach the conclusion that player i prefers to defect in all periods if $0 = E[X(h_0^j)] < M^j(0) - O(T^4 \exp(-K_j T^{1/3}))$ and follow a strategy from class Z^i if $0 = E[X(h_0^j)] > M^j(0) + O(T^4 \exp(-K_j T^{1/3}))$.

Note that when $\alpha^j = O(T^{-2})$, then $M^j(0) = 0$ when¹¹

$$\beta^j = K_C^i + O(T^{-1}), \quad \text{where} \quad K_C^i = \frac{g^i(D^i C^j) - g^i(C^i C^j)}{\pi(1^j | C^j C^i) - \pi(1^j | C^j D^i)}. \quad (17)$$

So, for the class of j 's strategies Z^j and reward functions that satisfy Definition 2, player i 's best response is either to play \hat{D}^i or a strategy from class Z^i . For some values of α^j and β^j , player i is just indifferent between playing \hat{D}^i and the best strategy from class Z^i , \hat{C}^i . If for each player $j = 1, 2$, strategy $\hat{C}^j \in Z^j$ and reward function W_C^j are chosen appropriately so that the opponent is indifferent between \hat{D}^i and \hat{C}^i , then the players' strategies and reward functions satisfy the conditions of Proposition 2. Those strategies can then be used to construct an equilibrium with any value pair from the set $[G_D^1, G_C^1] \times [G_D^2, G_C^2]$.

Proposition 4 shows that there exist T -period strategies \hat{C}^1, \hat{C}^2 and positive constants α^1, α^2 and β^1, β^2 such that, for $i = 1, 2$, both \hat{D}^i and \hat{C}^i are optimal in response to \hat{C}^j when the reward function W_C^j is given by Definition 2. We verify the existence of such strategies and reward functions via a fixed-point argument.

Proposition 4. *Let $\alpha^1 = \alpha^2 = T^{-2}$. Then for all sufficiently large T , there are reward functions W_C^i with $\beta^i = K_C^j + O(T^{-1})$ and T -period strategies \hat{C}^i from class Z^i for $i = 1, 2$, such that both \hat{D}^1 and \hat{C}^1 are best responses to \hat{C}^2 with the reward function W_C^2 , and similarly for \hat{D}^2 and \hat{C}^2 .*

Proof. See Appendix. □

To conclude, we must show that efficiency is attainable as $\delta \rightarrow 1$.

Proposition 5. *Suppose that $T = O((1 - \delta)^{-1/2})$. If player j follows a strategy \hat{C}^j from class Z^j and assigns a reward function W_C^j that satisfies Definition 2 with $\alpha = O(T^{-2})$ and $\beta = O(\text{const})$, then*

$$(1 - \delta)G_C^i \rightarrow g^i(C^i C^j). \quad (18)$$

Proof. See Appendix. □

¹¹Note the similarity of K_C^i to K_D^i defined in the last section.

4.1 The Proof of Proposition 3.

We use simplified notation in this proof, denoting

$$p(a^i) = \pi(1^j | C^j a^i) \quad \text{and}$$

$$p(a^i, y^i) = \pi(1^j | C^j a^i, y^i).$$

So when player j cooperates, $p(a^i)$ is the probability that j receives a good signal when player i chooses action a^i , and $p(a^i, y^i)$ is that probability conditional also on player i receiving signal y^i .

Let us show that for all $h_t^i \in \Phi_t^i$, it is optimal for player i to cooperate in period t if $\mu(h_t^i) > M^j(t) + O(T^4 \exp(-K_j T^{1/3}))$, where we define

$$\mu(h_t^i) = E[X(h_t^j) | h_t^i, \Phi_t^j] \quad \text{and} \quad \sigma^2(h_t^i) = \text{Var}[X(h_t^j) | h_t^i, \Phi_t^j].$$

Consider the case when $\mu(h_t^i) > M^j(t) + O(T^4 \exp(-K_j T^{1/3}))$. Conditional on the event $\neg \Phi_t^j$, player i 's future expected payoff does not depend on his actions. Therefore, to figure out his optimal actions, player i need only take into account his expected payoffs conditional on Φ_t^j .

Let us derive a *lower* bound for player i 's expected payoff (conditional on h_t^i and Φ_t^j) if he cooperates in period t and then acts optimally, and show that it is always greater than an *upper* bound for player i 's expected payoff (conditional on h_t^i and Φ_t^j) if he defects in period t .

A lower bound. Assume that player i cooperates in all future periods. Then, conditional on h_t^i and Φ_T^j , player i 's future expected payoff is

$$\frac{\delta^t - \delta^T}{1 - \delta} g^i(C^i C^j) + \delta^T E[V^j(X(h_T^j)) | h_t^i, \Phi_T^j] =$$

$$\frac{\delta^t - \delta^T}{1 - \delta} g^i(C^i C^j) + \delta^T V^j (E[X(h_T^j) | h_t^i, \Phi_T^j]) + \delta^T \alpha^j \text{Var}[X(h_T^j) | h_t^i, \Phi_T^j],$$

since $E[V^j(x)] = V^j(E[x]) + \alpha^j \text{Var}(x)$ for the quadratic function V^j . Since both V^j and U^j take values between G_C^i and $G_C^i - O(T)$ for various private histories h_T^j of player j , we can evaluate player i 's expected payoff conditional on h_t^i and Φ_T^j as

$$\frac{\delta^t - \delta^T}{1 - \delta} g^i(C^i C^j) + \delta^T V^j (E[X(h_T^j) | h_t^i, \Phi_T^j]) + \delta^T \alpha^j \text{Var}[X(h_T^j) | h_t^i, \Phi_T^j] + Pr[\neg \Phi_T^j | h_t^i, \Phi_t^j] O(T).$$

Lemma 2 shows the probability of event $\neg \Phi_T^j$, conditional on h_t^i and Φ_t^j , to be extremely small and evaluates $E[X(h_T^j) | h_t^i, \Phi_T^j]$ and $\text{Var}[X(h_T^j) | h_t^i, \Phi_T^j]$.

Lemma 2. *Suppose that player j is following a strategy from class Z^j . If $h_t^i \in \Phi_t^i$ and player i cooperates in the remaining periods $t, \dots, T-1$ of the review phase, then*

$$Pr[-\Phi_T^j \mid h_t^i, \Phi_t^j] \leq O(T \exp(-K_j T^{1/3})), \quad (19)$$

$$E[X(h_T^j) \mid h_t^i, \Phi_T^j] = \mu(h_t^i) + \frac{\delta^t - \delta^T}{1 - \delta} p(C^i) + O(T^2 \exp(-K_j T^{1/3})) \quad (20)$$

and

$$Var[X(h_T^j) \mid h_t^i, \Phi_T^j] = \sigma(h_t^i) + \frac{\delta^{2t} - \delta^{2T}}{1 - \delta^2} p(C^i)(1 - p(C^i)) + O(T^3 \exp(-K_j T^{1/3})). \quad (21)$$

Proof. See Appendix. □

Therefore, a lower bound on player i 's payoff if he cooperates in period t is

$$\begin{aligned} & \frac{\delta^t - \delta^T}{1 - \delta} g^i(C^i C^j) + \delta^T V^j \left(\mu(h_t^i) + \frac{\delta^t - \delta^T}{1 - \delta} p(C^i) \right) + \\ & \delta^T \alpha^j \left(\sigma^2(h_t^i) + \frac{\delta^{2t} - \delta^{2T}}{1 - \delta^2} p(C^i)(1 - p(C^i)) \right) + O(T^2 \exp(-K_j T^{1/3})), \end{aligned} \quad (22)$$

since $\alpha^j \leq O(T^{-2})$.

An upper bound. To bound from above player i 's payoff from defecting in period t , consider the following modification to player j 's strategy and reward function in the event Φ_t^j . Suppose that player j

1. cooperates from time t onward and
2. assigns to player i reward $V^j(X(h_T^j))$, even in the event $-\Phi_T^j$.¹²

Note that this modification may only improve player i 's payoff. Lemma 3 evaluates player i 's payoff with this modification.

Lemma 3. *Suppose that at time t player j cooperates in the remaining periods of the review phase and assigns to player i a convex quadratic reward function $V^j(X(h_T^j))$. Then it is*

¹²Such a reward function would violate the feasibility constraint, but we use it only to derive an upper bound on player i 's payoff if he defects in period t .

optimal for player i to cooperate in all remaining periods if $\mu(h_t^i) > M^j(t)$ and to defect in all remaining periods otherwise. Given this strategy, player i 's future expected payoff is

$$\hat{V}^i(t, \mu(h_t^i), \sigma^2(h_t^i)) = \max\{\hat{V}_{C^i}(t, \mu(h_t^i), \sigma^2(h_t^i)), \hat{V}_{D^i}(t, \mu(h_t^i), \sigma^2(h_t^i))\} \quad (23)$$

where for $a \in \{D^i, C^i\}$

$$\hat{V}_a(t, \mu, \sigma^2) = \delta^T V^j \left(\mu + \frac{\delta^t - \delta^T}{1 - \delta} p(a) \right) + \delta^T \alpha^j \left(\sigma^2 + \frac{\delta^{2t} - \delta^{2T}}{1 - \delta^2} p(a)(1 - p(a)) \right) + \frac{\delta^t - \delta^T}{1 - \delta} g^i(aC^j).$$

Proof. See Appendix. □

Lemma 3 implies that conditional on Φ_t^i , player i 's expected payoff from defecting once and acting optimally thereafter is bounded from above by

$$\begin{aligned} & \delta^t g^i(D^i C^j) + \pi(1^j | C^j D^i) \hat{V}^i(t, \mu(h_t^i) + \delta^t p(D^i, 1^i), \sigma^2(h_t^i) + \delta^{2t} p(D^i, 1^i)(1 - p(D^i, 1^i))) + \\ & \pi(0^j | C^j D^i) \hat{V}^i(t, \mu(h_t^i) + \delta^t p(D^i, 0^i), \sigma^2(h_t^i) + \delta^{2t} p(D^i, 0^i)(1 - p(D^i, 0^i))). \end{aligned} \quad (24)$$

Comparing payoffs from cooperating and defecting in period t . By Lemma 5 in the Appendix, payoff (24) is smaller than our lower bound

$$\hat{V}_{C^i}(t, \mu(h_t^i), \sigma^2(h_t^i)) + O(T^2 \exp(K_j T^{1/3})) \quad (25)$$

(see (22)) when

$$\delta^T \alpha^j (\mu(h_t^i) - M^j(t)) (p(C^i) - p(D^i)) \geq O(T^2 \exp(-K_j T^{1/3})). \quad (26)$$

Inequality (26) holds when $\mu(h_t^i) - M^j(t) > O(T^4 \exp(-K_j T^{1/3}))$, since $\alpha = O(T^{-2})$ and $\delta^T = O(1)$. We conclude that when $\mu(h_t^i) > M^j(t) + O(T^4 \exp(-K_j T^{1/3}))$, it is optimal for player i to cooperate in period t .

Assumption (15), i.e. $(p(D^i) + p(C^i))/2 \in (p(D^i, 1^i), p(C^i, 0^i))$ is required to show that for all δ sufficiently close to 1, if $\mu(h_t^i) > M^j(t) + O(T^4 \exp(-K_j T^{1/3}))$ then player i has incentives to cooperate in all future periods in the event Φ_T^i . Indeed, note that

$$M^j(t+1) - M^j(t) = \delta^t \frac{\delta}{1 + \delta} (p(D^i) + p(C^i)) + \delta^t \frac{1 - \delta}{1 + \delta} \in (\delta^t p(D^i, 1^i), \delta^t p(C^i, 0^i))$$

when δ is sufficiently close to 1 and $T = O((1 - \delta)^{-1/2})$. By Lemma 6 in the Appendix,

$$\mu(h_t^i) = \sum_{s=0}^{t-1} \delta^s p(a_s^i, y_s^i) + O(T^2 \exp(-K_j T^{1/3})).$$

Now, if $\mu(h_t^i) > M^j(t) + O(T^4 \exp(-K_j T^{\frac{1}{3}}))$ and player i cooperates in period t , then

$$\mu(h_{t+1}^i) = \mu(h_t^i) + \delta^t p(C^i, y^i) + O(T^2 \exp(-K_j T^{1/3})) > M^j(t+1) + O(T^4 \exp(-K_j T^{\frac{1}{3}}))$$

when δ is sufficiently close to 1. Therefore, if $M^j(0) < -O(T^4 \exp(-K_j T^{\frac{1}{3}}))$ it is optimal for player i to follow a strategy from class Z^i , i.e. cooperate in all future periods as long as $h_t^i \in \Phi_t^i$.

Similarly, we can show that if $M^j(0) > O(T^4 \exp(-K_j T^{\frac{1}{3}}))$, then it is optimal for player i to defect in all future periods. The central claim to reach this conclusion is that for all h_t^i such that player i has defected in the past, if $\mu(h_t^i) < M^j(t) - O(T^4 \exp(-K_j T^{1/3}))$, then it is optimal for player i to defect in period t . To justify this claim, we can derive an upper bound on player i 's expected payoff if he cooperates in period t using Lemma 3, and a lower bound if i defects in period t using an analogue of Lemma 2:

Lemma 4. *Suppose that player j is following a strategy from class Z^j . If h_t^i involves defection in all past periods and player i defects in the remaining periods $t, \dots, T-1$ of the review phase, then*

$$Pr[-\Phi_T^j | h_t^i, \Phi_t^j] \leq O(T \exp(-K_j T^{1/3})), \quad (27)$$

$$E[X(h_T^j) | h_t^i, \Phi_T^j] = \mu(h_t^i) + \frac{\delta^t - \delta^T}{1 - \delta} p(D^i) + O(T^2 \exp(-K_j T^{1/3})) \quad (28)$$

and

$$Var[X(h_T^j) | h_t^i, \Phi_T^j] = \sigma(h_t^i) + \frac{\delta^{2t} - \delta^{2T}}{1 - \delta^2} p(D^i)(1 - p(D^i)) + O(T^3 \exp(-K_j T^{1/3})). \quad (29)$$

Proof. See Appendix. □

This completes the proof of Proposition 3.

Remark. If assumption (15) fails, then Proposition 3 holds with a modified version of the reward function W_C^j . Specifically, in the event Φ_T^j we let

$$W_C^j(h_T^j) = V^j(X(h_T^j)) - L^j - \sum_{s=0}^{T-1} l_s^j y_s^j, \quad \text{where} \quad V^j(x) = G_C^i - \beta^j (X_C^j - x) + \alpha^j (X_C^j - x)^2$$

as before. Then

$$M^j(t+1) - M^j(t) = \frac{\delta^{t+1}}{(1+\delta)} (p(D^i) + p(C^i)) + \frac{1-\delta}{2\alpha^j} \left(\frac{\sum_{s=t+1}^{T-1} l_s^j}{\delta^{t+1} - \delta^T} - \frac{\sum_{s=t}^{T-1} l_s^j}{\delta^t - \delta^T} \right) + \delta^t \frac{1-\delta}{2(1+\delta)}.$$

We can ensure that

$$M^j(t+1) - M^j(t) = \delta^t \frac{p(D^i, 1^i) + p(C^i, 0^i)}{2} \in (\delta^t p(D^i, 1^i), \delta^t p(C^i, 0^i))$$

by choosing $L^j = O(const)$ and $l_s^j = O(T^{-1})$ appropriately.

5 Conclusions.

We show that patient players can attain efficiency in games with private monitoring even when signals are imperfect and correlated among players. Existing literature has found efficient equilibria only under the limiting assumptions of almost perfect, almost public or conditionally independent signals.

Let us summarize the main elements of our construction. Along the way, we comment on how our methods extend to general games and signal structures.

In our equilibrium, each player i is indifferent between two strategies \hat{D}^i and \hat{C}^i in every review phase, independently of the beliefs about his opponent's strategy. The indifference condition holds because of a reward function at the end of each review phase, implemented by the opponent's mixing between \hat{D}^i and \hat{C}^i .

Strategy \hat{D}^i involves defection in every period. When player i follows \hat{D}^i , he implements a reward function that is linear in the discounted count of good signals he receives during the review phase. This reward function makes player j indifferent between all strategies. This property is convenient, because at this point we have not constructed strategy \hat{C}^i yet.

A linear reward function rewards player j even if j defects in all periods of the review phase. Assumption 2 makes sure that such a function does not create too much value. For a general game, in which we attempt to sustain an efficient action profile (C^i, C^j) with the threat of a static Nash equilibrium (D^i, D^j) , there is a linear reward function that makes player j weakly prefer both D^j and C^j to all other actions during the review phase if j 's actions are identifiable through i 's signals. We need an analogue of Assumption 2 for a general game to make sure that the strategy \hat{D}^i destroys enough value.

Strategy \hat{C}^i involves cooperation in all periods in the event Φ_T^i , which occurs with probability nearly 1. In the event Φ_T^i , player i assigns a linear reward function V perturbed by a small quadratic term. The quadratic term ensures that player j prefers to continue defecting if he has defected once, and to continue cooperating if he has cooperated once (in the event Φ_T^j). With a general signal structure, such a construction is possible, for example, if player i has a *sufficiently private* signal y^i about j 's cooperation or defection,

such that

$$\pi(y^i|C^i, D^j, y^j) < \pi(y^i|C^i, C^j, \tilde{y}^j)$$

for any two signals y^j, \tilde{y}^j of player j . Then, if player j always cooperates, his beliefs about the discounted counts of i 's signals do not overlap with his beliefs if he always defects, independently of the private signals that i receives.

The event $\neg\Phi_T^i$ includes histories for which the reward function V exceeds the upper bound G_C^j on rewards that player i can assign. In the event $\neg\Phi_T^i$ we can define player i 's reward function in such a way that makes player j indifferent between all strategies, independently of the actions of player i . Then only rewards and punishments in the event Φ_T^i are relevant for the incentives of player j . To prevent the unraveling of incentives, player j must assign a very small probability to the event $\neg\Phi_T^i$ in the event Φ_T^j , when player i is cooperating in every period. Such a conclusion requires an additional assumption for general games. For our Prisoners' Dilemma, we can construct the events Φ_T^i and Φ_T^j appropriately because the good signals received by each player are positively correlated.

We conclude that the ideas of this paper extend directly to general games with private monitoring under appropriate assumptions. In fact, although the construction of efficient equilibria can be quite complex, attaining efficiency under weaker assumptions seems feasible since the ingredients of the construction are quite amenable to adjustments. We conjecture that the Folk Theorem holds for general games with imperfect private monitoring with full support, under only one assumption that each player can identify the actions of his opponent from his private signal.¹³

A Appendix.

A.1 Proof of Proposition 1.

Proof. For any T -period strategy of player i , his expected total payoff is

$$\begin{aligned} E \left[\sum_{s=0}^{T-1} \delta^s g^i(a_s^i D^j) + \delta^T W_D^j(h_T^j) \right] &= E \left[\sum_{s=0}^{T-1} \delta^s (g^i(a_s^i D^j) + K_D^i y_s^j) \right] + \delta^T G_D^i \\ &= \sum_{s=0}^{T-1} \delta^s (g^i(D^i D^j) + K_D^i \pi(1^j|D^i D^j)) + \delta^T G_D^i = G_D^i, \end{aligned}$$

¹³We say that signals have full support if each profile of private signals realizes with positive probability given any action profile.

since the expectation of $g^i(a_s^i D^j) + K_D^i y_s^j$ is the same regardless of whether player i cooperates or defects in period s (by the definition of K_D^i). Therefore, any T -period strategy of player i is an optimal response. Notice that

$$G_D^i = \frac{g^i(D^i D^j) + K_D^i \pi(1^j | D^j D^i)}{1 - \delta}.$$

Now, from Assumption 2, we have the following:

$$\begin{aligned} \pi(1^j | D^j D^i)(g^i(C^i C^j) - g^i(C^i D^j)) &< \pi(1^j | D^j C^i)(g^i(C^i C^j) - g^i(D^i D^j)) \Rightarrow \\ g^i(D^i D^j)\pi(1^j | D^j C^i) - g^i(C^i D^j)\pi(1^j | D^j D^i) &< g^i(C^i C^j)(\pi(1^j | D^j C^i) - \pi(1^j | D^j D^i)) \Rightarrow \\ G_D^i = \frac{g^i(D^i D^j) + K_D^i \pi(1^j | D^j D^i)}{1 - \delta} &= \frac{g^i(D^i D^j)\pi(1^j | D^j C^i) - g^i(C^i D^j)\pi(1^j | D^j D^i)}{(1 - \delta)(\pi(1^j | D^j C^i) - \pi(1^j | D^j D^i))} < \frac{g^i(C^i C^j)}{1 - \delta}. \end{aligned}$$

□

A.2 Proof of Lemma 1.

Proof. In this proof, we use simplified notation

$$p(a^i) = \pi(1^j | C^j a^i) \quad \text{and} \quad p(a^i, y^i) = \pi(1^j | C^j a^i, y^i).$$

We assume that player i cooperates in all remaining periods to bound $Pr[-\Phi_s^i | h_t^i]$ from above, since cooperation by player i only increases this probability. Then, conditional on h_t^i, y_u^j for $u = 0, \dots, T-1$ are independent. We have $y_u^j = 1$ with probability $p(a_u^i, y_u^i)$ for $u < t$ and with probability $p(C^i)$ for $u \geq t$. When $s \leq t$, the mean of $X(h_s^j)$ conditional on h_t^i is

$$\sum_{u=0}^{s-1} \delta^u p(a_u^i, y_u^i) \leq \sum_{u=0}^{s-1} \delta^u p(C^i, y_u^i) = X(h_s^i)(p(C^i, 1^i) - p(C^i, 0^i)) + \frac{1 - \delta^s}{1 - \delta} p(C^i, 0^i) \leq$$

$$\left(\frac{1 - \delta^s}{1 - \delta} \pi(1^i | C^i C^j) + T^{1/6} \sigma_C^j \right) (p(C^i, 1^i) - p(C^i, 0^i)) + \frac{1 - \delta^s}{1 - \delta} p(C^i, 0^i) =$$

$$T^{1/6} \sigma_C^j (p(C^i, 1^i) - p(C^i, 0^i)) + \frac{1 - \delta^s}{1 - \delta} p(C^i),$$

since $\pi(1^i | C^i C^j)(p(C^i, 1^i) - p(C^i, 0^i)) + p(C^i, 0^i) = p(C^i)$. Similarly, when $s > t$, the mean of $X(h_s^j)$ conditional on h_t^i is also less than or equal to

$$T^{1/6} \sigma_C^j (p(C^i, 1^i) - p(C^i, 0^i)) + \frac{1 - \delta^s}{1 - \delta} p(C^i).$$

By Hoeffding's inequality (see Hoeffding (1963))¹⁴

$$\Pr \left[X(h_s^j) \geq \frac{1 - \delta^s}{1 - \delta} p(C^i) + T^{1/6} \sigma_C^j \mid h_t^i \right] \leq \exp \left(- \frac{2(T^{1/6} \sigma_C^j (1 - p(C^i, 1^i) + p(C^i, 0^i)))^2}{\sum_{u=0}^{s-1} \delta^{2u}} \right) \leq \exp(-K_j T^{1/3}),$$

since $(\sigma_C^j)^2 = \sum_{u=0}^{T-1} \delta^{2u} p(C^i)(1 - p(C^i))$.

Therefore, for any $s = 1, \dots, T$,

$$\Pr [-\Phi_s^j \mid h_t^i] \leq \sum_{u=0}^{s-1} \Pr \left[X(h_u^j) \geq \frac{1 - \delta^u}{1 - \delta} p(C^i) + T^{1/6} \sigma_C^j \mid h_t^i \right] \leq T \exp(-K_j T^{1/3}) \quad (30)$$

when player j cooperates in every period of the review phase. \square

A.3 Proof of Proposition 2.

Proof. For players $i = 1, 2$, define recursive strategies \bar{C}^i and \bar{D}^i of the infinitely repeated game as follows. Let us divide the timeline into T -period review phases. Strategy \bar{C}^i starts with the T -period substrategy \hat{C}^i , and \bar{D}^i starts with \hat{D}^i . In all but the initial review phase, the player's T -period strategy depends on his private history and strategy in the previous review phase. If player i has played \hat{C}^i in the previous review phase and has observed private history h_T^i , then in the new review phase he follows the strategy

$$\begin{cases} \hat{C}^i & \text{with probability } (W_C^i(h_T^i) - G_D^j)/(G_C^j - G_D^j) \\ \hat{D}^i & \text{with probability } (G_C^j - W_C^i(h_T^i))/(G_C^j - G_D^j), \end{cases}$$

thereby assigning to the opponent an expected payoff of $W_C^i(h_T^i)$. Similarly, if player i has followed \hat{D}^i in the previous review phase and has observed private history h_T^i , then in the new review phase player i mixes between \hat{D}^i and \hat{C}^i to deliver to his opponent a continuation payoff of $W_D^i(h_T^i)$.

Notice that the strategies \bar{C}^i and \bar{D}^i have different starting regimes but the same transition rule between review phases (depending on the previous-phase strategy and private history).

¹⁴Hoeffding's inequality implies that whenever $\delta^s S_s^j \in [0, \delta^s]$ are independent random variables (conditionally on h_t^i), we have

$$\Pr [X(h_s^j) - E[X(h_s^j) \mid h_t^i] \geq M \mid h_t^i] \leq \exp \left(- \frac{2M^2}{\sum_{u=0}^{s-1} \delta^{2s}} \right).$$

Let us show that both \bar{C}^i and \bar{D}^i are best responses to \bar{C}^j and \bar{D}^j . From the properties of these strategies outlined in the statement of the proposition, it follows immediately that G_C^i is the payoff in response to \bar{C}^j from any strategy that involves \hat{C}^i or \hat{D}^i in each review phase, and in particular strategies \bar{C}^i and \bar{D}^i . Similarly, G_D^i is the payoff in response to \bar{D}^j from any of those strategies.

Let us show that G_C^i and G_D^i are the maximal expected payoffs that player i can achieve in response to \bar{C}^j and \bar{D}^j . If not, let \bar{A}_C and \bar{A}_D be strategies that achieve the maximal expected payoffs of $F_C^i \geq G_C^i$ and $F_D^i \geq G_D^i$ (with at least one strict inequality) in response to \bar{C}^j and \bar{D}^j , respectively. Without loss of generality, assume that $F_C^i - G_C^i \geq F_D^i - G_D^i$.

Consider player i playing \bar{A}_C in response to \bar{C}^j . At the end of the first review phase, conditional on h_T^i and h_T^j , player i 's expected payoff from the rest of the game cannot be greater than

$$\begin{aligned} & \delta^T \frac{W_C^j(h_T^j) - G_D^i}{G_C^i - G_D^i} F_C^i + \delta^T \frac{G_C^i - W_C^j(h_T^j)}{G_C^i - G_D^i} F_D^i \leq \\ & \delta^T (F_C^i - G_C^i) + \delta^T \frac{W_C^j(h_T^j) - G_D^i}{G_C^i - G_D^i} G_C^i + \delta^T \frac{G_C^i - W_C^j(h_T^j)}{G_C^i - G_D^i} G_D^i = \delta^T (F_C^i - G_C^i + W_C^j(h_T^j)). \end{aligned}$$

Then, player i 's expected payoff at time 0 cannot be greater than

$$E \left[\sum_{s=0}^{T-1} \delta^s g^i(a_s^i a_s^j) + \delta^T (F_C^i - G_C^i + W_C^j(h_T^j)) \mid \bar{A}_C, \hat{C}^j \right] \leq \delta^T (F_C^i - G_C^i) + G_C^i$$

by (10). This is less than F_C^i , a contradiction. We conclude that both \bar{C}^i and \bar{D}^i are best responses to \bar{C}^j and \bar{D}^j .

Now, for any pair of payoffs $(w_1, w_2) \in [G_D^1, G_C^1] \times [G_D^2, G_C^2]$, one Nash equilibrium that achieves it is

$$\left(\frac{w_1 - G_D^2}{G_C^2 - G_D^2} \bar{C}^1 + \frac{G_C^2 - w_1}{G_C^2 - G_D^2} \bar{D}^1, \frac{w_2 - G_D^1}{G_C^1 - G_D^1} \bar{C}^2 + \frac{G_C^1 - w_2}{G_C^1 - G_D^1} \bar{D}^2 \right).$$

This Nash equilibrium can be made into a Perfect Bayesian Equilibrium by defining the players' actions appropriately after off-equilibrium path private histories. \square

A.4 Proof of Lemma 2.

Proof. Let us alter player j 's strategy to let him cooperate in all periods. Then for all $s = 1 \dots T$, the entire joint distribution of s -period private histories in the event Φ_s^j remains unaffected, since j 's strategy has been altered only outside the event Φ_s^j . Both (20) and

(21) are unaffected by this change because they are conditioned on Φ_T^j (so Φ_s^j is true for any s). The probabilities (19) are also unaffected because once $\neg\Phi_s^j$ is true for some s , then $\neg\Phi_u^j$ is true for all $u > s$ and therefore the actions that are taken in the event $\neg\Phi_s^j$ do not affect the probabilities of $\neg\Phi_u^j$ for $u > s$. Therefore, throughout the proof, without loss of generality we assume player j is cooperating in all periods.

Fix a history $h_t^i \in \Phi_t^i$. Then, by Lemma 1,

$$Pr[\neg\Phi_T^j | h_t^i] \leq \exp(-K_j T^{1/3}). \quad (31)$$

We can immediately justify (19):

$$Pr[\neg\Phi_T^j | h_t^i, \Phi_t^j] = \frac{Pr[\neg\Phi_T^j | h_t^i] - Pr[\neg\Phi_t^j | h_t^i]}{1 - Pr[\neg\Phi_t^j | h_t^i]} \leq Pr[\neg\Phi_T^j | h_t^i] \leq T \exp(-K_j T^{1/3}).$$

For the estimates (20) and (21), we assume that player i cooperates in all remaining periods $s = t, \dots, T-1$. Observe that $X(h_T^j) \in [0, T]$, and that $E[X(h_T^j) | h_t^i, \Phi_t^j]$ is between

$$\frac{E[X(h_T^j) | h_t^i, \Phi_t^j] - T Pr[\neg\Phi_T^j | h_t^i, \Phi_t^j]}{1 - Pr[\neg\Phi_T^j | h_t^i, \Phi_t^j]} \quad \text{and} \quad \frac{E[X(h_T^j) | h_t^i, \Phi_t^j]}{1 - Pr[\neg\Phi_T^j | h_t^i, \Phi_t^j]}.$$

Since $Pr[\neg\Phi_T^j | h_t^i, \Phi_t^j] \leq T \exp(-K_j T^{1/3})$ by Lemma 1 and $E[X(h_T^j) | h_t^i] = O(T)$, we have

$$E[X(h_T^j) | h_t^i, \Phi_t^j] = E[X(h_T^j) | h_t^i, \Phi_t^j] + O(T^2 \exp(-K_j T^{1/3}))$$

and also

$$E[X(h_T^j) | h_t^i, \Phi_t^j]^2 = E[X(h_T^j) | h_t^i, \Phi_t^j]^2 + O(T^3 \exp(-K_j T^{1/3})).$$

Similarly, $E[X(h_T^j)^2 | h_t^i, \Phi_t^j]$ is between

$$\frac{E[X(h_T^j)^2 | h_t^i, \Phi_t^j] - T^2 Pr[\neg\Phi_T^j | h_t^i, \Phi_t^j]}{1 - Pr[\neg\Phi_T^j | h_t^i, \Phi_t^j]} \quad \text{and} \quad \frac{E[X(h_T^j)^2 | h_t^i, \Phi_t^j]}{1 - Pr[\neg\Phi_T^j | h_t^i, \Phi_t^j]}.$$

Thus,

$$E[X(h_T^j)^2 | h_t^i, \Phi_t^j] = E[X(h_T^j)^2 | h_t^i, \Phi_t^j] + O(T^3 \exp(-K_j T^{1/3})).$$

It follows that

$$Var[X(h_T^j)^2 | h_t^i, \Phi_t^j] = E[X(h_T^j)^2 | h_t^i, \Phi_t^j] - E[X(h_T^j) | h_t^i, \Phi_t^j]^2 =$$

$$E[X(h_T^j)^2 | h_t^i, \Phi_t^j] - E[X(h_T^j) | h_t^i, \Phi_t^j]^2 + O(T^3 \exp(-K_j T^{1/3})) = \\ \text{Var}[X(h_T^j) | h_t^i, \Phi_t^j] + O(T^3 \exp(-K_j T^{1/3})).$$

Equations (20) and (21) follow, because when player i chooses the same action a^i from period t onwards, then

$$E[X(h_T^j) | h_t^i, \Phi_t^j] = \mu(h_t^i) + \frac{\delta^t - \delta^T}{1 - \delta} p(C^i) \quad \text{and} \quad (32)$$

$$\text{Var}[X(h_T^j) | h_t^i, \Phi_t^j] = \sigma^2(h_t^i) + \frac{\delta^{2t} - \delta^{2T}}{1 - \delta^2} p(C^i)(1 - p(C^i)). \quad (33)$$

□

A.5 Proof of Lemma 3.

Proof. Suppose at time t player i believes that X_t^j has mean and variance μ and σ^2 . Then $\hat{V}_{a^i}(t, \mu, \sigma^2)$ is player i 's expected payoff from choosing action $a^i \in \{C^i, D^i\}$ in the remaining periods of the review phase, since i 's beliefs about $X(h_T^j)$ have mean and variance

$$\mu + \frac{\delta^t - \delta^T}{1 - \delta} p(a^i) \quad \text{and} \quad \sigma^2 + \frac{\delta^{2t} - \delta^{2T}}{1 - \delta^2} p(a^i)(1 - p(a^i))$$

respectively, and $E[V^j(x)] = V^j(E[x]) + \alpha^j \text{Var}[x]$ for the quadratic function V^j . $M^j(t)$ is defined so that $\hat{V}_{C^i}(t, \mu, \sigma^2) = \hat{V}_{D^i}(t, \mu, \sigma^2)$ when $\mu = M^j(t)$. Since

$$\frac{d}{d\mu} \hat{V}_{C^i}(t, \mu, \sigma^2) > \frac{d}{d\mu} \hat{V}_{D^i}(t, \mu, \sigma^2),$$

for all t , μ and σ , it follows that $\hat{V}_{C^i}(t, \mu, \sigma^2) \geq \hat{V}_{D^i}(t, \mu, \sigma^2)$ if and only if $\mu \geq M^j(t)$. Therefore, the conclusion of Lemma 3 follows if we show that $\hat{V}^i(t, \mu, \sigma^2)$ is player i 's expected payoff from the optimal strategy.

Let us prove this game by backward induction on t . Trivially, it holds for $t = T$. Assuming that it holds for time $t + 1$, let us prove it for time t .

Suppose that $\mu \geq M^j(t)$. If player i cooperates once, he prefers to cooperate in all future periods and gains the payoff of $\hat{V}_{C^i}(t, \mu, \sigma^2)$, because

$$\mu + \delta^t p(C^i, 1^i) > \mu + \delta^t p(C^i, 0^i) \geq \mu + \delta^t \frac{p(C^i) + p(D^i)}{2} \geq M^j(t + 1).$$

This payoff exceeds player i 's payoff from defecting once and acting optimally in all remaining periods by Lemma 5 below. Therefore, $\hat{V}_{C^i}(t, \mu, \sigma^2)$ is the maximal expected payoff that player i can achieve at time t if $\mu \geq M^j(t)$. A similar argument implies that $\hat{V}_{D^i}(t, \mu, \sigma^2)$ is player i 's maximal expected payoff if $\mu \leq M^j(t)$. \square

Lemma 5. *When $\mu \geq M^j(t)$, then*

$$\begin{aligned} \hat{V}_{C^i}(t, \mu, \sigma^2) &\geq 2\delta^T \alpha^j (\mu - M^j(t)) (p(C^i) - p(D^i)) + \\ &\pi(1^j | C^j D^i) \hat{V}^i(t+1, \mu + \delta^t p(D^i, 1^i), \sigma^2 + \delta^{2t} p(D^i, 1^i) (1 - p(D^i, 1^i))) \\ &+ \pi(0^j | C^j D^i) \hat{V}^i(t+1, \mu + \delta^t p(D^i, 0^i), \sigma^2 + \delta^{2t} p(D^i, 0^i) (1 - p(D^i, 0^i))) + \delta^t g^i(D^i C^j). \end{aligned} \quad (34)$$

Proof. Note that (34) holds with equality at $\mu = M^j(t)$ by the definition of $M^j(t)$. To prove (34) for $\mu > M^j(t)$ let us show that the derivative of the left hand side with respect to μ exceeds the derivative of the right hand side by at least $2\delta^T \alpha^j (p(C^i) - p(D^i))$.¹⁵ The derivative of the left hand side with respect to μ is

$$\delta^T (V^j)'(\mu + \frac{\delta^t - \delta^T}{1 - \delta} p(C^i)) = 2\delta^T \alpha^j (\mu + \delta^t p(C^i)) + \frac{\delta^{t+1} - \delta^T}{1 - \delta} p(C^i) - X_C^j + \delta^T \beta^j. \quad (35)$$

The derivative of the right hand side is

$$\begin{aligned} &\pi(1^j | C^j D^i) (V^j)'(\mu + \delta^t p(D^i, 1^i)) + \frac{\delta^{t+1} - \delta^T}{1 - \delta} p(a^i(1)) + \\ &\pi(0^j | C^j D^i) (V^j)'(\mu + \delta^t p(D^i, 0^i)) + \frac{\delta^{t+1} - \delta^T}{1 - \delta} p(a^i(0)) \leq \\ &\pi(1^j | C^j D^i) (V^j)'(\mu + \delta^t p(D^i, 1^i)) + \frac{\delta^{t+1} - \delta^T}{1 - \delta} p(C^i) + \\ &\pi(0^j | C^j D^i) (V^j)'(\mu + \delta^t p(D^i, 0^i)) + \frac{\delta^{t+1} - \delta^T}{1 - \delta} p(C^i) = \\ &2\delta^T \alpha^j (\mu + \delta^t p(D^i)) + \frac{\delta^{t+1} - \delta^T}{1 - \delta} p(C^i) - X_C^j + \delta^T \beta^j, \end{aligned} \quad (36)$$

where $a^i(1)$ and $a^i(0)$ are the actions player i chooses to follow in all remaining periods after he defects once and observes signal 1^i or 0^i . The difference between (35) and (36) is $2\delta^T \alpha^j (p(C^i) - p(D^i))$. \square

¹⁵The derivative of the right hand side exists everywhere except for two kink points.

A.6 Proof of Lemma 4.

By an argument analogous to Lemma 1, we can conclude that if player j cooperates in every period and h_t^i involves only defection by player i , then

$$Pr[-\Phi_T^j | h_t^i] \leq \exp(-K_j T^{1/3}).$$

To see that this argument goes through, the key observation is that the mean of $X(h_s^j)$ conditional on h_t^i is

$$\sum_{u=0}^{s-1} \delta^u p(D^i, y_u^i) \leq \sum_{u=0}^{s-1} \delta^u p(D^i, 1^i) < T^{1/6} \sigma_C^j (p(C^i, 1^i) - p(C^i, 0^i)) + \frac{1 - \delta^s}{1 - \delta} p(C^i).$$

Estimates (27), (28) and (29) follow from an argument identical to Lemma 2, with C^i replaced by D^i in (32) and (33).

A.7 Lemma 6.

Lemma 6. *Suppose that player j is following a strategy from class Z^j . If $h_t^i \in \Phi_t^i$, then*

$$E[X(h_t^j) | h_t^i, \Phi_t^j] = \sum_{s=0}^{t-1} \delta^s p(a_s^i, y_s^i) + O(T^2 \exp(-K_j T^{1/3})) \quad (37)$$

Proof. As in the proof of Lemma 2, we can assume that player j is cooperating in all periods to evaluate $E[X(h_t^j) | h_t^i, \Phi_t^j]$. Observe that $X(h_t^j) \in [0, t)$, and that $E[X(h_t^j) | h_t^i, \Phi_t^j]$ is between

$$\frac{E[X(h_t^j) | h_t^i] - t Pr[-\Phi_t^j | h_t^i]}{1 - Pr[-\Phi_t^j | h_t^i]} \quad \text{and} \quad \frac{E[X(h_t^j) | h_t^i]}{1 - Pr[-\Phi_t^j | h_t^i]}.$$

Since $E[X(h_t^j) | h_t^i] = O(t)$ and $Pr[-\Phi_t^j | h_t^i] \leq T \exp(-K_j T^{1/3})$ by Lemma 1, we have

$$E[X(h_t^j) | h_t^i, \Phi_t^j] = E[X(h_t^j) | h_t^i] + O(T^2 \exp(-K_j T^{1/3})) = \sum_{s=0}^{t-1} p(a_s^i, y_s^i) + O(T^2 \exp(-K_j T^{1/3})).$$

□

A.8 Proof of Proposition 4.

Proof. We will use Kakutani's fixed point theorem to prove Proposition 4.

Consider player i 's best response strategies $\hat{C}^j \in Z^j$ with reward functions W_C^j given by Definition 2 for $\alpha^j = T^{-2}$ and some $\beta^j = K_C^i + O(T^{-1})$. Proposition 3 implies that there exists a constant $L > 0$ such that i 's best response is \hat{D}^i whenever $\beta^j \leq K_C^i - L$, and some $\hat{C}^i \in Z^i$ whenever $\beta^j \geq K_C^i + L$, for all sufficiently large T .

Consider the lowest value of β^j for which at least one strategy from class Z^i is at least as good as \hat{D}^i in response to a given strategy $\hat{C}^j \in Z^j$. For that value of β^j , denote by $\Phi^i(\hat{C}^j)$ the set of all such strategies from class Z^i . By continuity of payoffs in strategies, $\Phi^i(\hat{C}^j)$ is nonempty and player i is indifferent between any strategy in $\Phi^i(\hat{C}^j)$ and \hat{D}^i . By linearity of payoffs in mixed strategies (here we think about mixtures over pure strategies from Z^i), the set $\Phi^i(\hat{C}^j)$ is convex.

Let us prove that the correspondence Φ^i is upper semi-continuous. Consider sequences $\hat{C}_n^i \rightarrow \hat{C}^i$ and $\hat{C}_n^j \rightarrow \hat{C}^j$ such that $\hat{C}_n^i \in \Phi^i(\hat{C}_n^j)$ for all n . Let us show that $\hat{C}^i \in \Phi^i(\hat{C}^j)$. Denote by β_n^j the lowest value of β^j for which \hat{C}_n^i is at least as good as \hat{D}^i in response to \hat{C}_n^j . Without loss of generality, assume that $\beta_n^j \rightarrow \tilde{\beta}^j$ (otherwise we can take a convergent subsequence). Then, by continuity, among all strategies from Z^i , \hat{C}^i gives player i the highest payoff in response to \hat{C}^j when $\beta^j = \tilde{\beta}^j$. This payoff equals player i 's payoff from \hat{D}^i . It follows that $\hat{C}^i \in \Phi^i(\hat{C}^j)$ if we show that for any $\beta^j < \tilde{\beta}^j$, \hat{D}^i is strictly better than any strategy from Z^i in response to \hat{C}^j . Suppose not, i.e. $\hat{C}' \in Z^i$ is better than \hat{D}^i for some $\beta^j < \tilde{\beta}^j$. Since player i 's payoff is linear in β^j and \hat{D}^i is his strict best response for all sufficiently small β^j by Proposition 3, it follows that \hat{C}' is *strictly* better than \hat{D}^i in response to \hat{C}^j for $\beta^j = \tilde{\beta}^j > \beta^j$, a contradiction.

We conclude that the correspondence $(\hat{C}^1, \hat{C}^2) \rightarrow (\Phi^1(\hat{C}^2), \Phi^2(\hat{C}^1))$ from $Z^1 \times Z^2$ to itself is convex-valued, nonempty-valued, and upper hemi-continuous. By Kakutani's fixed point theorem, there are strategies \hat{C}^1 and \hat{C}^2 such that $\hat{C}^1 \in \Phi^1(\hat{C}^2)$ and $\hat{C}^2 \in \Phi^2(\hat{C}^1)$. Then for $i = 1, 2$ in response to \hat{C}^j , a player i is indifferent between \hat{C}^i and \hat{D}^i for W_C^j defined by an appropriate value of $\beta^j = K_C^i + O(T^{-1})$. This completes the proof of Proposition 4, since by Proposition 3, it is always optimal to follow \hat{D}^i or a strategy from Z^i in response to any strategy from Z^j with a reward function W_C^j given by Definition 2. \square

A.9 Proof of Proposition 5.

Proof. We rely on our derivations and notation from the proof of Proposition 3. Using the lower bound (25) on player i 's payoff,

$$\begin{aligned}
G_C^i &\geq \hat{V}_{C^i}(0, 0, 0) + O(T^2 \exp(-K_j T^{1/3})) = \\
&\frac{1 - \delta^T}{1 - \delta} g^i(C^i C^j) + \delta^T V^j \left(\frac{1 - \delta^T}{1 - \delta} p(C^i) \right) + \delta^T \alpha^j \left(\frac{1 - \delta^{2T}}{1 - \delta^2} p(C^i) (1 - p(C^i)) \right) + O(T^2 \exp(-K_j T^{1/6})) > \\
&\frac{1 - \delta^T}{1 - \delta} g^i(C^i C^j) + \delta^T (G_C^i - \beta^j T^{1/6} \sigma_C^j) + \delta^T \alpha^j (\sigma_C^j)^2 + O(T^2 \exp(-K_j T^{1/6})) \Rightarrow \\
(1 - \delta^T) G_C^i &> \frac{1 - \delta^T}{1 - \delta} g^i(C^i C^j) - \delta^T \beta^j T^{1/6} \sigma_C^j + O(T^2 \exp(-K_j T^{1/3})) \Rightarrow \\
(1 - \delta) G_C^i &> g^i(C^i C^j) - \beta^j O(T^{-1/3}).
\end{aligned}$$

□

References

- Abreu, Dilip, Paul Milgrom & David Pearce. 1991. "Information and Timing in Repeated Partnerships." *Econometrica* 59(6):1713–1733.
- Aoyagi, Masaki. 2002. "Collusion in Dynamic Bertrand Oligopoly with Correlated Private Signals and Communication." *Journal of Economic Theory* 102:229–248.
- Bhaskar, V. & Ichiro Obara. 2002. "Belief-Based Equilibria in the Repeated Prisoners' Dilemma with Private Monitoring." *Journal of Economic Theory* 102:40–69.
- Compte, Oliver. 1998. "Communication in Repeated Games with Imperfect Private Monitoring." *Econometrica* 66(3):597–626.
- Ely, Jeffrey, Johannes Hörner & Wojciech Olszewski. 2005. "Belief-Free Equilibria in Repeated Games." *Econometrica* 73(2):377–415.
- Ely, Jeffrey & Juuso Valimäki. 2002. "A Robust Folk Theorem for the Prisoner's Dilemma." *Journal of Economic Theory* 102:84–105.
- Fudenberg, David & David Levine. 2002. "The Nash Threats Folk Theorem with Communication and Approximate Common Knowledge in Two Player Games." *Harvard Institute of Economic Research* .
- Fudenberg, David, David Levine & Eric Maskin. 1994. "The Folk Theorem with Imperfect Public Information." *Econometrica* 62(5):997–1040.
- Hoeffding, Wassily. 1963. "Probability Inequalities for Sums of Bounded Random Variables." *Journal of the American Statistical Association* 58(301):13–30.
- Hörner, Johannes & Wojciech Olszewski. 2006*a*. "The Folk Theorem for Games with Private Almost-Perfect Monitoring." *Econometrica* 74(65):1499–1544.
- Hörner, Johannes & Wojciech Olszewski. 2006*b*. "How Robust is the Folk Theorem with Imperfect Public Monitoring." Working Paper.
- Kandori, Michihiro. 2002. "Introduction to Repeated Games with Private Monitoring." *Journal of Economic Theory* 102:1–15.
- Kandori, Michihiro & Hitoshi Matsushima. 1998. "Private Observation, Communication and Collusion." *Econometrica* 66(3):627–652.
- Mailath, George & Larry Samuelson. 2006. *Repeated Games and Reputation: Long-Run Relationships*. New York, NY: Oxford University Press.
- Mailath, George & Stephen Morris. 2002. "Repeated Games with Almost-Public Monitoring." *Journal of Economic Theory* 102:189–228.

- Matsushima, Hitoshi. 2004. "Repeated Games with Private Monitoring: Two Players." *Econometrica* 72(3):823–852.
- Obara, Ichiro. 2007. "Folk Theorem with Communication." Working Paper.
- Piccione, Michele. 2002. "The Repeated Prisoner's Dilemma with Imperfect Private Monitoring." *Journal of Economic Theory* 102:70–83.
- Sekiguchi, Tadashi. 1997. "Efficiency in Repeated Prisoner's Dilemma with Private Monitoring." *Journal of Economic Theory* 76:345–361.
- Stigler, George. 1964. "A Theory of Oligopoly." *Journal of Political Economy* 72:44–61.