Efficiency in the Repeated Prisoner's Dilemma with Imperfect Private Monitoring.

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

September 15, 2010

Abstract

We prove that there exist equilibrium payoffs arbitrarily close to the efficient payoff in the two-player prisoner's dilemma with low discounting under imperfect private monitoring, provided that the monitoring structure satisfies two restrictions. We assume no communication, and no public randomization device.

1 Introduction.

Cooperation takes information. Precisely how little it takes is, however, an open issue, and a central theme in the literature on repeated games. Progress has largely consisted in weakening these informational requirements, from perfect monitoring (Fudenberg and Maskin, 1986) to imperfect public monitoring (Abreu, Pearce & Stacchetti, 1990, Fudenberg, Levine & Maskin, 1994). This paper provides a further step in this literature, by showing that cooperation can be sustained in the two-player prisoner's dilemma under imperfect private monitoring, under some restrictions. More precisely, we show that payoffs arbitrarily close to the efficient payoff can be achieved in equilibrium by sufficiently patient players, provided that the monitoring structure satisfies two conditions. First, monitoring should not be too noisy, in the sense that there should be some chance that a defecting player observes a signal that is sufficiently more likely when his

*Economics Department, Stanford University, kyna.fong@stanford.edu.

†Paris School of Economics, and Mathematics Department, London School of Economics, ogossner@gmail.com.

‡Economics Department, Yale University, johannes.horner@yale.edu.

§Economics Department, Princeton University, sannikov@gmail.com.

opponent defects than when he cooperates. Second, for each player, there must exist a particular type of statistic that is informative about his opponent's action, such that the pair of statistics is positively correlated.

Therefore, this paper is the first to show that cooperation is not a non-generic phenomenon under private monitoring. Previous contributions have established important limiting results. As the monitoring structure converges to perfect monitoring, the efficient payoff –indeed, the set of feasible and individually rational payoffs– can be supported in the two-player prisoner's dilemma (Sekiguchi, 1997, Bhaskar and Obara, 2002, Piccione, 2002, Ely and Välimäki, 2002) a result later extended to all finite games (Hörner and Olszewski, 2006). Further, the same result holds under some assumptions when the monitoring structure approaches imperfect public monitoring (Hörner and Olszewski, 2009), a result that builds on earlier findings (Mailath and Morris, 2002). Finally, Matsushima (2004) shows that the folk theorem also holds in the two-player prisoner's dilemma provided the monitoring structure is private, but conditionally independent, yielding an elegant counterpoint to Matsushima (1991). An excellent summary of some of these ideas can be found in Mailath and Samuelson (2006), as well as in Kandori (2002)'s survey.

These results provide significant robustness checks for the well-known folk theorems under perfect or imperfect public monitoring, and develop useful techniques paving our way. However, because the monitoring structures they consider are extreme cases, they are of limited value for applications in which monitoring is truly private. In industrial organization, the prevalence of such environments has already been emphasized (see Stigler, 1964).

To understand both the structure of our proof and the role of our assumptions, it is instructive to first describe the difficulties in generalizing earlier constructions. When monitoring is imperfect, it is necessary to aggregate information. Following Radner (1986), this can be achieved by dividing the infinite horizon into *review phases* of length T (see also Compte, 1998; Kandori and Matsushima, 1998). At the end of each phase, the continuation strategy is chosen as a function of some initial state and some final summary statistic, or *score*. From one phase to the next, the strategy profile is *belief-free*. That is, at the end of each review phase, each player's continuation strategy is optimal independently of the private history of his opponent (and so independently of the player's own history as well). However, the equilibrium itself is not belief-free: within a round, incentives depend on a player's *recent history*, that is, on his earlier observations during that round. Indeed, it is known that belief-free equilibria cannot support a nearly efficient outcome if the monitoring structure is bounded away from perfect monitoring (see Ely, Hörner and Olszewski, 2005).

Up to this point, our construction follows Matsushima (2004). In Matsushima (2004), players use one of two strategies within each round. One of these strategies always cooperates, and the other always defects. At the end of each round, each player chooses which strategy to use so as to enforce some continuation payoff, or reward, assigned to his opponent. The key in his construction is that signals are independent across players, conditional on an action profile. This implies that, within a round, a player's belief about the score observed by his opponent, and so his continuation strategy itself, is independent of his recent signals.

Difficulties appear once correlation across signals is allowed. During each round, a player's history of signals affects his belief about the signals observed by his opponent; and so about his score; and so about his continuation payoff at the end of the round. This affects his incentives. In general, it is not possible to provide incentives to always cooperate within a round, while preserving efficiency. Efficiency requires that the expected continuation payoff is close to the maximal one when a player always cooperates within a round. This means that a player cannot be rewarded for a score that is unusually high, an event that he might infer from his own signals. After some histories, cooperation *must* break down. This further complicates learning, because a player now also learns about his score indirectly, through the inferences he draws about the actions taken by his opponent. Observations are no longer i.i.d. over time. This provides opportunities for *strategic manipulation*, as a player's actions now affects his opponent's continuation strategy within a round.

Our proof relies on two critical insights. First, when a player observes an exceptionally high score, say n standard deviations above the mean, he expects that his opponent's private score is only ρn above the mean, where $\rho \in (0,1)$ is the positive correlation across scores (this is where one restriction on the monitoring structure must be imposed). Therefore, even when a player stops providing incentives for cooperation to his rival because his score attains some critical threshold, he keeps having incentives to cooperate himself, because he assigns very low probability to the score observed by his opponent being close to the critical threshold. This only works, however, if observations can be treated as i.i.d. random variables, that is, if each player views his opponent's action as constant over time. The second insight is that, if a player punishes his opponent for scores above the threshold (through his choice of a continuation payoff at the end of the round) in such a way that, conditional on this event, his opponent is indifferent over all continuation strategies within the round, each player can safely condition on his opponent's score being below the threshold, and therefore, on his opponent's action being constant.

As mentioned, cooperation must break down after some histories, and the incentives after

those histories depend on the fine details of the monitoring structure. Accordingly, the proof is partly non-constructive, and we simply show that there exists some strategy profile which cooperates "almost always." Yet this unspecified cooperative strategy must also be a best-reply to the opponent's defective strategy, as follows from the requirement that the strategies be belief-free from one round to the next. To ensure this, we specify the future continuation payoff assigned to his opponent by a player using the defective strategy in such a way that, conditional on this event, *all* strategies within the round are optimal, including, necessarily, the cooperative one. This severely restricts this reward function, and to make sure that the resulting range of continuation payoffs is feasible, the second restriction on the monitoring structure must be imposed: some signal must be sufficiently informative. It is worth noting that, while not innocuous, this restriction is automatically satisfied whenever nontrivial belief-free equilibria in the two-player prisoner's dilemma exist (see Ely, Hörner and Olszewski, 2005).

Some of the difficulties we encounter are specifically due to discounting. Lehrer (1990) provides a remarkable analysis of the undiscounted case. Also, Fudenberg and Levine (1991) prove a folk theorem when the solution concept used is approximate optimality. Finally, there is a growing literature on repeated games with imperfect private monitoring and communication. As mentioned earlier, some of these papers use similar ideas and techniques. See Ben-Porath and Kahneman (1996), Compte (1998), Kandori and Matsushima (1998), Aoyagi (2002) and Obara (2009). Building on Obara (2009), Sugaya (2010) provides a remarkable result in the case in which there are sufficiently many signals. While the class of games and monitoring structures they consider are significantly larger than ours, it is worth pointing out that they do not include ours. In particular, our result establishes efficiency in some cases for which this was heretofore unknown even with communication.

The paper is organized as follows. Section 2 introduces the model and states the main result. Section 3 presents a brief overview of the argument and develops the basic theoretical ideas behind the construction. Section 4 presents the formal proofs.

2 Notation and Result

We consider the infinitely repeated prisoner's dilemma with private monitoring. Each player $i = 1, 2$ chooses an action $a_t^i \in \{C^i, D^i\}$ in every period $t \geq 1$. Players do not observe each other's actions. Rather, at the end of each period, player i observes a private signal y_t^i from a finite set Y^i with $N \geq 2$ elements. There is no communication and no public randomization

device.

For each action profile, every profile of private signals realizes according to a joint probability distribution $\pi(y^i y^j | a^i a^j)$.[1] We assume that the matrix of joint probabilities $(\pi(\cdot|a^i a^j))$ has full rank and full support for each action profile (a^i, a^j). We use $\pi(y^i|a^i a^j)$ to denote the marginal probability that player i receives signal y^i given action profile (a^i, a^j), 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, and given (a^i, a^j).

Denote by $g^i(a^i a^j)$ the expected stage-game payoff of player i, and by $g(a^i a^j)$ for the payoff vector.[2] As mentioned, the stage-game payoffs are those of the prisoner's dilemma, i.e.

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

and that the cooperative action profile $C^i C^j$ is efficient, i.e., it maximizes the sum of the players' payoffs over all action profiles.

A t-period private history of player i is a sequence of player i's past actions and signals, denoted by $h_t^i = (a_1^i, y_1^i, a_2^i, y_2^i \dots a_t^i, y_t^i)$. Let H_t^i $(i = 1, 2, t \geq 2)$ be the set of t-period private histories for player i. For notational convenience, we define H_0^i $(i = 1, 2)$ as an arbitrary singleton set. A behavior strategy for player i is a function $s^i : \bigcup_{t=0}^{\infty} H_t^i \to [0, 1]$ that specifies the probability with which player i plays C^i after each private history $h_t^i \in H_t^i$ for all $t \geq 1$. We denote the set of strategies for player i by S^i.

Players discount future payoffs at a common rate δ. Given a strategy profile $s = (s^1, s^2) \in S^1 \times S^2$, player i's expected payoff (or payoff, for short) is

$$\mathbb{E}_s \left[\sum_{t=1}^{\infty} \delta^{t-1} g^i(a_t) \right], \tag{1}$$

where $\mathbb{E}_s[\cdot]$ refers to the expected value with respect to the probability distribution of action profiles induced by s, and a_t is the realized action profile in period t. We refer to the repeated game with private monitoring and discount factor δ by G_δ. The *average payoff* for player i is player i's payoff, multiplied by the factor $(1 - \delta)$.

Our objective is to prove that there exist asymptotically efficient sequential equilibria of the game G_δ. That is, we wish to construct a sequence of equilibria s_δ such that the average payoff

[1] Whenever we refer to players i and j, we assume $i, j \in \{1, 2\}$ and $i \neq j$.

[2] As usual, this can be interpreted as the expectation of a function that only depends on player i's action and his private signal, so that player i's realized payoff carries no further information about player j's action. See Kandori (2002) for details.

vector converges to $g(C^1C^2)$ as $\delta \to 1$. We shall establish that there exist asymptotically efficient Nash equilibria of the game. Of course, a Nash equilibrium does not satisfy sequential rationality in general. However, because of the full support assumption, for every Nash equilibrium, there exists a sequential equilibrium with the same outcome. This is because any player's information set off the equilibrium path must follow the player's own deviation. See Sekiguchi (1997, Proposition 3) for a formal statement and proof of this claim.

Our main result holds under two assumptions on the monitoring structure. The first assumption states that when player j is defecting, there exists a signal $\hat{y}^j \in Y^j$ that has a sufficiently high likelihood ratio to test for player i's cooperation:

Assumption 1 (Minimal informativeness)**.** *For $j = 1, 2$ there exists a signal $\hat{y}^j \in Y^j$ such that*

$$\frac{\pi(\hat{y}^j|D^jC^i)}{\pi(\hat{y}^j|D^jD^i)} > \frac{g^i(C^iC^j) - g^i(C^iD^j)}{g^i(C^iC^j) - g^i(D^iD^j)}.$$

Let M^i denote the matrix of conditional probabilities $(\pi(y^j|C^iC^j, y^i))_{y^iy^j}$, that expresses player i's private beliefs about player j's signal when the action profile C^iC^j is played. In our equilibrium construction, each player i assigns a score $\lambda^i(y^i)$ for each received signal y^i. Our second assumption guarantees that these scores can be chosen in a convenient way, namely that when player j plays C^j, the expectation of the score λ^i that his opponent assigns is higher than when player j plays D^j.

Assumption 2 (Positively correlated scores)**.** *There exists an eigenvector λ^1 of M^1M^2 associated to a positive eigenvalue such that, if we let $\lambda^2 = M^2\lambda^1$, the expectation of λ^i for $i = 1, 2$ is higher under C^iC^j than under C^iD^j.*

In order to better understand Assumption 2, consider the case in which $N = 2$ so that each player has two signals. Label one signal of player i as 1^i where, for every action of player i, 1^i is always more likely when player j plays C^j than when player j plays D^j. The signal 1^i can then be interpreted as a "good" signal about j's cooperation. In this case, it is straightforward to see that Assumption 2 reduces to the statement that good signals are positively correlated when both players cooperate, i.e., for $i = 1, 2$:

$$\pi(1^i|C^iC^j, 1^j) \geq \pi(1^i|C^iC^j).$$

It is worth pointing out that both of our assumptions involve existential qualifiers and are therefore more "likely" to be satisfied as the number of signals grows.[3]

[3]This assertion can be formalized by showing that the measure of monitoring structures satisfying our assumptions increases in the number of signals; more precisely, this measure converges to one exponentially fast.

We can now state our main result.

Theorem 1. *Under Assumptions 1 and 2, there exist asymptotically efficient equilibria: for every $\varepsilon > 0$, there exists $\underline{\delta} < 1$, for every $\delta \in (\underline{\delta}, 1)$, there exists a sequential equilibrium of G_δ whose average payoff for player $i = 1, 2$ exceeds $g^i(C^1C^2) - \varepsilon$.*

3 An Overview of the Argument

This section provides some intuition for our construction, as well as for the role of our two assumptions.

Following Radner (1986) and Matsushima (2004), our equilibrium relies on breaking up the infinite horizon into finite *review phases*. 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

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

Longer review phases allow for better information aggregation, which helps reduce the use of inefficient punishments that occur on the equilibrium path. At the same time, it is important that δ^T converge to 1 as $\delta \to 1$, so that each review phase has a negligible contribution to the overall payoff in the supergame.

As in Matsushima (2004), the equilibrium is *periodically* belief-free, which means that, at the beginning of each review phase, there exist optimal continuation strategies for player i that are independent of his private history. More precisely, any continuation strategy that adheres to one of two strategies of the T-finitely repeated game in each review phase is optimal. These strategies are denoted $\mathcal{C}^i$ and $\mathcal{D}^i$. Strategy $\mathcal{C}^i$ involves cooperation after almost all private histories in a review phase, and strategy $\mathcal{D}^i$ consists of defection after *all* histories of the review phase.

Player i creates incentives for his opponent to select from those strategies through a transition rule that determines which strategy, $\mathcal{D}^i$ or $\mathcal{C}^i$, 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. Thus this reward function, which we denote by W_C^j when $\mathcal{C}^j$ is played and by W_D^j when $\mathcal{D}^j$ is played, creates incentives across review phases. Denote by $[G_D^j, G_C^j]$ the range of rewards and punishments (i.e., continuation payoffs) that can be assigned to player j by player i's mixing between $\mathcal{D}^i$ and $\mathcal{C}^i$.

We may thus focus on the T-finitely repeated game in which each player's payoff is augmented by a terminal reward that depends on his opponent's private history. To ensure that both $\mathcal{C}^i$ and $\mathcal{D}^i$ are optimal, we construct pairs of strategies and reward functions (with range in $[G_D^j, G_C^j]$) with the property that both the strategy $\mathcal{C}^i$ and $\mathcal{D}^i$ are optimal in the T-finitely repeated game, whether player j uses $(\mathcal{C}^j, W_C^j)$, or $(\mathcal{D}^j, W_D^j)$. Efficiency in the repeated game will then be achieved if, provided both players have used $\mathcal{C}^1, \mathcal{C}^2$ in a review phase, the reward for player i must be arbitrarily close to its maximum G_C^i, so that both players use those strategies again in the following phase, with probability arbitrarily close to one.

As explained below, our construction will not only be periodically belief-free, but also *conditionally* belief-free: conditional on player j using the pair $(\mathcal{D}^j, W_D^j)$, *all* strategies of player i in the T-finitely repeated game will be optimal. This will only be feasible under some assumption, namely, Assumption **1**. But it will allow us to focus on the case in which player j uses the pair $(\mathcal{C}^j, W_C^j)$, because player i might assume as well for the sake of computing his best responses.

Let us start by considering the cooperative strategy $\mathcal{C}^j$ and the reward function W_C^j. A natural way for player j to reward player i is to compute a *score* of i's performance that depends on the signals that j receives. For instance, player j could count the number of times he received a "good" signal about player i's behavior. More generally, start with an assignment λ^j of signals y^j to real numbers, whose expected value is higher when player i cooperates than when he defects. By adding up these values over all T periods, player j obtains a score that determines player i's reward: if higher scores lead to appropriately higher rewards, player i can be provided with incentives to cooperate.

Suppose we try to guarantee that player i cooperates in every period of the review phase. In order to motivate cooperation in every period after any private history h_i^i, W_C^j must reward player i even when he is extremely lucky and achieves the best possible scores. In expectation, such a reward function will specify a reward $O(T)$ below the maximum (efficient) level G_C^i. Hence, on average, player j will switch to $\mathcal{D}^j$ at the end of the review phase with probability bounded away from zero, which destroys efficiency. So it is impossible for W_C^j to induce cooperation in all periods, yet guarantee efficiency. See the left panel of Figure 1.

As a solution, we "shift" the reward function up so that in expectation, for some $0 < k < 1$ only a term of the order $O(T^k)$ in value is destroyed through transitions to $\mathcal{D}^j$. That loss in efficiency of $O(T^k)$ per T periods is negligible as T becomes large. See the center panel of Figure 1. We take $k = 2/3 > 1/2$ so that the probability that player i's score, as computed by j, ever exceeds the critical threshold is arbitrarily small. Below this threshold, incentives to cooperate

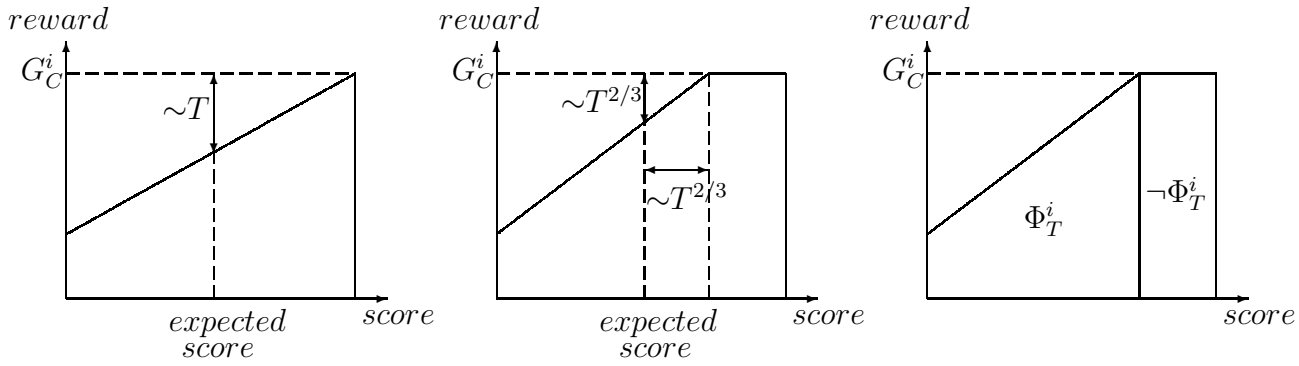

Figure 1: Scores and rewards.

will be provided, but not above. For the purpose of this discussion, we refer to these two events as Φ_t^j and $\neg\Phi_t^j$. The event Φ_t^j corresponds to scores that have remained throughout below some critical threshold so far, while $\neg\Phi_t^j$ gathers the histories of j according to which player i has already "overperformed." See the right panel of Figure 1. We hasten to add that the formal definition of these events, provided in the next section, is somewhat more complicated, and the reward function is actually not "flat" for scores above the threshold. Our informal discussion here should be viewed as merely suggestive of the actual construction.

If signals were conditionally independent, players would never learn about their score. Because we do not assume they are, players update their beliefs about their opponent's private history, and therefore about their own score, from their private history. This matters, because, as explained, there are histories after which player j does not provide incentives for player i to cooperate. There will be histories after which a player defects, given his inferences about his incentives.

There are at least two reasons for why player i cares about the possibility that player j defects. First, it affects player i's flow payoff, and so distorts player i's incentives (perhaps he will be motivated to take actions allowing him to prevent this from happening). Second, it affects player i's learning about his score, because player j's action affects the distribution of signals. This implies that, in general, player i cannot treat the signals that he receives as identically distributed, and in fact, not even as independently distributed (because player j's action in period t depends on his previous signals).

If player j anticipates that his score is likely to be such that playing C^j is no longer optimal, he will defect. If player i anticipates this, he might want to defect as well. Unravelling must be prevented. This requires us to understand players' inferences. To keep those manageable, we

specify rewards in the event $\neg\Phi_T^j$ so that, for the sake of computing best-replies, player i may condition on the event that h_t^j is not yet in $\neg\Phi_t^j$. And we make sure that player j never plays D^j before this is the case.

To ensure that player i can condition on the event Φ_t^j, we specify W_C^j so that, as soon as (if ever) the history of player j belongs to $\neg\Phi_t^j$, further variations to player i's reward are computed according to the *bi-linear* test, defined as follows.

Definition 1. *Fix a collection of values $K(a^j, y^j) \geq 0$ such that player i's expected payoff*

$$g^i(a^i a^j) - \sum_{y^j} \pi(y^j | a^i a^j) K(a^j, y^j)$$

is independent of $a^i a^j$. A bi-linear test *assigned by player j rewards the constant $-K(a^j, y^j)$ whenever player j plays action a^j and receives signal y^j.*

Under our assumption on π, such values can be found. If player j subtracts such a (appropriately discounted) constant from the terminal reward for each of the remaining periods, then, conditional on this event, player i is indifferent over all action profiles; indeed, he is then indifferent over all sequences over action profiles for the remaining periods. More formally, define $\tau \leq T$ as the stopping time in the review phase such that $h_t^j \in \Phi_t^j$ for all $t \leq \tau$ and $h_{\tau+1}^j \in \neg\Phi_{\tau+1}^j$.[4] Taking discounting into account, the "continuation" reward assigned by player j for periods $\tau+1$ through T, as a function of his private history h_T^j, will be

$$-\delta^{-T} \sum_{t=\tau+1}^{T} \delta^t K(a_t^j, y_t^j) \tag{2}$$

where δ^{-T} discounts the reward back to time T of the review phase. This specification of W_C^j allows us to condition on the event Φ_t^j.

Because player i can condition on the event Φ_t^j, and assuming for now that player j cooperates on that event, player i can treat the signals that he receives as i.i.d. Given his inferences, will he then find it optimal himself to cooperate as long as his history belongs to Φ_t^i? This is where Assumption *2* plays a key role. It ensures both that player i's score (about j) is a sufficient statistic for his beliefs about j's score about him, and that beliefs are *contracting*: player i always believes that j's score about him is closer to its mean than i's score about j. See Figure 2. This guarantees that, as long as player i's score about j is in Φ_t^i, he views it as extremely unlikely that player j's score about him will ever exit Φ_t^j, provided that he keeps cooperating. Formally, we have:

[4]Let $\tau = T$ if $h_t^j \in \Phi_t^j$ for all $t \leq T$.

Lemma 1. *Under Assumption 2, for each $j = 1, 2$ there exists a collection of weights $\{\lambda^j(y^j), y^j \in Y^j\}$ with $\lambda^j(y^j) \in [0, 1]$, and a constant $0 < \beta < 1$ such that*

$$\mathbb{E}_{C^iC^j}[\lambda^j(y^j)] > \mathbb{E}_{D^iC^j}[\lambda^j(y^j)], \tag{3}$$

and

$$\mathbb{E}_{C^iC^j}[\lambda^j(y^j)|y^i] - \bar{\lambda}^j = \beta(\lambda^i(y^i) - \bar{\lambda}^i). \tag{4}$$

where $\bar{\lambda}^j := \sum_{y^j \in Y^j} \lambda^j(y^j)\pi(y^j|C^iC^j)$ is the unconditional mean of λ^j.

We fix throughout the paper such a collection of weights λ^1, λ^2. The proof of this lemma is in appendix. Condition (3) ensures that the weights are *capable* of motivating player i to cooperate since the expected increase in the score j assigns is higher when i cooperates than when i defects. Condition (4) ensures that, when both players are cooperating, given player i's private signal, his best predictor of the score assigned to him is a linear and increasing contraction of the score i himself is assigning. It follows that $\lambda^i(y^i)$ is a positively correlated sufficient statistic for player i's beliefs about $\lambda^j(y^j)$.

There is another place where Assumption **2** plays a key role in our construction. If player i conditions on player j using strategy $\mathcal{C}^j$ (as will be the case), but player j actually happens to use $\mathcal{D}^j$, player i's score about j will be significantly lower than what i would have expected. However, because the constant β is positive, and given the asymmetric structure of the reward schemes W_C^i, W_C^j (which provides incentives for unusually low, but not high scores), player i still expects player j's score about him to be such that he will continue to cooperate almost always, so that player i finds it optimal do so as well. In this way, the likely outcome of strategy $\mathcal{C}^i$ will involve cooperation in almost all periods, whether player j uses strategy $\mathcal{C}^j$ or $\mathcal{D}^j$.

We hope that this sketch will have provided the reader with some intuition regarding how we can construct W_C^i so as to ensure that playing C^i as long as $h_t^i \in \Phi_t^i$ is optimal. Of course, this does not provide a full description of player i's strategy $\mathcal{C}^i$. This will require the application of a fixed-point theorem. We have not explained either how we ensure that player i is actually indifferent between such a strategy and the strategy $\mathcal{D}^i$ that consists ni defecting always: the flexibility that we have in defining the "slope" of the reward function will be the key, as explained in Section 4.

So far, we have sketched how to ensure that both $\mathcal{C}^i$ and $\mathcal{D}^i$ are best responses to $(\mathcal{C}^j, W_C^j)$. How do we make sure that these are also best responses to $(\mathcal{D}^j, W_D^j)$? Because we do not know $\mathcal{C}^i$ explicitly (its specification on the events $\neg\Phi_t^i, t \geq 1$, is unknown), it appears difficult to "fine-tune" the reward function W_D^j. This is why our construct is conditionally belief-free: we specify

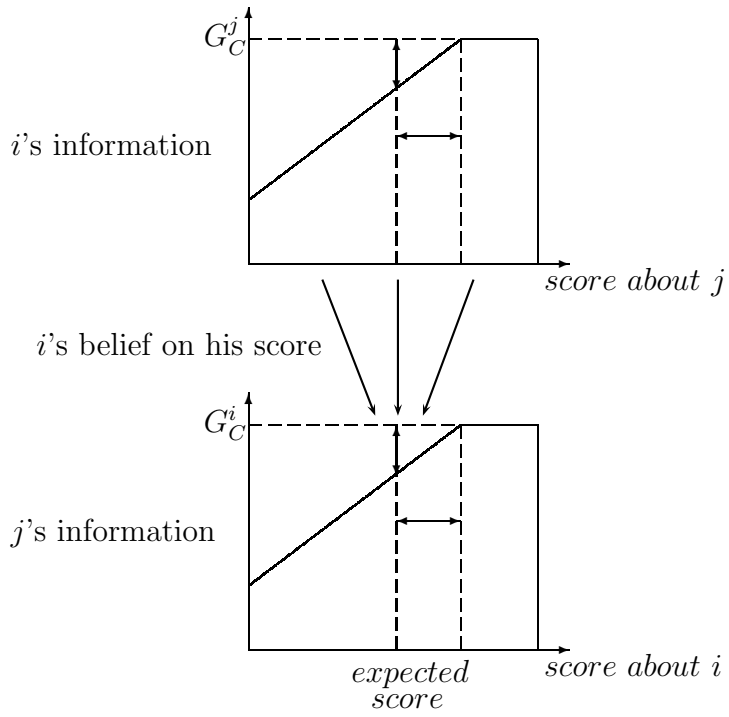

Figure 2: Player i's beliefs about j's score about him, given his score about j.

the reward function W_D^j so that *all* strategies are best-replies to $(\mathcal{D}^j, W_D^j)$. This means that W_D^j is very similar to the bi-linear test, but it is not quite the same: because player j plays the constant action D, it is not necessary for the constants to ensure that player i is indifferent across all action profiles, but only across those in which player j plays D. When player j follows $\mathcal{D}^j$, he rewards player i a constant amount K_D^i for each signal $\hat{y}^j$ received (as identified in Assumption **1**; fix one if several such signals exist). The value of K_D^i is chosen to make player i just indifferent between cooperating and defecting when player j defects. This is the *linear* test, defined next.

Definition 2. *A* linear test *rewards a constant $K_D^i(\hat{y}^j) \geq 0$ for each signal received that is equal to $\hat{y}^j$, where $K_D^i(\hat{y}^j)$ satisfies*

$$g^i(C^iD^j) + \pi(\hat{y}^j|D^jC^i)K_D^i(y^j) = g^i(D^iD^j) + \pi(\hat{y}^j|D^jD^i)K_D^i(y^j). \tag{5}$$

Under Assumption **1**, a linear test exists. We choose the reward function W_D^j to be a linear test. Let $(1-\delta)G_D$ be equal to the expected per-period payoff of player i when facing this linear test, i.e. $(1-\delta)G_D = g^i(D^iD^j) + \pi(\hat{y}^j|D^jD^i)K_D^i(\hat{y}^j)$. Formally, taking discounting into account, we can write the reward function W_D^j as

$$W_D^j(h_T^j) = G_D^i + K_D^i \sum_{t=1}^{T} \delta^{t-T} 1(y_t^j = \hat{y}^j), \tag{6}$$

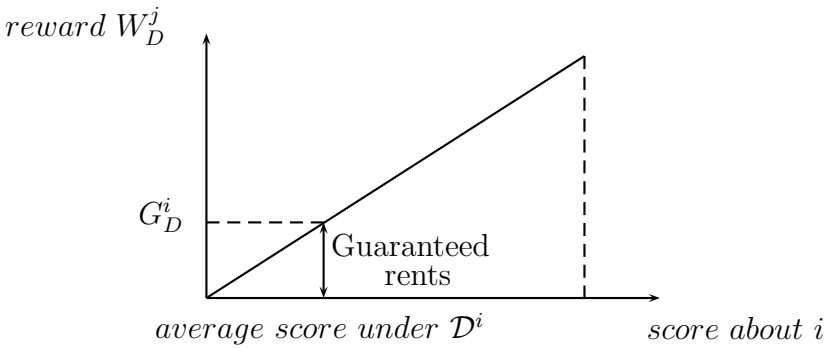

Figure 3: Inefficiency of the reward function W_D^j.

where h_T^j is player j's private history during the review phase, and $1(y_t^j = \hat{y}^j)$ is the indicator function that y_t^j, the time-t signal in history h_T^j, is equal to $\hat{y}^j$.

Notice that the linear test rewards player i on average even if he defects in every period. As a result, when player j is punishing player i by playing $\mathcal{D}^j$, player i's expected payoff G_D^i is bounded strictly above $g^i(D^iD^j)/(1-\delta)$. Assumption **1** is needed to ensure that this construction still leaves room to punish player i, i.e. that the resulting rents are such that $G_D^i < g^i(C^iC^j)/(1-\delta) \sim G_C^i$. See Figure 3.

We now turn to the formal construction. All proofs are in appendix.

4 Formal Statements

4.1 Inferences

Efficiency requires that players have incentives to play C almost always, and that provided they do, they expect to receive a reward (continuation payoff from the end of the phase onward) arbitrarily close to the maximum reward. As explained above, this means that they cannot be given incentives to always exert effort.

To achieve this, we specify that this reward be increasing in the appropriate score, but only up to a point. The range of scores over which this reward provides incentives must include the average score that a player receives if he keeps on cooperating. In fact, it should include all scores that are likely to arise if he does. That is, incentives should be provided for all scores that extend above the average score within a bound that is small relative to the length of a phase (so as to ensure that the expected reward is close to the maximum reward) but large relative to all

but the most unlikely statistical deviations from the mean score (so as to preserve incentives to cooperate almost always). To be concrete, a "slack" of order $T^{2/3}$ will do.

What is the appropriate score? Obviously, this score should be higher, on average, if a player cooperates than if he defects. Assumption **2** provides us with such a measure, λ^i, that satisfies a further property to be discussed shortly. However, because players do not cooperate after every history, there might be a value to learning about own's performance, and just because defecting leads to a lower expected score does not imply that the player's belief about his own score when he has defected will be first-order stochastically dominated by his belief about his performance when he has cooperated. This stronger, but desirable property need not hold in general. But it is easy to construct an alternative score which does. Instead of using $\lambda^i(\cdot)$ as the actual increment to a player's score, we might use this as the probability with which the increment to the score is 1, rather than 0. The (real-valued) sum of the former values is referred to as the *virtual* score, while the latter (integer-valued) sum of values is what we call the *real* score. Because increments in the real score are either 0 or 1, a higher expected increment necessarily corresponds to an improvement in the sense of first-order stochastic dominance.

A shortcoming of real scores, as opposed to virtual scores, is that player i's real score about j is no longer a sufficient statistic for i's belief about j's score about i. But it is easy enough to make sure that one measure tracks the other measure closely enough. We simply focus attention on the likely event in which the real and virtual scores are close to one another (say, within a range of order $T^{7/12}$, which makes it a very likely event) and specify rewards in the complementary, unlikely event in a way that allows players to view such events as irrelevant for the purpose of computing best-replies.[5]

This leads to the following definitions. Given h_t^i, the virtual score is

$$\Lambda_t^i := \sum_{\tau=3}^{t} \lambda^i \left(y_\tau^i\right),$$

and the real score is

$$L_t^i := \sum_{\tau=3}^{t} l_\tau^i,$$

where $\{l_\tau^i : \tau = 3, \ldots, t\}$ are independent Bernoulli random variables with mean $\lambda^i\left(y_\tau^j\right)$. [Observe that the summations start at $\tau = 3$: as explained in Subsection 4.2, the first two periods of a

[5]The rate 7/12 is smaller than 2/3, ensuring that this slack is negligible relative to the slack defined above, but this is unimportant.

Strategy	Period 1	Period 2	...	Period t	...	Period T
$\mathcal{C}^1$	C	$\frac{1}{2}C + \frac{1}{2}D$	...	mostly C	...	mostly C
$\mathcal{C}^2$	$\frac{1}{2}C + \frac{1}{2}D$	C	...	mostly C	...	mostly C

Figure 4: Strategies $\mathcal{C}^1$ and $\mathcal{C}^2$.

round play a special role in our construction.] Let $\bar{\lambda}^i$ denote the expected value of λ^i, conditional on both players cooperating.

We now introduce the three events of interest. The first, $\Phi_t'^i$, denotes the set of histories along which player i's observed signals about j lead to a real score that does not exceed the average score up to this period by more than the margin $T^{2/3}$. The second, $\Phi_t''^i$, refers to the histories along which real and virtual scores remain within $T^{7/12}$ of one another. Finally, Φ_t^i is the set of histories satisfying both requirements.

Definition 3. *For all $t = 1, \ldots, T$, let*

$$\begin{aligned} \Phi_t'^i &:= \left\{ h_t^i \in H_t^i : L_\tau^i \leq \tau \bar{\lambda}^i + T^{2/3} \text{ for all } \tau \leq t \right\}, \\ \Phi_t''^i &:= \left\{ h_t^i \in H_t^i : \left| L_\tau^i - \Lambda_\tau^i \right| \leq T^{7/12} \text{ for all } \tau \leq t \right\}, \\ \Phi_t^i &:= \Phi_t'^i \cap \Phi_t''^i. \end{aligned}$$

We now provide bounds on the relevant conditional probabilities of interest. These probabilities are predicated upon the assumption that the players' strategies belong to a particular class. Formally, we introduce

Definition 4. *Given the events $\{\Phi_t^j \ \ \forall t \leq T\}$, a T-period strategy of player j is from the* class *Z^j if it satisfies the following conditions:*

1. *In period $j \leq 3$, player j plays C.*
2. *In period $3 - j \leq 3$, player j plays C with probability $1/2$ and D with probability $1/2$.*
3. *In periods $t = 3, \ldots, T$, player j plays C so long as $h_t^j \in \Phi_t^j$*
4. *The action specified after h_t^j does not depend on the real score L_t^j.*

The motivation for the first and second point will be provided in Subsection 4.2. This definition is summarized in Figure 4.

The main statistical properties that will be needed are the following.

Lemma 2. *There exists $\alpha > 0$ such that, for all $t = 1, \ldots, T$:*

(i) conditional on both players using a strategy from Z^i,

$$\Pr\left[\neg\Phi_T^i\right] < \alpha^{-1}e^{-T^\alpha},$$

(ii) for all $h_t^i \in \Phi_t^i$, conditional on both players using a strategy from Z^i,

$$\Pr\left[\neg\Phi_T^j | h_t^i\right] < \alpha^{-1}e^{-T^\alpha},$$

(iii) for all $h_t^i \in H_t^i$, if j uses a strategy from Z^j, and player i always defects, for all $\tau = t, \ldots, T$,

$$\Pr\left[\neg\Phi_\tau''^j | h_t^i, \Phi_t^j\right] < \alpha^{-1}T^{-\alpha}.$$

The first bound ensures that cooperation is played in almost all periods. The second result ensures that, as long as player i's score about j remains in the event Φ_t^i of interest, he keeps assigning a probability arbitrarily close to one to his opponent's history belonging to this event, and remaining in it for all later periods –as long as both players keep cooperating on Φ_t^i. That is, player i is almost sure that his opponent will cooperate in all remaining periods, and that his own score observed by his opponent will not exceed the critical threshold. This is where Assumption **2** is really needed. Because player i's belief about j's score about him is always closer to its mean than i's score about j, player i views it as extremely unlikely that j's score about him will ever be outside Φ_t^j if his own score about j is not.

The final property ensures that the event Φ_t'' can also be ignored as relevant when player i defects, and this will be convenient when we shall show that defecting throughout is also a best response.

4.2 Incentives

By now, it should be clear how incentives will be provided for player i to cooperate as long as his score (about j) is in Φ_t^i, if he expects j to play the cooperative strategy $\mathcal{C}^j$ (that cooperates for every history in Φ_t^j). Indeed, it suffices for this that the reward be increasing in the score (as long as this score remains in Φ_t^j) at a rate that is strictly higher than the disutility of cooperating given that j cooperates, normalized by the difference (across actions) in the expected score. We can then pick the phase length to be large enough to make any other consideration (such as the possibility that a player's score leaves Φ_t^i) irrelevant. Similarly, if the slope is strictly lower, it is optimal to defect. In fact, *always* defecting would then be optimal, because there would

be no history of player j after which he would give i strict incentives to cooperate. Remember that our objective is to construct strategies that are belief-free from one round to the next: in particular, we must ensure that both the strategies $\mathcal{C}^i$, to be defined, must be best responses to both $(\mathcal{C}^j, W_C^j)$ and $(\mathcal{D}^j, W_D^j)$. The strategy $\mathcal{C}^i$ will belong to the class Z^i, but we shall not be able to describe its exact specification. The strategy $\mathcal{D}^i$ plays D always.

This raises two issues. First, to compute his best-responses, we shall find it convenient to specify W_D^j so that player i can condition on player j using strategy $\mathcal{C}^j$ (rather than $\mathcal{D}^j$). Second, we must ensure that player i is not only willing to play the cooperative strategy against $(\mathcal{C}^j, W_C^j)$, but is actually indifferent between $\mathcal{C}^i$ and $\mathcal{D}^i$.

The first issue is easy to take care of. Because $\mathcal{D}^j$ specifies defection always, and applies the linear test that makes player i indifferent across his own actions (given j's fixed action), player i might ignore this event, given that all strategies are equally good in that case. As explained above, the linear test is inefficient, because it treats signals independently, and thus fails to aggregate information, providing thereby a non-negligible reward to a player who in fact always defects. Assumption **1** is precisely what guarantees that the average payoff of player i when j plays $\mathcal{D}^j$ is still less than the efficient payoff $g^i(C^iC^j)$. Recall that in Section 2 we introduced the linear reward function:

$$W_D^j(h_T^j) = G_D^i + K_D^i \sum_{t=1}^{T} \delta^{t-T} 1(y_t^j\left(h_T^j\right) = \hat{y}^j), \quad \text{where} \quad (1-\delta)G_D^i = g^i(D^iD^j) + K_D^i \pi(\hat{y}^j|D^jD^i).$$

We may now state the following.

Proposition 1. *Suppose that $T = O((1-\delta)^{-1/2})$. If player j is playing the strategy $\mathcal{D}^j$ of defecting in every period and assigns to player i the reward function W_D^j, then:*

(i) player i is indifferent between all T-period strategies;

(ii) it holds that

$$\lim_{T\to\infty} (1-\delta)G_D^i < g^i(C^iC^j).$$

The second point motivates some finer points of our construction. Of course, we could also choose the rate which higher rates get rewarded (under $(\mathcal{C}^j, W^j)$) to increase at exactly the right rate that makes player i indifferent in the initial period between cooperating and defecting. However, this would imply that, no matter how unlikely the event Φ_t^j might be, the possibility that it might realize will still enter player i's calculations regarding player i's optimal choice in, say, the second period of the phase (because he was exactly indifferent in the first). In particular,

the signal that he receives in the first period might outweigh the action that he played, and he might find it optimal to switch from cooperation to defection, or vice-versa. This, of course, is not consistent with our desired specification of $\mathcal{C}^i$, or $\mathcal{D}^i$, and this is avoided as follows.

To make sure player i's initial action reinforces his incentives to take the same action in later periods, we make the rate at which j rewards i's score contingent on the signal that he receives (and the action he plays) in the initial periods. This is done such that player i's belief about this rate be (i) independent of i's own signal in that initial period, and (ii) above the critical threshold if he cooperated, and below it if he defected. If these rates are determined sequentially (say, the rate at which player i will be rewarded is determined in period i while player $j \neq i$ randomizes in the period in which i's rate gets determined), the possibility of finding such rates is ensured by the full rank assumption. Indeed, let

$$\tilde{M} := \begin{bmatrix} \pi(C^j y_1^j | C^i y_1^i) & \dots & \pi(C^j y_N^j | C^i y_1^i) & \pi(D^j y_1^j | C^i y_1^i) & \dots & \pi(D^j y_N^j | C^i y_1^i) \\ \vdots & & \vdots & \vdots & & \vdots \\ \pi(C^j y_1^j | C^i y_N^i) & \dots & \pi(C^j y_N^j | C^i y_N^i) & \pi(D^j y_1^j | C^i y_N^i) & \dots & \pi(D^j y_N^j | C^i y_N^i) \\ \pi(C^j y_1^j | D^i y_1^i) & \dots & \pi(C^j y_N^j | D^i y_1^i) & \pi(D^j y_1^j | D^i y_1^i) & \dots & \pi(D^j y_N^j | D^i y_1^i) \\ \vdots & & \vdots & \vdots & & \vdots \\ \pi(C^j y_1^j | D^i y_N^i) & \dots & \pi(C^j y_N^j | D^i y_N^i) & \pi(D^j y_1^j | D^i y_N^i) & \dots & \pi(D^j y_N^j | D^i y_N^i) \end{bmatrix},$$

computed under the assumption that player j randomizes equally between D^j and C^j. Note that the first N rows are player i's beliefs, conditional on each of his possible signals, about player j's action-signal pair in that period, when i himself plays C^i. The last N rows are his corresponding beliefs when he plays D^i. The following lemma states that weights can be found, which depend on player j's action-signal pair, so that player i's posterior belief about the expected weight does not depend on his private signal, but differs according to the action that he played (at least when the difference in those expected weights is low enough).

Lemma 3. *For any $b_C \in \mathbb{R}$, there exists $\bar{\varepsilon} > 0$ and $\kappa > 0$ such that, for all $\varepsilon \in (0, \bar{\varepsilon})$, if*

$b^i_D \in (b^i_C - \varepsilon, b^i_C + \varepsilon)$, *then the system*

$$\tilde{M} \begin{bmatrix} b(C^j y^j_1) \\ \vdots \\ b(C^j y^j_N) \\ b(D^j y^j_1) \\ \vdots \\ b(D^j y^j_N) \end{bmatrix} = \begin{bmatrix} b^i_C \\ \vdots \\ b^i_C \\ b^i_D \\ \vdots \\ b^i_D \end{bmatrix}, \tag{7}$$

has a solution $b(a^j y^j) \in (b^i_C - \kappa\varepsilon, b^i_C + \kappa\varepsilon)$ *for* $a^j \in \{D^j, C^j\}$ *and* $y^j \in Y^j$.

We may now state:

Proposition 2. *Suppose that* $T = O((1-\delta)^{-1/2})$, *and fix* $\varepsilon > 0$. *For any strategy of player* j *in* Z^j, *we can define a reward function* W^j_C *such that:*

(i) the maximum over all T*-period strategies of player* i

$$G^i_C := \max \mathbb{E}\left[\sum_{s=1}^{T} \delta^{s-1} g^i(a^i_s a^j_s) + \delta^T W^j_C(h^j_T)\right]$$

is achieved by both a strategy from class Z^i, *and by* $\mathcal{D}^i$.

(ii) it holds that

$$\lim_T (1-\delta) G^i_C > g^i(C^i C^j) - \varepsilon.$$

Note: This reward function is defined explicitly in Definition 5 in the proof of this proposition (in the Appendix).

It follows from these two propositions that, given some strategy $\mathcal{C}^j$ from class Z^j, and given $\mathcal{D}^j$, we can find W^j_C, W^j_D such that both $\mathcal{D}^i$ and some strategy from class Z^i are best responses. Furthermore, if player j uses $(\mathcal{C}^j, W^j_C)$, player i's average payoff from playing either best response is arbitrarily close to his efficient payoff $g^i(C^i C^j)$, when the horizon is long enough, while his average payoff is bounded below this level if player j uses $(\mathcal{D}^j, W^j_D)$.

4.3 Defining Strategies within a Phase

If player i knew that j's score about i lay outside of Φ^j_t, player j's action would be of no importance, given that the bilinear test makes player i indifferent over all action profiles. In fact,

player j's behavior outside of Φ_t^j is irrelevant for i's inferences, given i's history, and player i might as well condition on player j cooperating always. Nonetheless, this behavior still affects player i's overall payoff in the phase, because it affects the probability with which each of i's private histories realizes. Therefore, the exact definition of the reward function that j uses depends on j's entire strategy $\mathcal{C}^j$ (not just on its restriction to Φ_t^j), and of course, this definition must also depend on $\mathcal{C}^i$ if it is to make player i precisely indifferent between the strategies $\mathcal{C}^i$ and $\mathcal{D}^i$.

This implies that strategies $(\mathcal{C}^i, \mathcal{C}^j)$ and reward functions (W^i, W^j) must be defined jointly, and such a definition requires the application of a fixed-point theorem. However, in the application of a fixed-point theorem, there is a key observation that allows us to focus simply on correspondences between pairs of strategies, rather than both strategies and reward functions. As discussed in the proof of Proposition 2, we can define a set of reward functions W_C^j parametrized by a constant $\bar{c}^j$ such that for any given strategy $\mathcal{C}^j$ there is a unique reward function in that set that makes player i indifferent between following a strategy from class Z^i and strategy $\mathcal{D}^i$.

Applying Kakutani's fixed-point theorem leads us to the following proposition.

Proposition 3. *For all sufficiently large T, there are reward functions W_C^i and T-period strategies $\mathcal{C}^i$ from class Z^i for $i = 1, 2$, such that both $\mathcal{D}^i$ and $\mathcal{C}^i$ are best responses to both $(\mathcal{C}^i, W_C^i)$ and $(\mathcal{D}^i, W_D^i)$, for $i = 1, 2$. These strategies and reward functions satisfy the conclusions of Propositions 1 and 2.*

4.4 The Equilibrium of the Supergame

It remains to specify what strategies players use in the supergame. This part of the construction is standard. The infinite horizon is divided in review phases of length T, and in each of those phases players use either $\mathcal{C}^i$ or $\mathcal{D}^i$ as a function of their private history, so as to achieve the promised continuation payoff to their opponent. More precisely, to achieve efficiency, players use $\mathcal{C}^i$ in the first round, and from that point on, given the reward W_C^i or W_D^i that is promised at the end of a given phase, they randomize at the beginning of the next between both strategies so as to achieve the exact payoff in $[G_D^i, G_C^i]$ that is needed. Of course, by varying the choice of the strategy chosen in the initial period, every payoff in the square $[G_D^1, G_C^1] \times [G_D^2, G_C^2]$ can be achieved. This is formally established in the following proposition.

Proposition 4. *Suppose that for $i = 1, 2$ and some $T > 0$, there are T-period strategies $\mathcal{C}^i$ and $\mathcal{D}^i$ and reward functions $W_C^i : H_T \to [G_D^j, G_C^j]$ and $W_D^i : H_T \to [G_D^j, G_C^j]$ that satisfy the following conditions.*

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

$$G_C^i = \max \mathbb{E}\left[\sum_{s=1}^{T} \delta^{s-1} g^i(a_s^i a_s^j) + \delta^T W_C^j(h_T^j)\right], \tag{8}$$

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

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

$$G_D^i = \max \mathbb{E}\left[\sum_{s=1}^{T} \delta^{s-1} g^i(a_s^i a_s^j) + \delta^T W_D^j(h_T^j)\right], \tag{9}$$

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

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 sequential equilibrium of an infinitely repeated game with discount factor δ.

Our main result, Theorem 1, follows from Propositions 3 and 4.

5 Concluding Comments

This paper has established that efficiency can be achieved under imperfect private monitoring under certain conditions. Our result, then, raises three questions: (i) can the two assumptions be weakened? (ii) can the result be strengthened to a folk theorem? (iii) can the analysis be extended to a broader set of games?

Our construction, as much of the constructions in the repeated games with imperfect monitoring literature is belief-free by blocks in the following sense: during each block of size T, each player i is indifferent between two strategies $\mathcal{C}^i$ and $\mathcal{D}^i$, whether the opponent, player j, plays $\mathcal{C}^j$ or $\mathcal{D}^j$. This type of construction was first applied successfully by Matsushima (2004) under a conditional independence assumption, with $\mathcal{C}^i$ and $\mathcal{D}^i$ playing the constant action C and D respectively. Without conditional independence, it is not possible to have $\mathcal{C}^i$ play C constantly against $\mathcal{C}^j$ while maintaining efficiency. The reason for this is that in order to provide incentives for $\mathcal{C}^i$ to play C after any history of signals, the highest reward provided must be significantly larger than the average reward, which implies that the average reward must be significantly lower than the efficient payoff. Since not much is known *a priori* about $\mathcal{C}^i$ and we need that it is a best response to $\mathcal{D}^j$, we ensure that every strategy is a best-response to $\mathcal{D}^i$. This entails a lower bound on achievable payoffs in our construction, and Assumption **1** ensures that this restriction

still does not rule out the efficient payoff. Relaxing Assumption **1** would require a construction in which $\mathcal{D}^i$ together with its reward scheme is tailored to $\mathcal{C}^i$. The system of algebraic inequalities stating that an arbitrary strategy $\mathcal{C}^i$ is a best-response both to $\mathcal{C}^j$ and to $\mathcal{D}^j$ does not appear easier to satisfy than the corresponding system of equalities, which was our starting point. In order to pursue this relaxation of Assumption **2**, one needs to rely more deeply on some information on $\mathcal{C}^i$. Preliminary research suggests that one can construct reward schemes in order that $\mathcal{C}^i$ is a trigger strategy, but much works remains to be done in that direction.

We can obtain asymmetric payoffs that give one of the two players a payoff above $g^i(C^iC^j)$ by following the same methods as in Ely, Hörner and Olszewski (2005). Here is a sketch. Suppose that, in some fixed review phases that are regularly intersparsed among phases that are otherwise identical to those described above, players are not both supposed to be indifferent between $\mathcal{C}^i$ and $\mathcal{D}^i$. Rather, in those phases, player 2, say, has a strict incentive to play $\mathcal{D}^2$, while player 1 is indifferent between $\mathcal{C}^1$ and $\mathcal{D}^1$.[6] Because player 2 does not need to be willing to play a cooperative strategy, the reward function W_D^1 that is then used by player 1 need not be the linear test, so that this regime (in the sense of Ely, Hörner and Olszewski, 2005) is associated with a range of payoffs for player 2 that tends to $[g^2(D^1D^2)/(1-\delta), g^2(C^1D^2)/(1-\delta)]$ as $\delta \to 1$. We must, however, make an assumption that parallels Assumption **2**, obtained by replacing all references to the matrix M^i by the matrix $\hat{M}^i = (\pi(y^j|C^1D^2, y^i)_{y^iy^j}$. For player 1, then, this regime is associated with continuation payoffs that are not sustainable *per se*, as player 1 can secure $g^1(D^1D^2)$, yet he must be willing to play in a way that gives him a flow payoff equal to $g^1(C^1D^2)$. As in Ely, Hörner and Olszewski (2005), we must then make sure that the relative frequency of both types of regimes is such that the average of what a player i can secure across blocks (his lowest continuation payoff) is below what his opponent can make sure that player i gets, for some optimal strategy of player j. This puts an upper bound on the relative frequency of the asymmetric regime in which player 2's optimal strategy is $\mathcal{D}^2$, namely, from the constraint on player 1's range of equilibrium payoffs, it cannot exceed

$$\frac{g^1(C^1C^2) - \lim_\delta (1-\delta)G_D^1}{g^1(C^1C^2) - g^1(C^1D^2) + g^1(D^1D^2) - \lim_\delta (1-\delta)G_D^1},$$

which is in $(0,1)$. Indeed, the resulting payoff vector lies on the Pareto-frontier, and gives player 1 a payoff strictly below G_D^1. If the limit of $(1-\delta)G_D^1$ were precisely equal to $g^1(D^1D^2)$, this would give us the folk theorem, but as it stands, only a subset is obtained.

Obviously, the third question is more ambitious. Given that we do not yet have a character-

[6] Of course, the strategy $\mathcal{C}^1$ need not be exactly the same than in our current construction.

ization of the set of individually rational payoffs in general (see Gossner and Hörner, 2010, for some results in this direction), the case of two players appears to be the best place to start.[7]

[7]Alternatively, one may want to start with signal structures that are sufficiently rich, as in Sugaya (2010).

A Appendix

A.1 Proof of Lemma 1

Let λ'^1 be an eigenvector of M^1M^2 associated to the eigenvalue $\beta_0 > 0$ as in Assumption 2.

First we show that $\beta_0 < 1$. Since M^1M^2 is a stochastic matrix, $\beta_0 \leq 1$. Since M^1 and M^2 are stochastic, the only eigenvectors of M^1M^2 associated to the eigenvalue 1 are multiples of the constant vector 1_n, but this case is excluded by the requirement that the expectation of λ'^1 under C^1C^2 is higher than under C^1D^2.

Let $\beta = \sqrt{\beta_0}$, and $\lambda'^2 = \beta M^2\lambda'^1$. For $i, j = \{1, 2\}$, $M^j\lambda'^i = \beta\lambda'^j$. With $A = \min_{i\in 1,2} \min_{a_i} \lambda^i_{y^i}$ and $B = \max_{i\in 1,2} \max_{a_i}(\lambda^i_{y^i} - A) > 0$, we let $\lambda^i(y^i) = \frac{\lambda^i_{y^i} - A}{B}$.

Now we verify that the families of weights λ^1, λ^2 satisfy the requirements of Lemma 1. First, their definition ensures $\lambda^i(y^i) \in [0, 1]$ for every i and a^i. Second, from Assumption 2,

$$\mathbb{E}_{C^iC^j}[\lambda^j(y^j)] = \frac{1}{B}(\mathbb{E}_{C^iC^j}\lambda'^j_{y^j} - A) > \frac{1}{B}(\mathbb{E}_{D^iC^j}\lambda^j_{y^j} - A) = \mathbb{E}_{D^iC^j}[\lambda^j(y^j)].$$

Finally, let λ^i, $i = 1, 2$ denote the $(1, n)$ matrix given by $\lambda^i_{y^i} = \lambda^i(y^i)$, and let E^i_C be the $(1, n)$ matrix defined by $E^i_{C,y^i} = \mathbb{E}_{C^iC^j}[\lambda^j(y^j)|y^i] = \sum_{y^j} M^i_{y^i,y^j}\lambda^j_{y^j}$. In matrix notation:

$$\begin{aligned} E^i_C &= M^i\lambda^j = M^i\frac{1}{B}(\lambda'^j - A1_n) \\ &= \frac{\beta}{B}\lambda'^i - \frac{A}{B}1_n = \beta\frac{1}{B}(\lambda'^i - A1_n) + (\beta - 1)\frac{A}{B}1_n \\ &= \beta\lambda^i + (\beta - 1)\frac{A}{B}1_n. \end{aligned}$$

Hence for every y^i:

$$\begin{aligned} \mathbb{E}_{C^iC^j}[\lambda^j(y^j)|y^i] &= \beta\lambda^i(y^i) + (\beta - 1)\frac{A}{B} \\ \mathbb{E}_{C^iC^j}[\lambda^j(y^j)|y^i] - \bar{\lambda}^j &= \beta(\lambda^i(y^i) - \bar{\lambda}^i) + (\beta - 1)\frac{A}{B} + \beta\bar{\lambda}^i - \bar{\lambda}^j. \end{aligned} \tag{10}$$

Note that $\bar{\lambda}^j = \mathbb{E}_{C^iC^j}[\mathbb{E}_{C^iC^j}[\lambda^j(y^j)|y^i]]$, and $\bar{\lambda}^i = \mathbb{E}_{C^iC^j}[\lambda^i(y^i)]$. Taking expectations over y^i in (10) gives:

$$(\beta - 1)\frac{A}{B} + \beta\bar{\lambda}^i - \bar{\lambda}^j = 0,$$

and (10) becomes

$$\mathbb{E}_{C^iC^j}[\lambda^j(y^j)|y^i] - \bar{\lambda}^j = \beta(\lambda^i(y^i) - \bar{\lambda}^i),$$

which is the desired result.

A.2 Proof of Lemma 2

The proof of Lemma 2 relies on the following large deviations result, see e.g. Alon and Spencer (2008).

Lemma 4. *Let $y_1, \dots y_n$ be a mutually independent family of random variables with $\mathbb{E}[y_i] = \bar{y}_i$ and $|y_i - \bar{y}_i| \leq 1$. Then, for every $a > 0$*

$$\Pr[\sum_{t=1}^{n} y_t > \sum_{t=1}^{n} \bar{y}_t + a] \leq e^{-\frac{a^2}{2n}}.$$

We first estimate $\Pr[\neg\Phi_T''^i | h_T^i] = \Pr[\neg\Phi_T''^i | h_T^i h_T^j]$, for any h_T^i, h_T^j. From Lemma 4 above, for any τ,

$$\Pr[\left| L_\tau^i - \Lambda_\tau^i \right| > T^{7/12}] \leq 2e^{-\frac{1}{2}T^{2/12}}.$$

Hence

$$\Pr[\neg\Phi_T''^i | h_T^i] \leq 2Te^{-\frac{1}{2}T^{2/12}}. \tag{11}$$

Note that the probabilities in (i) and (ii) of the Lemma are unchanged if each player plays the constant strategy that specifies C after all histories. We therefore estimate these probabilities under this assumption.

Proof of (i) From Lemma 4,

$$\Pr[\exists\tau, L_\tau^i > \tau\bar{\lambda}^i + T^{2/3}] \leq Te^{-\frac{1}{2}T^{1/3}}.$$

Combining with (11) we obtain

$$\Pr[\neg\Phi_T^i] \leq 3Te^{-\frac{1}{2}T^{2/12}}.$$

Proof of (ii) Conditional on $h_t^i \in \Phi_t^i$, the distribution of $\lambda_1^j, \dots, \lambda_t^j$ is the one of mutually independent random variables. From Lemma 1, for any $\tau \leq T$

$$\begin{aligned} \mathbb{E}[L_\tau^j | h_t^i] &= \sum_{t' \leq \max(\tau,t)} [\bar{\lambda}^j + \beta(\lambda_t^i - \bar{\lambda}^i)] + (\tau - t)^+ \bar{\lambda}^j \\ &\leq \tau\bar{\lambda}^j + \beta\tau T^{2/3}. \end{aligned}$$

From Lemma 4,

$$\Pr[L_\tau^j > \tau\bar{\lambda}^j + T^{2/3}] \leq e^{-(1-\beta)^2 T^{1/3}},$$

and combining with (11)

$$\Pr\left[\neg\Phi_T^j|h_t^i\right] \leq Te^{-(1-\beta)^2T^{1/3}} + 2Te^{-\frac{1}{2}T^{2/12}}.$$

Proof of (iii) For any h_t^i, h_t^j we decompose

$$\Pr[\Phi_\tau''^j|h_t^i,h_t^j] = \Pr[\Phi_t''^j|h_t^i,h_t^j]\Pr[\Phi_\tau''^j|h_t^i,h_t^j,\Phi_t''^j] + \Pr[\neg\Phi_t''^j|h_t^i,h_t^j]\Pr[\Phi_\tau''^j|h_t^i,h_t^j,\neg\Phi_t''^j].$$

For $t \leq \tau$, $\Pr[\Phi_\tau''^j|h_t^i,h_t^j,\neg\Phi_t''^j] = 0$, hence

$$\Pr[\neg\Phi_\tau''^j|h_t^i,h_t^j,\Phi_t''^j] \leq \Pr[\neg\Phi_\tau''^j|h_t^i,h_t^j] \leq 2Te^{-\frac{1}{2}T^{2/12}}.$$

Now,

$$\Pr[\neg\Phi_\tau''^j|h_t^i,\Phi_t^j] = \sum_{h_t^j\in\Phi_t^j}\Pr[h_t^j|h_t^i,\Phi_t''^j,\Phi_t^j]\Pr[\neg\Phi_\tau''^j|h_t^i,h_t^j,\Phi_t''^j] \leq 2Te^{-\frac{1}{2}T^{2/12}}.$$

Hence the result.

A.3 Proof of Lemma 3

Note that $\tilde{M}$ is stochastic, i.e. if U denotes the unit vector, $\tilde{M}U = U$. Also, $\tilde{M}$ is generically (in the monitoring structure) invertible. Let V be such that $\tilde{M}V$ has its N first components equal to 0, and its N last equal to 1. Then, $b_C^iU + (b_D^i - b_C^i)V$ satisfies equation 7. For b_D^i sufficiently close to b_C^i, all coefficients of $b_C^iU + (b_D^i - b_C^i)V$ are strictly positive.

A.4 Proof of Proposition 1

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

$$\mathbb{E}\left[\sum_{s=1}^T \delta^{s-1}g^i(a_s^iD^j) + \delta^T W_D^j(h_T^j)\right] = \mathbb{E}\left[\sum_{s=1}^T \delta^{s-1}(g^i(a_s^iD^j) + K_D^i\,1(y_s^j=\hat{y}^j))\right] + \delta^T G_D^i$$

$$= \sum_{s=1}^T \delta^{s-1}(g^i(D^iD^j) + K_D^i\,\pi(\hat{y}^j|D^iD^j)) + \delta^T G_D^i = G_D^i,$$

because the expectation of $g^i(a_s^iD^j) + K_D^i\,1(y_s^j=\hat{y}^j)$ 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

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

Now, from Assumption **1**, we have the following:

$$\pi(\hat{y}^j|D^jD^i)(g^i(C^iC^j)-g^i(C^iD^j))<\pi(\hat{y}^j|D^jC^i)(g^i(C^iC^j)-g^i(D^iD^j)) \quad \Rightarrow$$

$$g^i(D^iD^j)\pi(\hat{y}^j|D^jC^i)-g^i(C^iD^j)\pi(\hat{y}^j|D^jD^i)<g^i(C^iC^j)(\pi(\hat{y}^j|D^jC^i)-\pi(\hat{y}^j|D^jD^i)) \quad \Rightarrow$$

$$G_D^i=\frac{g^i(D^iD^j)+K_D^i\pi(\hat{y}^j|D^jD^i)}{1-\delta}=\frac{g^i(D^iD^j)\pi(\hat{y}^j|D^jC^i)-g^i(C^iD^j)\pi(\hat{y}^j|D^jD^i)}{(1-\delta)(\pi(\hat{y}^j|D^jC^i)-\pi(\hat{y}^j|D^jD^i))}<\frac{g^i(C^iC^j)}{1-\delta}.$$

A.5 Proof of Proposition 2

We begin by defining the reward functions. To do so, we first define a linear test K_C^i by

$$g^i(C^iC^j)+\pi(\hat{y}^j|C^jC^i)K_C^i=g^i(D^iC^j)+\pi(\hat{y}^j|C^jD^i)K_C^i,$$

where i gets an additional K_C^i when $y^j=\hat{y}^j$ such that i is indifferent between both actions if j plays C^j. Also, recall the function $K(\cdot)$ from Definition 1 of the bi-linear test.

Definition 5. *Given $\varepsilon\in(0,\bar{\varepsilon}/2)$, let*

$$b_C^i:=b_0^i+\varepsilon, b_D^i:=b_0^i-\varepsilon,$$
$$\text{where } b_0^i:=\frac{g^i\left(D^iC^j\right)-g^i\left(C^iC^j\right)}{\sum_y^j\left(\pi\left(y^j|C^jC^i\right)-\pi\left(y^j|C^jD^i\right)\right)\lambda\left(y^j\right)},$$

and let $b(a^jy^j)$ denote the corresponding solution of equation (7) whose existence is shown in Lemma 3.

We define the reward function W_C^j that player j uses to reward player $i=3-j$ while following a strategy $\mathcal{C}^j$ from class Z^j as

$$W_C^j(h_T^j)=\bar{c}^j1\left(y_{3-j}^j=\hat{y}^j\right)+\delta^{j-T}K_C^i1\left(y_j^j=\hat{y}^j\right)+\sum_{t=3}^{\tau^j}b(a_{3-j}^jy_{3-j}^j)1\left(l_t^j=1\right)-\delta^{t-T}\sum_{t=\tau^j+1}^{T}K\left(a_t^jy_t^j\right),$$

where

$$\tau^j=\left\{\inf t:h_{t+1}^j\in\neg\Phi_{t+1}^j\right\}$$

is the random stopping time at which player j's history first leaves Φ_t^j, and $\bar{c}^j$ is a constant that depends on $\mathcal{C}^j$.

Let $\mathcal{W}_C^j$ denote the set of reward functions satisfying the above.

We select $\bar{c}^j$ as follows. Observe that given player i's first action and signal $(h^i_{1i} = (a^i_i, y^i_i))$, and given player j's strategy, player i's optimal continuation strategy in the T-stage repeated game is independent of the specification of $\bar{c}^j$ (since the latter depends only on y^j_{3-j}). We pick the unique $\bar{c}^j$ such that player i is just indifferent between playing C^i and D^i in period i, given that player j randomizes equally between both actions in that period (i.e. follows a strategy from Z^j). Observe that, because all values of $b(a^j_i y^j_i)$ are within $4\kappa\varepsilon$ of each other, if the event $h^j_t \in \neg\Phi^j_t$ is arbitrarily unlikely under the optimal strategy, then the value of $\bar{c}^j$ is of the order εT.

We now check the three claims whose validity the proposition asserts. Throughout, fix a strategy s^j in Z^j.

Claim: Some strategy in Z^i is optimal. The indifference in periods 1 and 2 follows from the definition of W^j_C, so let us assume that player i has played C in period i, and let us show that it is optimal to play C for $h^i_t \in \Phi^i_t$, for all periods $t \geq 3$. Let us define W'^j_C as

$$W'^j_C\left(h^j_T\right) = \delta^{j-T} K^{3-j}_C 1\left(y^j_j = \hat{y}^j\right) + c^j\left(y^j_{3-j}\right) + \sum_{t=3}^{T} b(a^j_{3-j} y^j_{3-j}) 1\left(l^j_t = 1\right),$$

and s'^j as the strategy in Z^j that cooperates in every period $t \geq 3$. That is, W'^j_C and W^j_C only differ in the specification of the rewards on the event $\neg\Phi^j_t$. Because of the definition of K, it follows that the payoff of any given strategy s^i against s'^j and $W'^j_C\left(h^j_T\right)$ is weakly higher than against s^j and $W^j_C\left(h^j_T\right)$. Because player i has played C in period i, and the expected value of $b(A^j_i y^j_i)$ conditional on C is b^i_C (independently of i's signal in period i) a continuation strategy $s^i|h^i_t$ is optimal against s'^j and $W'^j_C\left(h^j_T\right)$ if it is optimal against

$$W''^j_C\left(h^j_T\right) = \sum_{t=3}^{T} b^i_C 1\left(l^j_t = 1\right),$$

and since $b^i_C > b^i_0$, it follows that the unique continuation strategy that is optimal against s'^j and $W'^j_C\left(h^j_T\right)$ consists of playing C after history h^i_t. Furthermore, the gain from playing C rather than D is bounded away from 0, because $b^i_C > b^i_0$, independently of T. Observe now that, because $\Pr\left[\neg\Phi^j_T | h^i_t\right] < e^{-T^\alpha}$ for $h^i_t \in \Phi^i_t$, the continuation payoff from playing s'^i (i.e. playing C always) against s'^j and $W'^j_C\left(h^j_T\right)$, after a history $h^i_t \in \Phi^i_t$, tends to the continuation payoff against s^j and $W^j_C\left(h^j_T\right)$. That is, for T large enough, playing C after history $h^i_t \in \Phi^i_t$ is optimal against s^j and $W^j_C\left(h^j_T\right)$.

Claim: The strategy $\mathcal{D}^i$ in Z^i that plays D in every period is optimal. The indifference in periods 1 and 2 follows from the definition of W^j_C, so let us assume that player i has

played D in period i, and let us show that it is optimal to play D for $h_t^i \in \Phi_t^i$, $t \geq 3$. For this case, we define $W_C'''^j$ as:

$$W_C'''^j\left(h_T^j\right) = \sum_{t=3}^{\tilde{\tau}^1} b_D 1\left(l_t^j = 1\right) - \sum_{t=\tilde{\tau}^1+1}^{T} \delta^{t-T} K\left(a_t^j y_t^j\right),$$

where $\tilde{\tau}^j := \left\{\inf t : h_{t+1}^j \in \neg\Phi_{t+1}'^j\right\}$. That is, the only differences between $W_C'''^j$ and W_C^j are (i) the coefficient b_D^i which replaces $b\left(a_{3-j}^j y_{3-j}^j\right)$ and (ii) the region which triggers the bi-linear test; in the case of W_C^j, it is once h_T^j leaves the region Φ_{t+1}^j; in the case of $\tilde{W}_C^j$, it is when h_T^j leaves the region $\Phi_{t+1}'^j$ –a subset of Φ_{t+1}^j. Observe that replacing $b\left(a_{3-j}^j y_{3-j}^j\right)$ by b_D^i does not change the incentives of player i, because b_D^i is the expected value of $b\left(a_{3-j}^j y_{3-j}^j\right)$, conditional on i having played D in period i, independently of his signal in period i.

Observe that i's optimal continuation strategy against s'^j and $W_C'''^j$, conditional on the event $h_t^i \cap \Phi_t^j$, for *any* $h_t^i \in H_t^i$, consists in playing D always: indeed, the distribution of $\tilde{\tau}^j$ conditional on D always is (weakly) first-order stochastically dominated by the distribution of $\tilde{\tau}^j$ conditional on any other continuation strategy. Second, in any period in which $h_t^j \in \Phi_t^j$, and thus $h_t^j \in \Phi_t'^j$, the gain from playing D rather than C in the immediate period is bounded away from 0, because $b_0^i > b_D^i$.

Observe now that, because $\Pr\left[\Phi_\tau'^j \cap \neg\Phi_\tau''^j | h_t^i, \Phi_t^j\right] < T^{-\alpha}$ as long as players have played $D^i C^j$ in all periods $t' = 3, \ldots, t$, the distribution of τ^j conditional on the event $h_t^i \cap \Phi_t^j$ (given s^j) approaches the distribution of $\tilde{\tau}^j$ (given s'^j). So the payoff from playing against s^j and W_C^j tends to the payoff against s'^j and $W_C'''^j$ as $T \to \infty$. It follows that playing D is optimal for player i in period $t = 3$, and recursively, for any $t \geq 3$.

Claim: The payoff of player i is asymptotically efficient. We must show that, as $T \to \infty$,

$$(1-\delta) G_C^i \to g^i(C^i C^j).$$

As we have observed, c^i is of order $T\varepsilon$, and so is $\sum_{t=3}^{T}\left(b\left(a^i y^i\right) - b_0^i\right)$, for all $a^i = C, D$ and $y^i \in Y^i$. Finally, since playing some strategy from Z^i is optimal, $\Pr\left[\neg\Phi_T^i\right] < e^{-T^\alpha}$. Therefore, since $(1-\delta)T \to 0$, and rescaling $\varepsilon > 0$ if necessary,

$$(1-\delta) G_C^i > g^i(C^i C^j) - \varepsilon.$$

A.6 Proof of Proposition 3

We will use Kakutani's fixed point theorem to prove Proposition 3.

Fix some strategy $\hat{\mathcal{C}}^j \in Z^j$, and some $\epsilon \in (0, \frac{\bar{\epsilon}}{2})$. Consider the set of reward functions $\mathcal{W}_C^j$ defined by Definition 5 in the proof of Proposition 2. We know we can parametrize those reward functions by $\bar{c}^j$ and find a $\bar{c}$ such that if $\bar{c}^j > \bar{c}$ then i's best response is some $\hat{\mathcal{C}}^i \in Z^i$, and if $\bar{c}^j < \bar{c}$, i's best response is $\hat{\mathcal{D}}^i$.[8]

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

Let us prove that the correspondence Φ^i is upper hemi-continuous. Consider sequences $\hat{\mathcal{C}}_n^i \to \hat{\mathcal{C}}^i$ and $\hat{\mathcal{C}}_n^j \to \hat{\mathcal{C}}^j$ such that $\hat{\mathcal{C}}_n^i \in \Phi^i(\hat{\mathcal{C}}_n^j)$ for all n. Let us show that $\hat{\mathcal{C}}^i \in \Phi^i(\hat{\mathcal{C}}^j)$. Denote by $\bar{c}_n^j$ the lowest value of $\bar{c}^j$ for which $\hat{\mathcal{C}}_n^i$ is at least as good as $\hat{\mathcal{D}}^i$ in response to $\hat{\mathcal{C}}_n^j$. Without loss of generality, assume that $\bar{c}_n^j \to \hat{c}^j$ for some $\hat{c}^j$ (otherwise we can take a convergent subsequence). Then, by continuity, among all strategies from Z^i, $\hat{\mathcal{C}}^i$ gives player i the highest payoff in response to $\hat{\mathcal{C}}^j$ when $\bar{c}^j = \hat{c}^j$. This payoff equals player i's payoff from $\hat{\mathcal{D}}^i$. It follows that $\hat{\mathcal{C}}^i \in \Phi^i(\hat{\mathcal{C}}^j)$ if we show that for any $\bar{c}^j < \hat{c}^j$, $\hat{\mathcal{D}}^i$ is strictly better than any strategy from Z^i in response to $\hat{\mathcal{C}}^j$. Suppose not, i.e. $\hat{\mathcal{C}}' \in Z^i$ is better than $\hat{\mathcal{D}}^i$ for some $c' < \hat{c}^j$. Since player i's payoff is linear in $\bar{c}^j$ and $\hat{\mathcal{D}}^i$ is his strict best response for all sufficiently small $\bar{c}^j$ by the proof of Proposition 2, it follows that $\hat{\mathcal{C}}'$ is *strictly* better than $\hat{\mathcal{D}}^i$ in response to $\hat{\mathcal{C}}^j$ for $\bar{c}^j = \hat{c}^j > c'$, a contradiction.

We conclude that the correspondence $(\hat{\mathcal{C}}^1, \hat{\mathcal{C}}^2) \to (\Phi^1(\hat{\mathcal{C}}^2), \Phi^2(\hat{\mathcal{C}}^1))$ 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 $\hat{\mathcal{C}}^1$ and $\hat{\mathcal{C}}^2$ such that $\hat{\mathcal{C}}^1 = \Phi^1(\hat{\mathcal{C}}^2)$ and $\hat{\mathcal{C}}^2 = \Phi^2(\hat{\mathcal{C}}^1)$. Then for $i = 1, 2$ in response to $\hat{\mathcal{C}}^j$, a player i is indifferent between $\hat{\mathcal{C}}^i$ and $\hat{\mathcal{D}}^i$ for W_C^j defined by an appropriate value of $\bar{c}^j$. This completes the proof of Proposition 3, since by Proposition 2, it is always optimal to follow $\hat{\mathcal{D}}^i$ or a strategy from Z^i in response to any strategy from Z^j with a reward function $W_C^j \in \mathcal{W}_C^j$.

A.7 Proof of Proposition 4

For players $i = 1, 2$, define recursive strategies $\bar{\mathcal{C}}^i$ and $\bar{\mathcal{D}}^i$ of the infinitely repeated game as follows. Let us divide the timeline into T-period review phases. Strategy $\bar{\mathcal{C}}^i$ coincides with $\mathcal{C}^i$ over the first review phase, and $\bar{D}^i$ starts with $\mathcal{D}^i$. 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

[8] If $\bar{c} < 0$, the inequalities are reversed but the same argument for the proof holds.

player i has played $\mathcal{C}^i$ in the previous review phase and has observed private history h_T^i, then in the new review phase he follows the strategy

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

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

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

Let us show that both $\bar{\mathcal{C}}^i$ and $\bar{\mathcal{D}}^i$ are best responses to $\bar{\mathcal{C}}^j$ and $\bar{\mathcal{D}}^j$. 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 $\bar{\mathcal{C}}^j$ from any strategy that involves $\mathcal{C}^i$ or $\mathcal{D}^i$ in each review phase, and in particular strategies $\bar{\mathcal{C}}^i$ and $\bar{\mathcal{D}}^i$. Similarly, G_D^i is the payoff in response to $\bar{\mathcal{D}}^j$ 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 $\bar{\mathcal{C}}^j$ and $\bar{\mathcal{D}}^j$. 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 $\bar{\mathcal{C}}^j$ and $\bar{\mathcal{D}}^j$, 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 $\bar{\mathcal{C}}^j$. 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

$$\delta^T \frac{W_C^j(h_T^j) - G_D^i}{G_C^i - G_D^i} F_C^i + \delta^T \frac{G_C^i - W_C^j(h_T^j)}{G_C^i - G_D^i} F_D^i \leq$$

$$\delta^T (F_C^i - G_C^i) + \delta^T \frac{W_C^j(h_T^j) - G_D^i}{G_C^i - G_D^i} G_C^i + \delta^T \frac{G_C^i - W_C^j(h_T^j)}{G_C^i - G_D^i} G_D^i = \delta^T (F_C^i - G_C^i + W_C^j(h_T^j)).$$

Then, player i's expected payoff at time 1 cannot be greater than

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

by (8). This is less than F_C^i, a contradiction. We conclude that both $\bar{\mathcal{C}}^i$ and $\bar{\mathcal{D}}^i$ are best responses to $\bar{\mathcal{C}}^j$ and $\bar{\mathcal{D}}^j$.

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

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

This Nash equilibrium can be made into a sequential equilibrium by defining the players' actions appropriately after off-equilibrium path private histories.

References

[1] Abreu, D., D. Pearce, and E. Stacchetti (1990). "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, **58**, 1041–1063.

[2] Alon, N. and J. H. Spencer (2008). *The probabilistic method*, 3rd. ed. Wiley-Interscience, Hoboken, New Jersey.

[3] Aoyagi M. (2002). "Collusion in Dynamic Bertrand Oligopoly with Correlated Private Signals and Communication," *Journal of Economic Theory*, **102**, 229–248.

[4] Ben-Porath, E. and M. Kahneman (1996). "Communication in Repeated Games with Private Monitoring," *Journal of Economic Theory*, **70**, 281–297.

[5] Bhaskar, V., and I. Obara (2002). "Belief-Based Equilibria in the Repeated Prisoners' Dilemma with Private Monitoring," *Journal of Economic Theory*, **102**, 40–69.

[6] Compte, O. (1998). "Communication in Repeated Games with Imperfect Private Monitoring," *Econometrica*, **66**, 597–626.

[7] Ely, J. and J. Välimäki (2002). "A Robust Folk Theorem for the Prisoner's Dilemma," *Journal of Economic Theory*, **102**, 84–105.

[8] Ely, J., J. Hörner and W. Olszewski (2005). "Belief-free Equilibria in Repeated Games," *Econometrica*, **73**, 377–415.

[9] Fudenberg D. and D. Levine (1991). "An Approximate Folk Theorem with Imperfect Private Information," *Journal of Economic Theory*, **54**, 26–47.

[10] Fudenberg, D., D. Levine, and E. Maskin (1994). "The Folk Theorem with Imperfect Public Information," *Econometrica*, **62**, 997–1040.

[11] Fudenberg D. and E. Maskin (1986). "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information," *Econometrica*, **54**, 533–554.

[12] Gossner, O. and J. Hörner (2010). "When is the lowest equilibrium payoff in a repeated game equal to the minmax payoff?" *Journal of Economic Theory*, **145**, 63–84.

[13] Hörner, J. and W. Olszewski (2006). "The Folk Theorem for Games with Private Almost-Perfect Monitoring," *Econometrica*, **74**, 1499–1544.

[14] Hörner, J. and W. Olszewski (2009). "How Robust is the Folk Theorem with Imperfect Public Monitoring?" *Quarterly Journal of Economics*, **124**, 1773–1814.

[15] Kandori, M. (2002). "Introduction to Repeated Games with Private Monitoring, *Journal of Economic Theory*, **102**, 1–15.

[16] Lehrer, E. (1990). "Nash Equilibria of n-player Repeated Games with Semi-Standard Information," *International Journal of Game Theory*, **19**, 191–217.

[17] Mailath, G. J., and S. Morris (2002). "Repeated Games with Almost-Public Monitoring," *Journal of Economic Theory*, **102**, 189–228.

[18] Mailath, G. and L. Samuelson (2006). *Repeated Games and Reputations: Long-Run Relationships*. Oxford University Press, New York, NY.

[19] Matsushima, H. (1991). "On the theory of repeated games with private information : Part I: anti-folk theorem without communication," *Economics Letters*, **35**, 253–256.

[20] Matsushima, H. (2004): "Repeated Games with Private Monitoring: Two Players," *Econometrica*, **72**, 823–852.

[21] Obara, I. (2009). "Folk Theorem with Communication," *Journal of Economic Theory*, **144**, 120–134.

[22] Piccione, M. (2002). "The Repeated Prisoner's Dilemma with Imperfect Private Monitoring," *Journal of Economic Theory*, **102**, 70–83.

[23] Radner R. (1986). "Repeated Partnership Games with Imperfect Monitoring and No Discounting," *Review of Economic Studies*, **53**, 43–58.

[24] Sekiguchi, T. (1997). "Efficiency in Repeated Prisoner's Dilemma with Private Monitoring," *Journal of Economic Theory*, **76**, 345–361.

[25] Stigler, G. (1964). "A Theory of Oligopoly," *Journal of Political Economy*, **72**, 44–61.

[26] Sugaya, T. (2010). "Belief-Free Review-Strategy Equilibrium without Conditional Independence," mimeo, Princeton University.