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

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

December 3, 2007

Abstract

We prove that there exists 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 & 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 of the prisoner’s dilemma 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

---

∗Stanford GSB, contact: fong.kyna at gsb.stanford.edu.
†PSE, Paris and MEDS/KSM.
‡KSM/MEDS and Yale.
§UC Berkeley, contact: sannikov at econ.berkeley.edu.
some chance that a defecting player observes a signal that is sufficiently more likely when
his 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 re-
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 twoplayer prisoner’s dilemma (Sekiguchi (1997), Bhaskar & Obara (2002), Piccione (2002), Ely & Välimäki (2002), a result later extended to all finite games (Hörner & Olszewski 2006a). Further, the same result holds under some assumptions when the monitoring structure approaches imperfect public monitoring (Hörner & Olszewski 2006b), a result that generalizes earlier findings (Mailath & 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’s (2002) 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 (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 (1985), this can be
achieved by dividing the infinite horizon into review phases, or rounds, of length $T$ (see also
Abreu, Milgrom and Pearce (1991); 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 (Ely,
Hörner & 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 cooperate,
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 $\rho n$ 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 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 (Ely et al. 2005).

Figure 1

Some of the difficulties we encounter are 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), Fudenberg and Levine (2007) and Obara (2007). 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. That is, 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 Setup and Notation.

We consider an infinitely repeated Prisoner’s Dilemma with private monitoring and no communication. Each player \(i = 1, 2\) chooses an action \(a_i^t \in \{D_i, C_i\}\) 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 \( y_i^t \) from a finite set \( Y^i \) of \( N \geq 2 \) elements.

For each action profile, every profile of private signals realizes with a strictly positive joint probability, which we denote by \( \pi(y^i y^j | a^i a^j) \).\(^1\) We assume that the matrix of joint probabilities \( \pi(y^i y^j | a^i a^j) \) has maximal rank (i.e. rank \( 2(N-1) \)). 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 \( y^j \).

Denote by \( g^i(a^i a^j) \) the expected stage-game payoff of player \( i \).\(^2\) The stage-game payoffs are those of a Prisoner’s Dilemma, i.e.

\[
g^i(D^i C^j) > g^i(C^i C^j) > g^i(D^i D^j) > g^i(C^i D^j).
\]

Players discount future payoffs at a common rate \( \delta \). So, player \( i \)’s expected payoff in the repeated game is

\[
E \left[ \sum_{t=1}^{\infty} \delta^{t-1} g^i(a^i_t a^j_t) \right].
\] (1)

A \( t \)-period private history of player \( i \) is a sequence of private actions and private signals, denoted by \( h^i_t = (a^i_1, y^i_1, a^i_2, y^i_2, \ldots a^i_t, y^i_t) \). Let \( H^i_t \) be the set of \( t \)-period private histories for player \( i \). A private strategy of player \( i \) is a function \( s^i : \bigcup_{t=1}^{\infty} H^i_t \to [0,1] \) that gives the probability with which player \( i \) plays \( C^i \) after each private history \( h^i_t \in H^i_t \) for all \( t \geq 0 \).

A profile of private strategies \((s^1, s^2)\) 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 \((s^1, s^2)\) is a sequential equilibrium of the repeated game if each player’s strategy maximizes the conditional expectation of his payoff

\[
E \left[ \sum_{t=1}^{\infty} \delta^{t-1} g^i(a^i_t a^j_t) \mid h^i_t \right],
\]

for any private history \( h^i_t \in H^i_t \). 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 sequential equilibrium by modifying the players’ actions after zero-probability private histories.

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:

\(^1\)Whenever we refer to players \( i \) and \( j \), we assume \( i, j \in \{1, 2\} \) and \( i \neq j \).

\(^2\)As usual, the actual payoff realization of player \( i \) in a given period depends on his action and his private signal.
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 \( M^i \) be the matrix of conditional probabilities \((\pi(y^j|C^iC^j, y^i))_{y^i,y^j}\), which 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 \( \lambda^i \) that his opponent assigns is higher than when player \( j \) plays \( D^j \).

Assumption 2 (Positively correlated scores). There exists an eigenvector \( \lambda^1 \) of \( M^1M^2 \) such that, if we let \( \lambda^2 = M^2\lambda^1 \), the expectation of \( \lambda^i \) 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 when \( N = 2 \) 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 at least as 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.
\[
\pi(1^i|C^iC^j, 1^j) \geq \pi(1^i|C^iC^j).
\]

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. \(^3\)

We can now state our main result.

**Theorem 1.** Under Assumptions 1 and 2, there exists a sequential equilibrium of the repeated game that approaches efficiency as the discount factor goes to 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.

\(^3\)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.
For each $\delta$, 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. 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. At the same time, it is important that $\delta^T$ converge to 1 as $\delta \to 1$ 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: $C^i$ and $D^i$. Strategy $C^i$ involves cooperation in almost all $T$ periods, and strategy $D^i$ consists of defection in all $T$ periods. Player $i$ creates incentives for his opponent through a transition rule that determines which strategy, $D^j$ or $C^i$, 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 $W^j_\delta$ when $C^j$ is played and by $W^j_D$ when $D^j$ is played, creates incentives across review phases.\(^5\) Denote by $[G^j_D, G^j_C]$ the range of rewards and punishments that can be assigned to player $j$ by player $i$’s mixing between $D^i$ and $C^i$.

To ensure that both $C^i$ and $D^i$ 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 $D^j$ for sure, player $i$ is indifferent between any sequence of actions.

2. When the opponent is playing $C^j$ for sure, both strategies $C^i$ and $D^i$ are optimal.

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


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

\(^6\)So our construction is belief-free across review phases, as in Horner and Olszewski (2006a) and Matsushima (2004).
Players attain payoffs arbitrarily close to full cooperation as \( \delta \to 1 \) because with probability near 1 the players again play \( \hat{C}^1 \) and \( C^2 \) at the end of a review phase in which they have just played \( C^1 \) and \( C^2 \). 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, \( D^j \). This strategy is simple: it consists of defection in all \( T \) periods. When player \( j \) follows \( D^j \), he rewards player \( i \) a constant amount \( K^i_D \) for each signal \( \hat{y}^j \) received (as identified in Assumption 1). The value of \( K^i_D \) is chosen to make player \( i \) just indifferent between cooperating and defecting when player \( j \) defects. We introduce the following definition.

**Definition 1.** A linear test for the signal \( y^j \) rewards a constant \( K^i_D(y^j) \geq 0 \) for each signal received that is equal to \( y^j \), where \( K^i_D(y^j) \) is defined implicitly by

\[
g^i(C^iD^j) + \pi(y^j|D^jC^i)K^i_D(y^j) = g^i(D^iD^j) + \pi(y^j|D^jD^i)K^i_D(y^j).
\]

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

\[
W^j_D(h^j_T) = G_D + K^i_D \sum_{t=1}^{T} \delta^{t-T} 1(y^j_t = \hat{y}^j),
\]

where \( h^j_T \) is player \( j \)'s private history during the review phase, and \( 1(y^j_t = \hat{y}^j) \) is the indicator function that \( y^j_t \), the time-\( t \) signal in history \( h^j_T \), is equal to \( \hat{y}^j \). In the following, we refer to \( W^j_D \) as the linear test.

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 \( \hat{D}^j \), player \( i \)'s expected payoff \( G^i_D \) is bounded strictly above \( g^i(D^iD^j)/(1 - \delta) \). We need Assumption 1 to ensure that this construction still leaves room to punish player \( i \), i.e. that \( G^i_D < g^i(C^iC^j)/(1 - \delta) \).

Let us turn our attention now to the cooperative strategy \( C^j \) and the reward function \( W^j_C \). A natural way for player \( j \) to reward player \( i \) is to keep a “score” of \( i \)'s performance that depends on the signals \( j \) receives. This score then determines player \( i \)'s reward from the review phase. To construct that score, in Lemma 1 we identify a collection of weights \( \{\lambda(y^j), y^j \in Y^j\} \) that satisfy some convenient properties and whose existence is guaranteed by Assumption 2. One can think of these weights as the expected score increase player \( j \) assigns as a function of the signal he receives.
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

\[
E_{C \mid i \in C} \lambda^j(y^j) > E_{D \mid i \in C} \lambda^j(y^j)
\]

and

\[
E_{C \mid i \in C} [\lambda^j(y^j)|y^i] - \bar{\lambda}^j = \beta (\lambda^i(y^i) - \bar{\lambda}^i).
\]

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

Proof. See Appendix.

Condition (4) 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 (5) is one of the most important insights in our equilibrium construction: 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 \( \lambda^j(y^j) \). Thus, we reduce the dimensionality of each player’s inference problem to a single dimension.

Loosely speaking, we specify the reward function \( W_C^j \) to be a function of the weights \( \lambda^j(y^j) \). 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 \( h_i \), \( W_C^j \) 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. Hence, efficiency will be destroyed through excessive transitions to \( D^j \) at the end of the review phase. So it is impossible for \( W_C^j \) to induce cooperation in all periods while maintaining efficiency. See Figure 3.

As a solution, we “shift” the reward function up so that in expectation, for some \( 0 < k < 1 \) only \( O(T^k) \) in value is destroyed through transitions to \( D^j \). That loss in efficiency of \( O(T^k) \) per \( T \) periods is negligible as \( T \) becomes large. See Figure 3.

There is still the issue of incentives, however. In fact, we have now created a potential problem since when player \( j \)'s assessment of player \( i \)'s performance is exceptionally high, player \( i \)'s incentives to cooperate break down because the reward function \( W_C^j \) becomes flat. To solve that problem, our strategy is to tightly bound the event in which player \( i \)'s incentives break down, and show that the probability of this event occurring becomes very small as \( T \) becomes large.
To that end, we distinguish two events $\Phi^j_t$ and $\neg \Phi^j_t$. We label the event $\Phi^j_t$ such that if $h^i_t \in \Phi^j_t$, then $W^j_{C}$ gives player $i$ incentives to cooperate. In particular, $\Phi^j_t$ corresponds to the set of histories in which the scores $\lambda^j(y^j)$ that player $j$ has computed about player $i$’s performance are not too large. See Figure 3.

Lemma 2 guarantees that we can choose $\Phi^j_t$ to satisfy the statistical properties we need.

**Lemma 2.** For each $1 \leq t \leq T$, there exists an event $\Phi^j_t \subset H^j_t$ such that for some $\alpha > 0$,

1. If both players are cooperating in every period of the review phase, then
   \[ Pr[\neg \Phi^j_T] \leq e^{-\alpha T}, \]  
   \[ (6) \]
2. If player $j$ is cooperating in every period of the review phase, then
   \[ Pr[\neg \Phi^j_T | h^i_t \in \Phi^j_t] \leq e^{-\alpha T}. \]  
   \[ (7) \]

Property (6) tells us that if both players are cooperating so long as $\Phi^j_T$ and $\Phi^j_T$ hold, then each will almost always cooperate over the entire review phase. This guarantees that the loss of efficiency due to defections within a review phase is small. Meanwhile, Property (7) tells us that so long as the score player $i$ is computing is not too high (i.e. $\Phi^j_t$ is true), then the probability that his incentives to cooperate break down because player $j$ leaves $\Phi^j_T$ becomes arbitrarily small. Thus, we can ensure that conditional on player $i$’s own score not being too high, player $i$ has incentives to cooperate. The intuition for why this is true is related to condition (5), which tells us that even when the score player $i$ is computing himself is extreme, his beliefs about the score player $j$ is computing are not quite as extreme.

There will arise the concern that player $i$’s inferences about the relative likelihoods of $\Phi^j_t$ and $\neg \Phi^j_t$ may affect player $i$’s incentives to cooperate. To mitigate this problem, we specify the reward function in the event $\neg \Phi^j_t$ such that in choosing what strategy to play, player $i$ can always condition on the event $\Phi^j_t$ being true. In particular, we alter the reward function $W^j_{C}$ such that, as soon as $\neg \Phi^j_t$ holds for any $t$, for the remainder of the review phase player $j$ will assign rewards that make player $i$ just indifferent between all actions. As a result, the possibility of the event $\neg \Phi^j_t$ becomes irrelevant for player $i$’s incentives.

The alternative reward function in the event $\neg \Phi^j_t$ for some $t \leq T$ is defined in the same spirit as the linear test above. The key difference is that instead of ensuring that all actions
Table:

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Period 1</th>
<th>Period 2</th>
<th>...</th>
<th>Period t</th>
<th>...</th>
<th>Period T</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C^1$</td>
<td>$C$</td>
<td>$\frac{1}{2} C + \frac{1}{2} D$</td>
<td>$C$</td>
<td>mostly $C$</td>
<td>...</td>
<td>mostly $C$</td>
</tr>
<tr>
<td>$C^2$</td>
<td>$\frac{1}{2} C + \frac{1}{2} D$</td>
<td>$C$</td>
<td>...</td>
<td>mostly $C$</td>
<td>...</td>
<td>mostly $C$</td>
</tr>
</tbody>
</table>

Figure 1: Strategies $C^1$ and $C^2$.

are a best reply when the opponent defects, this test ensures that all actions are a best reply no matter what action the opponent plays. We call this test the bi-linear test and define it as follows.

**Definition 2.** 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 $y^j$.

Define $\tau \leq T$ as the stopping time in the review phase such that $h^j_t \in \Phi^j_t$ for all $t \leq \tau$ and $h^j_{\tau+1} \not\in \Phi^j_{\tau+1}$. We alter $W^j_C$ such that from period $\tau + 1$ onwards, the reward assigned by player $j$ is a bi-linear test. More formally, taking discounting into account, the continuation reward assigned by player $j$ for periods $\tau + 1$ through $T$, as a function of his private history $h^j_T$, will be

$$-\delta^{-T} \sum_{t=\tau+1}^{T} \delta^t K(a^j_t, y^j_t)$$

where $\delta^{-T}$ discounts the reward back to time 0 of the review phase. This amendment to $W^j_C$ allows us to condition only on the event $\Phi^j_t$ when considering player $i$’s incentives.

Recall that when player $j$ is playing $C^j$ and assigning the reward function $W^j_C$, both $C^i$ and $D^i$ must be best responses for player $i$. In particular, we need to guarantee that once player $i$ starts cooperating (respectively defecting), he will have incentives to continue doing so for the remainder of the review phase. We achieve that property through an initial “lock-in” stage.

We distinguish the initial two periods of the review phase to play a special role in $C^j$, and separate them from the remainder of the review phase, which follows play as described above. More specifically, play in the first two periods will determine the reward function faced, and hence the incentives faced, by the players in the remaining periods (so long as the relevant event $\Phi^j_t$ is satisfied). The construction of $C^j$ is illustrated in Figure 1.

---

\footnote{Let $\tau = T$ if $h^j_t \in \Phi^j_t$ for all $t \leq T$.}
In period 1, player 1 “announces” whether he intends to play $C^1$ or $D^1$ during the review phase by playing $C$ or $D$ respectively in period 1. At the same time, if player 2 is playing $C^2$, he mixes between $C$ and $D$.

Player 2 then chooses the “slope” of $W^2_C$ as a function of his private action and private signal in period 1. If player 1 plays $C$ in period 1, then in expectation player 2 will choose the slope of $W^2_C$ to be large enough that player 1 wants to continue to cooperating. On the other hand, if player 1 plays $D$ in period 1, then player 1 will expect player 2 to choose the slope of $W^2_C$ such that player 1 wants to continue defecting. In period 2, the players switch roles. Player 2 “announces” his strategy for the review phase, while player 1 who is following strategy $C^1$ mixes between $C$ and $D$ and determines the slope of $W^1_C$ as a function of his private action and private signal in period 2. We add a constant to $W^j_i$ to guarantee that player $i$ is indifferent between playing $C$ and $D$ in period $i$.

The maximal rank assumption of the joint probability matrix is convenient in determining the slopes that the players are assigned in their “announcement” period, i.e. period $i$ for player $i$. In particular, maximal rank allows us to specify slopes such that player $i$’s incentives (before the stopping time $\tau$) do not depend at all on the signal he receives in period $i$. The only thing that affects his incentives is the action he takes in that period.

To summarize, we define the strategies $C^j$ for $j = 1, 2$ as belonging to the following class of strategies.

**Definition 3.** Given the events $\{\Phi^j_t \ \forall t \leq T\}$, a $T$-period strategy of player $j$ is from the class $Z^j$ if it satisfies the following conditions:

1. In period $j$, player $j$ plays $C$.

2. In period $3 - j$, 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 $\Phi^j_s$ holds for all $s \leq t$.

The reader may wonder how the possibility of the stopping time $\tau$ and the switch to an alternative reward function affects player $i$’s incentives to defect in every period if he follows the strategy $D^i$. When player $i$ chooses his action in period $t$, his decision is guided by three considerations: (1) the reward function before the stopping time $\tau$, which we argue above gives him incentives to defect always, (2) the reward function after the stopping time $\tau$, which is the bi-linear test that makes him indifferent between cooperating and defecting in each period, and (3) the extent to which player $i$’s action the occurrence of the

---

8Recall that strategy $D^2$ specifies defection in every period.

9Recall that the expectation of $W^2_C$ is a function of the scores $\lambda^2$ that player 2 computes in the remaining periods as a function of his received signals. Hence the slope of $W^2_C$ loosely reflects the marginal reward assigned by player 2 for higher scores.
stopping time \( \tau \). To guarantee that playing \( D^i \) is a best response, we need to ensure that it not only maximizes player \( i \)'s expected payoff (relative to playing \( C^i \)), but also minimizes the probability of the stopping time occurring. As such, the scores we use are in fact a modification of the “scores” \( \lambda^j(y^j) \) that we have introduced previously. Details for the interested reader are found in Section 4.

Notice that in our discussion we have not relied on a specific strategy \( C^j \) from the class \( Z^j \). In particular, we have not specified player \( j \)'s strategy after the stopping time \( \tau \) when \( \neg \Phi^j_t \) first occurs. This portion of the strategy will be determined by a fixed-point argument.

We conclude then that under our construction, both playing \( D^i \) and following some strategy \( C^i \) from class \( Z^i \) are best responses for player \( i \) to any strategy \( C^j \) from class \( Z^j \) together with an appropriately defined reward function \( W^j_C \). A fixed point argument shows that there exists a pair of strategies \( C^1 \in Z_1 \) and \( C^2 \in Z_2 \) and a pair of reward functions \( W^1_C \) and \( W^2_C \) such that both strategies are mutual best responses.

4 Formal Proofs.

In this section, we formally construct sequential equilibria of the repeated Prisoner’s Dilemma that approach efficiency as \( \delta \to 1 \). 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 \( C^i \) or \( D^i \). Both of these strategies are optimal independently of the beliefs about the strategy of the opponent. While \( D^i \) involves defection in every period, \( C^i \in Z^i \) involves cooperation in every period with probability near 1. A player’s strategy choice creates equilibrium incentives for his opponent.

Proposition 1 shows that, to construct efficient equilibria along these lines, we just need to find appropriate \( T \)-period strategies and reward functions.

**Proposition 1.** Suppose that for \( i = 1, 2 \) and some \( T > 0 \), there are \( T \)-period strategies \( C^i \) and \( D^i \) and reward functions \( W^i_C : H_T \to [G^i_D, G^i_C] \) and \( W^i_D : H_T \to [G^i_D, G^i_C] \) that satisfy the following conditions.

First, when player \( j = 1, 2 \) is following strategy \( C^j \), then

\[
G^i_C = \max E \left[ \sum_{s=0}^{T-1} \delta^s g^i(a^i_s a^j_s) + \delta^T W^i_C(h^i_T) \right],
\]

(9)

where the maximum, taken over all \( T \)-period strategies of player \( i \), is achieved by both \( C^i \) and \( D^i \).
Second, when player $j$ is following strategy $D^j$, then

$$G^i_D = \max E \left[ \sum_{s=0}^{T-1} \delta^s g^i(a^i_s a^j_s) + \delta^T W^j_D(h_T^j) \right],$$

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

Then any pair of payoffs $(w_1, w_2) \in [G^1_D, G^1_C] \times [G^2_D, G^2_C]$ is achievable by a sequential Equilibrium of an infinitely repeated game with discount factor $\delta$.\(^{10}\)

**Proof.** See Appendix. \( \square \)

To construct equilibria using Proposition 1, for $i = 1, 2$, strategies $C^i$ and $D^i$ together with reward functions $W^i_C$ and $W^i_D$ have to satisfy two sets of constraints

1. **incentive constraints**, i.e. both $C^i$ and $D^i$ have to maximize (9) and (10) and
2. **feasibility constraints**, i.e. the reward functions must take values in the ranges $[G^i_D, G^i_C]$, which are defined by (9) and (10).

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

$$(1 - \delta)G^i_C \to g^i(C^i C^j) \quad \text{as} \quad \delta \to 1. \quad (11)$$

**Strategies $C^j$ and reward functions $W^j_D$.** In Section 2 we introduced the linear reward function:

$$W^j_D(h_T^j) = G^i_D + \delta^{-T} K^i_D \sum_{t=1}^{T} 1(y_t^i(h_T^j) = \hat{y}^j), \quad \text{where} \quad (1-\delta)G^i_D = g^i(D^i D^j) + K^i_D \pi(\hat{y}^j | D^i D^j),$$

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

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

(ii) we have

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

\(^{10}\)When we let $\delta \to 1$, for simplicity our notation suppresses the dependence of the strategies, reward functions and bounds $G^i_D$ and $G^i_C$ on $\delta$. 

14
Proof. For any $T$-period strategy of player $i$, his expected total payoff is

$$E \left[ \sum_{s=0}^{T-1} \delta^s g^i(a_s^i D^i) + \delta^T W_D^i(h_T^i) \right] = E \left[ \sum_{s=0}^{T-1} \delta^s (g^i(a_s^i D^i) + K_D^i 1(y_s^i = \hat{y}^i)) \right] + \delta^T G_D^i$$

$$= \sum_{s=0}^{T-1} \delta^s (g^i(D^i D^i) + K_D^i \pi(\hat{y}^i|D^i D^i)) + \delta^T G_D^i = G_D^i,$$

since the expectation of $g^i(a_s^i D^i) + K_D^i 1(y_s^i = \hat{y}^i)$ 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

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

Now, from Assumption 1, we have the following:

$$\pi(\hat{y}^i|D^i D^i)(g^i(C^i C^i) - g^i(C^i D^i)) < \pi(\hat{y}^i|D^i C^i)(g^i(C^i C^i) - g^i(D^i D^i)) \Rightarrow$$

$$g^i(D^i D^i)\pi(\hat{y}^i|D^i C^i) - g^i(C^i D^i)\pi(\hat{y}^i|D^i D^i) < g^i(C^i C^i)(\pi(\hat{y}^i|D^i C^i) - \pi(\hat{y}^i|D^i D^i)) \Rightarrow$$

$$G_D^i = \frac{g^i(D^i D^i) + K_D^i \pi(\hat{y}^i|D^i D^i)}{1 - \delta} = \frac{g^i(D^i D^i)\pi(\hat{y}^i|D^i C^i) - g^i(C^i D^i)\pi(\hat{y}^i|D^i D^i)}{(1 - \delta)(\pi(\hat{y}^i|D^i C^i) - \pi(\hat{y}^i|D^i D^i))} < \frac{g^i(C^i C^i)}{1 - \delta}. \quad \Box$$

**Strategies $C^j$ and reward functions $W_C^j$.** To complete the proof that equilibrium efficiency is attainable in the limit as $\delta \to 1$, for each $\delta$ and $i = 1, 2$ we need to construct strategies $C^j \in Z^j$ and reward functions $W_C^j$ that satisfy (9), the feasibility constraints and (11). This task is significantly more challenging than the construction of $D^j$. 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.\(^{11}\)

We first define need to define the event $\Phi_i^t$. Without loss of generality, we shall assume that the weights $\lambda^j : Y^j \to \mathbb{R}$ are in $(0, 1)$, as we can always add a constant to all of them and divide them by a positive constant. Given $h_t^i$, let

$$\Lambda_t^i = \sum_{\tau=t}^{T} \lambda^j \left( y_{\tau}^i \right).$$

\(^{11}\)Such 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.
and define

\[ L_i^t = \sum_{\tau=3}^{t} l_i^\tau, \]

where \( \{l_i^\tau : \tau = 3, \ldots, t\} \) are independent Bernoulli random variables with mean \( \lambda_i(y_j^\tau) \). (Observe that the summations start at \( \tau = 3 \): as explained in Section 3, the first two periods of a round play a special role in our construction.) Let \( \bar{\Lambda}_i^t \) denote the expected value of \( \Lambda_i^t \), conditional on both players cooperating in all periods.

**Definition 4.** For all \( t = 1, \ldots, T \), let

\[ \Phi_i^n : = \left\{ h_i^t \in H_i^t : L_i^{\tau} \leq \bar{\Lambda}_i^T + T^{2/3} \text{ for all } \tau \leq t \right\}, \]

\[ \Phi_i^m : = \left\{ h_i^t \in H_i^t : \left| L_i^{\tau} - \bar{\Lambda}_i^{\tau} \right| \leq T^{7/12} \text{ for all } \tau \leq t \right\}, \]

\[ \Phi_i^t : = \Phi_i^n \cap \Phi_i^m. \]

The definitions of scores introduced above deserves a few comments. We have introduced both a virtual score \( \Lambda_i^t \in \mathbb{R}_+ \) that follows directly from the values of \( \lambda_i \) obtained from Assumption 2, and an actual score, \( L_i^t \in \mathbb{N} \) that each player computes independently, based on \( \Lambda_i^t \). The reason why we introduced the actual score, in addition to the virtual one, is the following: while the expectations of \( \lambda_i \) is higher under \( C_iC_j \) than \( C_iD_j \), the distribution under \( C_iC_j \) need not first-order stochastically dominate the distribution under \( C_iD_j \). As incentives do not only depend on the average reward under \( \Phi_i^t \), but also on the probability of the event \( \neg \Phi_i^t \), it is useful for our purposes to have stochastic dominance.

We shall use the following strengthening of Lemma 2. Recall that a strategy of player \( j \) is from class \( Z^j \) if it plays \( C \) in the \( j \)th period and randomizes equally across actions in period \( 3 - j \). cooperation is always prescribed in period \( t \geq 3 \) in the event \( \Phi_i^t \).

**Lemma 3.** There exists \( \alpha > 0 \) such that, for all \( t = 1, \ldots, T \):

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

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

(ii) for all \( h_i^t \in \Phi_i^t \), conditional on both players using a strategy from \( Z^i \),

\[ \Pr \left[ \neg \Phi_T^j | h_i^t \right] < e^{-T^\alpha}, \]

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

\[ \Pr \left[ \Phi_T^j \cap \neg \Phi_T^m | h_i^t, \Phi_i^t \right] < T^{-\alpha}. \]
Before defining the reward function, we must introduce some further notation. Let

\[ M = \begin{bmatrix}
  \pi(C^j y^j_1 | C^i y^i_1) & \cdots & \pi(C^j y^j_N | C^i y^i_1) & \pi(D^j y^j_1 | C^i y^i_1) & \cdots & \pi(D^j y^j_N | C^i y^i_1) \\
  \vdots & & \vdots & \vdots & & \vdots \\
  \pi(C^j y^j_1 | C^i y^i_N) & \cdots & \pi(C^j y^j_N | C^i y^i_N) & \pi(D^j y^j_1 | C^i y^i_N) & \cdots & \pi(D^j y^j_N | C^i y^i_N) \\
  \pi(C^j y^j_1 | D^j y^j_1) & \cdots & \pi(C^j y^j_N | D^j y^j_1) & \pi(D^j y^j_1 | D^j y^j_1) & \cdots & \pi(D^j y^j_N | D^j y^j_1) \\
  \vdots & & \vdots & \vdots & & \vdots \\
  \pi(C^j y^j_1 | D^j y^j_N) & \cdots & \pi(C^j y^j_N | D^j y^j_N) & \pi(D^j y^j_1 | D^j y^j_N) & \cdots & \pi(D^j y^j_N | D^j y^j_N) 
\end{bmatrix} \]

Lemma 4. For any reals \( b_C, b_D \), there exists \( \tilde{\varepsilon} > 0 \) and \( M > 0 \) such that, for all \( \varepsilon \in (0, \tilde{\varepsilon}) \), if \( b_D \in (b_C - \varepsilon, b_D) \), then the equation

\[ M \begin{bmatrix} b(C^j y^j_1) \\ \vdots \\ b(D^j y^j_N) \end{bmatrix} = \begin{bmatrix} b_C \\ \vdots \\ b_D \end{bmatrix}, \tag{12} \]

has a solution in \( b(A^j y^j) \in (b_C - M \varepsilon, b_D + M \varepsilon) \) for \( A^j \in \{D^j, C^j\} \) and \( y^j \in Y^j \).

Proof. See Appendix.

Given \( \varepsilon \in (0, \tilde{\varepsilon}/2) \), let

\[ b_C := b_0 + \varepsilon, b_D := b_0 - \varepsilon, \]

where \( b_0 := \frac{g^j(D^j C^j) - g^j(C^i C^j)}{\sum_{y^j} (\pi(y^j | C^j C^i) - \pi(y^j | C^j D^i)) \lambda(y^j)} \),

and let \( b(A^j y^j) \) denote the corresponding solution of equation (12) whose existence is shown in the Lemma above. Also, define \( K^i_C \) by

\[ g^j(C^i C^j) + \pi(y^j | C^j D^i) K^i_C = g^j(D^i C^j) + \pi(y^j | C^j D^i) K^i_C, \]

so that \( i \) is indifferent between both actions if \( j \) plays \( C^j \), but \( i \) gets an additional \( K^i_C \) in the event in which \( y^j = \hat{y}^j \).

We define the reward function as follows. Consider player 1 first. Let

\[ W^2_C(h_T^2) = c^2(y^2_1) + \delta^{2-T} K^i_C 1(y^2_2 = \hat{y}_2) + \sum_{t=3}^{T} b(a^2_i y^2_i) 1(L_i^2 = 1) - \delta^{2-T} \sum_{t=\tau^2+1}^{T} K(a^2_i y^2_i), \]
where the function $K$ comes from the definition of the bi-linear test, $\tau^2 = \{ \inf t : h_{t+1}^2 \in \Phi_{t+1}^2 \}$ is the random stopping time at which player 2’s history first leaves $\Phi_t^2$, and the function $c^2$ is defined as follows. Observe that player 1’s first action and signal $(h_1^1 = (a_1^1, y_1^1))$, and given player 2’s strategy, player 1’s optimal continuation strategy in the $T$-stage repeated game is independent of the specification of $c^2$ (since the latter only depends on $y_1^1$). We pick $c^2$ such that player 1 is indifferent between playing $C$ and $D$ in the initial period, given that player 2 randomizes equally between both actions in that period. Observe that, because all values of $b(a_1^1 y_1^1)$ are within $M\varepsilon$ of each other, if the event $h_{t}^2 \in \Phi_{t}^2$ is arbitrarily unlikely under the optimal strategy, the range of values of $c^2$ is of the order $\varepsilon T$.

Consider player 2 now. Then:

$$W_C^1(h_T^1) = \delta^{1-T} K_C^3 1 (y_1^1 = \hat{y}^1) + \sum_{t=3}^{\tau^1} b(a_2^t y_2^t) 1 (L_t^1 = 1) - \delta^{1-T} \sum_{t=\tau^1+1}^{T} K (a_t^1 y_t^1),$$

where $\tau^1$ is defined as before, and $c^1$ is now defined so as to ensure that player 2 is indifferent between both actions in the second period, given his optimal continuation strategy, for a given strategy of player 1 (which in particular randomizes equally between both actions in the second period). Here as well, the range of $c^1$ is of order $\varepsilon T$, provided the event $h_t^1 \in \Phi_t^1$ is arbitrarily unlikely under the optimal continuation strategy.

We may now state:

**Proposition 3.** Suppose that $T = O((1 - \delta)^{-1/2})$, and fix $\varepsilon > 0$. Fix a strategy of player $j$ in $Z^j$. If player $j$ assigns to player $i$ the corresponding reward function $W_C^j$, then:

(i) for $T$ large enough, both $D^i$ and some strategy in $Z^i$ are optimal.

(ii) we have

$$\lim_{T} (1 - \delta) G_C^i > g^i(C^iC^j) - \varepsilon.$$

**Proof:** Throughout, fix a strategy $s^j$ in $Z^j$.

**Some strategy in $Z^i$ is optimal.** The indifference in periods 1 and 2 follows from the definition of $W_C^j$, 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 $h_t^i \in \Phi_t^i$, for all periods $t \geq 3$. Let us define $W_C^{ij}$ as

$$W_C^{ij}(h_T^j) = \delta^{j-T} K_C^{3-j} 1 (y_j^j = \hat{y}^j) + \sum_{t=3}^{T} b(a_j^t y_j^t) 1 (L_t^j = 1),$$

18
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 $\Phi'_j$. Because of the definition of $K$, it follows that the payoff of any given strategy $s^i$ against $s^j$ and $W^j_C (h^j_T)$ is weakly higher than against $s^j$ and $W^j_C (h^j_T)$.

Observe now that, because $\text{Pr} [\Phi'_j|h^j_i] > 0$, because $h^j_i$ does not change the incentives of player $i$, and the expected value of $b(A^j_i y^j_i)$ conditional on $C$ is $b_C$ (independently of $i$'s signal in period $i$) a continuation strategy $s^i|\tau^j_i$ is optimal against $s^j$ and $W^j_C (h^j_T)$ is optimal if it is optimal against

$$W^j_C (h^j_T) = \sum_{t=3}^{T} b_C (L^j_t = 1),$$

and since $b_C > b_0$, it follows that the unique continuation strategy that is optimal against $s^j$ and $W^j_C (h^j_T)$ consists in playing $C$ after history $h^j_i$. Furthermore, the gain from playing $C$ rather than $D$ is bounded away from $0$, because $b_C > b_D$, independently of $T$. Observe now that, because $\text{Pr} [\Phi'_j|h^j_i] < e^{-T\alpha}$ for $h^j_i \in \Phi'_i$, the continuation payoff from playing $s^j$ (i.e., playing $C$ always) against $s^j$ and $W^j_C (h^j_T)$, after a history $h^j_i \in \Phi'_i$, tends to the continuation payoff against $s^j$ and $W^j_C (h^j_T)$. That is, for $T$ large enough, playing $C$ against history $h^j_i \in \Phi'_i$ is optimal against $s^j$ and $W^j_C (h^j_T)$.

The strategy $D^j$ in $Z^j$ 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 $h^j_i \in \Phi'_i$, $t \geq 3$. For this case, we define $W^j_C$ as:

$$\hat{W}^j_C (h^j_T) = \sum_{t=3}^{\tilde{t}^1} b_D (L^j_t = 1) - \delta^{t-T} \sum_{t=\tilde{t}^1+1}^{T} K (a^j_i y^j_i),$$

where $\tilde{t}^1 = \{ \text{inf } t : h^j_{t+1} \in \Phi^j_{t+1} \}$. That is, the only differences between $\hat{W}^j_C$ and $W^j_C$ are (i) the coefficient $b_D$ which replaces $b(a^j_{3-j} y^j_{3-j})$ and (ii) the region which triggers the bi-linear test; in the case of $W^j_C$, it is when $h^j_T$ leaves the region $\Phi^j_{t+1}$; in the case of $\hat{W}^j_C$, it is when $h^j_T$ leaves the region $\Phi^j_{t+1}$ - a subset of $\Phi^j_{t+1}$. Observe that replacing $b(a^j_{3-j} y^j_{3-j})$ by $b_D$ does not change the incentives of player $i$, since $b_D$ is the expected value of $b(a^j_{3-j} y^j_{3-j})$, 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 $\hat{W}^j_C$, conditional on the event $h^j_i \cap \Phi^j_i$, for any $h^j_i \in H^j_i$ consists in playing $D$ always: indeed, the distribution of $\tau^j$ conditional on $D$ always is (weakly) first-order stochastically dominated by the distribution of $\tau^j$ conditional on any other continuation strategy. Second, in any period in which $h^j_i \in \Phi^j_i$, and thus $h^j_i \in \Phi^j_i$, the gain from playing $D$ rather than $C$ in the immediate period is bounded away from $0$, because $b_0 > b_C$.

Observe now that, because $\text{Pr} [\Phi^j_i \cap \Phi^j_{t+1} | h^j_i, \Phi^j_i] < T^{-\alpha}$ as long as players have played $D^jC^j$ in all periods $t' = 3, \ldots, t$, the distribution of $\tau^j$ conditional on the event $h^j_i \cap \Phi^j_i$...
(given $s^i$) approaches the distribution of $\tau^i$ (given $s^j$). So the payoff from playing against $s^j$ and $W^j_C$ tends to the payoff against $s^j$ and $\tilde{W}^j_C$ as $T \to \infty$. It follows that playing $D$ is optimal for player $i$ in period $t = 3$, and recursively, for any $t \geq 3$.

**Efficiency.** It remains to show that:

$$(1 - \delta)G^i_C \to g^i(C^iC^j).$$

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

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

□

So, for the class of $j$‘s strategies $Z^j$ and reward functions that satisfy Proposition 3, player $i$‘s best response is either to play $D^i$ or a strategy from class $Z^i$, $C^i$. If for each player $j = 1, 2$, strategy $C^j \in Z^j$ and reward function $W^j_C$ are chosen appropriately so that the opponent is indifferent between $D^i$ and $C^i$, then the players’ strategies and reward functions satisfy the conditions of Proposition 1. Those strategies can then be used to construct an equilibrium with any value pair from the set $[G^1_D, G^1_C] \times [G^2_D, G^2_C]$.

Proposition 4 shows that there exist $T$-period strategies $C^1, C^2$ and functions $W^1_C, W^2_C$ such that, for $i = 1, 2$, both $D^i$ and $C^i$ are optimal in response to $C^j$ when the reward function $W^j_C$ is given by Proposition 3. We verify the existence of such strategies and reward functions via a fixed-point argument.

**Proposition 4.** For all sufficiently large $T$, there are reward functions $W^1_C$ and $T$-period strategies $C^i$ from class $Z^i$ for $i = 1, 2$, such that both $D^1$ and $C^1$ are best responses to $C^2$ with the reward function $W^2_C$, and similarly for $D^2$ and $C^2$.

**Proof.** TO BE COMPLETED □

### 4.1 Proof of Second lemma “Large deviations”

We rely on the following “large deviations”, result. See, e.g., Theorem A.1.16, in “The Probabilistic Method”, by Alon, Spencer, with the contribution of Erdös.\(^\text{12}\)

\(^\text{12}\)see the beautiful ö? :)

20
Lemma 5. Let $y_i, 1 \leq i \leq n,$ be mutually independent with all $E[y_i] = 0$ and all $|y_i| \leq A.$ Set $S = y_1 + \ldots + y_n.$ Then,

$$P[S > a] < e^{-(a/A)^2/2n}$$

Assume that $h^i_t \in \phi^i_t.$ Define $y_s = \lambda_s^i - z\lambda^i_t$ for $s \leq t$ and $y_s = \lambda^i_s$ for $s > t.$ Conditional on $h^i_t,$ the $y_s$ are mutually independent random variables, with expectation 0 and values between $-A$ and $A$ for some $A$ ($A$ does depend on the $\lambda$s, hence on the monitoring structure, only).

From lemma 5, for $s \leq t$ and every $a$

$$P[X^j_s - zX^i_t > a|h^i_t] < e^{-(a/A)^2/2s} \leq e^{-(a/A)^2/2T}$$

and for $s > t,$

$$P[X^j_s - zX^i_t > a|h^i_t] < e^{-(a/A)^2/2s} \leq e^{-(a/A)^2/2T}$$

Let us redefine $\phi^i_t$ as the intersection of events

$$X(h^i_s) \leq T^{2/3}$$

for $s \leq t.$

For every $h^i_t \in \phi^i_t$ and every $s \leq T,$

$$P[X^j_s > T^{2/3}|h^i_t] < e^{-(1-z)T^{1/3}/2A^2}$$

so that

$$P[-\phi^i_T|h^i_t] < Te^{-KT^{1/3}}$$

with $K = \frac{1-\xi}{2A^2}.$

4.2 Proof of Third lemma “Mixing”

Let $M$ denote the matrix on the left hand side of equation 12. $M$ is stochastic, i.e. if $U$ denotes the unit vector, $MU = U.$ Also, $M$ is generically (in the monitoring structure) invertible. Let $V$ be such that $MV$ has its $N$ first components equal to 0, and its $N$ last equal to 1. Then, $b_CU + (b_D - b_C)V$ satisfies equation 12. For $b_D$ sufficiently close to $b_C,$ all coefficients of $b_CU + (b_D - b_C)V$ are strictly positive.
A Appendix.

A.1 Proof of Proposition 1.

Proof. For players $i = 1, 2$, define recursive strategies $\hat{C}^i$ and $\hat{D}^i$ of the infinitely repeated game as follows. Let us divide the timeline into $T$-period review phases. Strategy $\hat{C}^i$ starts with the $T$-period substrategy $\hat{C}_T^i$, and $\hat{D}^i$ starts with $\hat{D}_T^i$. 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

$$
\begin{cases}
\hat{C}^i & \text{with probability } (W^i_C(h_T^i) - G^i_D)/(G^i_C - G^i_D) \\
\hat{D}^i & \text{with probability } (G^i_C - W^i_C(h_T^i))/(G^i_C - G^i_D),
\end{cases}
$$

thereby assigning to the opponent an expected payoff of $W^i_C(h_T^i)$. 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 $W^i_D(h_T^i)$.

Notice that the strategies $\hat{C}^i$ and $\hat{D}^i$ 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 $\hat{C}^i$ and $\hat{D}^i$ are best responses to $\hat{C}^j$ and $\hat{D}^j$. From the properties of these strategies outlined in the statement of the proposition, it follows immediately that $G^i_C$ is the payoff in response to $\hat{C}^j$ from any strategy that involves $\hat{C}^i$ or $\hat{D}^i$ in each review phase, and in particular strategies $\hat{C}^i$ and $\hat{D}^i$. Similarly, $G^j_D$ is the payoff in response to $\hat{D}^j$ from any of those strategies.

Let us show that $G^i_C$ and $G^j_D$ are the maximal expected payoffs that player $i$ can achieve in response to $\hat{C}^j$ and $\hat{D}^j$. If not, let $\hat{A}^i_C$ and $\hat{A}^i_D$ be strategies that achieve the maximal expected payoffs of $F^i_C \geq G^i_C$ and $F^i_D \geq G^i_D$ (with at least one strict inequality) in response to $\hat{C}^j$ and $\hat{D}^j$, respectively. Without loss of generality, assume that $F^i_C - G^i_C \geq F^i_D - G^i_D$.

Consider player $i$ playing $\hat{A}^i_C$ in response to $\hat{C}^j$. 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

$$
\delta^T \frac{W^i_C(h_T^j) - G^i_D}{G^i_C - G^i_D} F^i_C + \delta^T \frac{G^i_C - W^i_C(h_T^j)}{G^i_C - G^i_D} F^i_D \leq \delta^T (F^i_C - G^i_C) + \delta^T \frac{W^j_C(h_T^i) - G^i_D}{G^i_C - G^i_D} G^i_C + \delta^T \frac{G^i_C - W^j_C(h_T^i)}{G^i_C - G^i_D} G^i_D = \delta^T (F^i_C - G^i_C + W^j_C(h_T^i)).
$$
Then, player i’s expected payoff at time 0 cannot be greater than

\[ E \left[ \sum_{s=0}^{T-1} \delta^s g^i(a^i_s \alpha_s) + \delta^T (F^i_C - G^i_C + W^j_C(h^j_T)) \mid \hat{A}, \hat{C} \right] \leq \delta^T (F^i_C - G^i_C) + G^i_C \]

by (9). This is less than \( F^i_C \), 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 \((w_1, w_2) \in [G^1_D, G^1_C] \times [G^2_D, G^2_C]\), one Nash equilibrium that achieves it is

\[
\left( \frac{w_1 - G^2_D}{G^2_C - G^2_D} \bar{C}^1 + \frac{G^2_C - w_1}{G^2_C - G^2_D} \bar{D}^1, \frac{w_2 - G^1_D}{G^1_C - G^1_D} \bar{C}^2 + \frac{G^1_C - w_2}{G^1_C - G^1_D} \bar{D}^2 \right).
\]

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

\[ \square \]

### A.2 Proof of Proposition 4.

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

\[ \square \]
References


