

Sequential Equilibria in the Repeated Prisoners' Dilemma with Private Monitoring

V. Bhaskar*

Dept. of Economics

University of Essex

Wivenhoe Park

Colchester CO4 3SQ, UK

Email: vbhas@essex.ac.uk

March 1999; Revised May 1999

Abstract

We analyze the infinitely repeated prisoners' dilemma with imperfect private monitoring. The efficient outcome can be approximated in any prisoners' dilemma game, while every individually rational feasible payoff can be approximated in a class of prisoner dilemma games. Our results require that monitoring be sufficiently accurate but do not require very low discounting.

1 Introduction

We analyze the infinitely repeated prisoners' dilemma with imperfect private monitoring and discounting. The main contribution of this paper is to construct "belief-based" strategies, where a player's continuation strategy is a function only of his beliefs. This simplifies the analysis considerably, and allows us to explicitly construct sequential equilibria for such games, thus

*I am grateful to Michihiro Kandori for many discussions on this topic, and to the University of Tokyo for its hospitality while this paper was written.

enabling us to invoke the one-step deviation principle of dynamic programming. By doing so, we prove that one can approximate the efficient payoff in any prisoners' dilemma game provided that the monitoring is sufficiently accurate. Furthermore, for a class of prisoners' dilemma games, one can approximate every individually rational feasible payoff. These results require that monitoring be sufficiently accurate, but only require a uniform lower bound on the discount rate.

These results are closely related to an important paper by Sekiguchi [10], who shows that one can approximate the efficient payoff in such games provided that the monitoring is sufficiently accurate. Sekiguchi's result applies for a class of prisoners' dilemma payoffs, and relied on the construction of a Nash equilibrium which achieves approximate efficiency. Standard arguments can then be invoked to show that there exists a sequential equilibrium with the same outcome. However, in order to show that a strategy profile is a Nash equilibrium, Sekiguchi has to invoke a path dominance argument, which requires the payoff restrictions imposed. The explicit construction of sequential equilibrium shows that one does not require such a payoff restriction.

2 Approximating the Efficient Payoff

	C	D
C	1	$-l$
D	$1 + g$	0

We consider the prisoners' dilemma with the stage game payoffs given above, where the row indicates the player's own action and the column indicates his opponent's action. Players only observe their own actions, and also observe a private signal which is informative about their opponent's action. This signal belongs to the set $\Omega = \{c, d\}$, where c (resp. d) is more likely when the opponent plays C (resp. D). The signalling structure is assumed symmetric, in the sense that the probability of errors does not depend on the action profile played. Given any action profile $a = (a_1, a_2)$, $a_i \in A = \{C, D\}$, the probability that exactly one player receives a wrong signal is $\varepsilon > 0$, and the probability that both receive wrong signals is $\xi > 0$. Players maximize the expected sum of stage game payoffs discounted at rate δ . We also assume that at the end of each period, players observe the realization of a public randomization device uniformly distributed on the unit interval.

Our approach is closely related to Sekiguchi's [10]: we show that one can construct a mixed trigger strategy sequential equilibrium which achieves partial cooperation. With public randomization, one can modify this appropriately in order to approximate full cooperation. Our approach involves the construction of a "belief-based" strategy, i.e. a strategy which is a function of the player's beliefs about his opponent's continuation strategy. This results in a major simplification as compared to the more conventional notion of a strategy which is a function of the private information of the player.

We begin by defining *partial continuation strategies*. In any period t , define the partial continuation strategy σ_D as follows: play D at period t , and at period $t + 1$ play σ_D if the realized outcomes in period t are (Dc) or (Dd) . Define the partial continuation strategy σ_C as follows: in any period t play C ; at period $t + 1$ play σ_C if the realized outcomes in period t is (Cc) , and play σ_D if the realized outcome at t is (Cd) . We call σ_C and σ_D a partial continuation strategy since each of these fully specifies the player's actions in every subsequent period at every information set that arises given that he confirms to the strategy. In consequence, the (random) path and payoffs induced by any pair of partial continuation strategies is well defined. However, a partial continuation strategy does not specify the player's actions in the event that he deviates from the strategy at some information set. This is deliberate, since our purpose is to construct the full strategies that constitute a sequential equilibrium. Note also that for any player i , only the partial continuation strategy of player j is relevant when computing i 's payoffs in any equilibrium.

Let $V_{ab}(\delta, \varepsilon, \xi)$, $a, b \in \{C, D\}$ denote the repeated game payoff of σ_a against σ_b — these payoffs are well defined since the path induced by each pair is well defined. We have that $V_{DD} > V_{CD}$, for all parameter values. Furthermore, if $\delta > \frac{g}{1+g}$, and $(\varepsilon + \xi)$ is sufficiently small, then $V_{CC} > V_{DC}$. Suppose that player i believes that his opponent is playing either σ_C or σ_D , and is playing σ_C with probability μ . Then the difference between his payoff from playing σ_C and his payoff from playing σ_D is given by

$$\Delta V(\mu; \delta, \varepsilon, \xi) = \mu(V_{CC} - V_{DC}) - (1 - \mu)(V_{DD} - V_{CD}) \quad (1)$$

Hence $\Delta V(\mu)$ is increasing and linear in μ and there is a unique value, $p(\delta, \varepsilon, \xi)$, at which it is zero. Suppose now that at $t = 1$ both players are restricted to choosing between σ_C and σ_D . There is a mixed equilibrium of the restricted game, where each player plays the strategy σ which plays σ_D

with probability $1-p$ and σ_C with probability p . Note that $p(\delta, \varepsilon, \xi)$ increases to 1 as we decrease δ towards its lower bound $\frac{g}{1+g}$. Let δ be such that $p > \frac{1}{2}$.

For future reference we emphasize that equation (1) applies to any period — if a player believes that his opponent's continuation strategy is σ_C with probability μ and σ_D with probability $1-\mu$, then he prefers σ_C to σ_D if $\mu > p$ and prefers σ_D to σ_C if $\mu < p$. Note also that if a player's opponent begins at $t=1$ with a strategy in $\{\sigma_C, \sigma_D\}$, his continuation strategy also belongs to this set, since σ_D induces only σ_D , while σ_C may induce either σ_C or σ_D , depending upon the private history that the opponent has observed.

We define the following four belief revision operators. Starting with any initial belief μ , we can define the player's new beliefs when he takes action a and receives signal ω . His new belief (i.e. the probability that j 's continuation strategy is σ_C) will be given by $\chi_{a\omega}(\mu)$. We have four belief operators, χ_{Cc} , χ_{Cd} , χ_{Dc} , χ_{Dd} , where each $\chi_{a\omega} : [0, 1] \rightarrow [0, 1]$ is defined, using Bayes rule, as follows

$$\chi_{Cc}(\mu) = \frac{\mu(1-2\varepsilon-\xi)}{\mu(1-\varepsilon-\xi) + (\varepsilon+\xi)(1-\mu)} \quad (2)$$

$$\chi_{Cd}(\mu) = \frac{\mu\varepsilon}{\mu(\varepsilon+\xi) + (1-\varepsilon-\xi)(1-\mu)} \quad (3)$$

$$\chi_{Dc}(\mu) = \frac{\mu\varepsilon}{\mu(1-\varepsilon-\xi) + (\varepsilon+\xi)(1-\mu)} \quad (4)$$

$$\chi_{Dd}(\mu) = \frac{\mu\xi}{\mu(\varepsilon+\xi) + (1-\varepsilon-\xi)(1-\mu)} \quad (5)$$

Starting with any initial belief $\hat{\mu}$ at the beginning of the game, a player's belief at any private history, i.e. after an arbitrary sequence $(a\omega)_{r=1}^t$, can be computed by iterated application of the appropriate belief operators. Let $\Xi(\hat{\mu})$ be the set of *possible beliefs*, i.e. $\mu \in \Xi(\hat{\mu}) \Leftrightarrow \exists \langle \mu_r \rangle_{r=1}^t : \mu_1 = \hat{\mu}, \mu_t = \mu$ and $\mu_{r+1} = \chi_{(a\omega)_r}(\mu_r), (a\omega)_r \in A \times \Omega, 1 \leq r \leq t-1$. Let τ be a (full) strategy, which is defined at every information set, i.e. after arbitrary private histories. Clearly, τ is a best response to σ after every private history if and only if it is optimal to play τ at every belief $\mu \in \Xi(p)$, i.e. at all possible beliefs given the initial belief p .

We now examine the properties of these belief operators. First, each is a strictly increasing function.¹ Fig. 1 graphs the first two belief opera-

¹The derivative is strictly positive since the derivative of the numerator is strictly positive while the derivative of the denominator is zero for every $\chi_{a\omega}$.

tors, χ_{C_c} and χ_{C_d} , which play a particularly important role in the analysis.

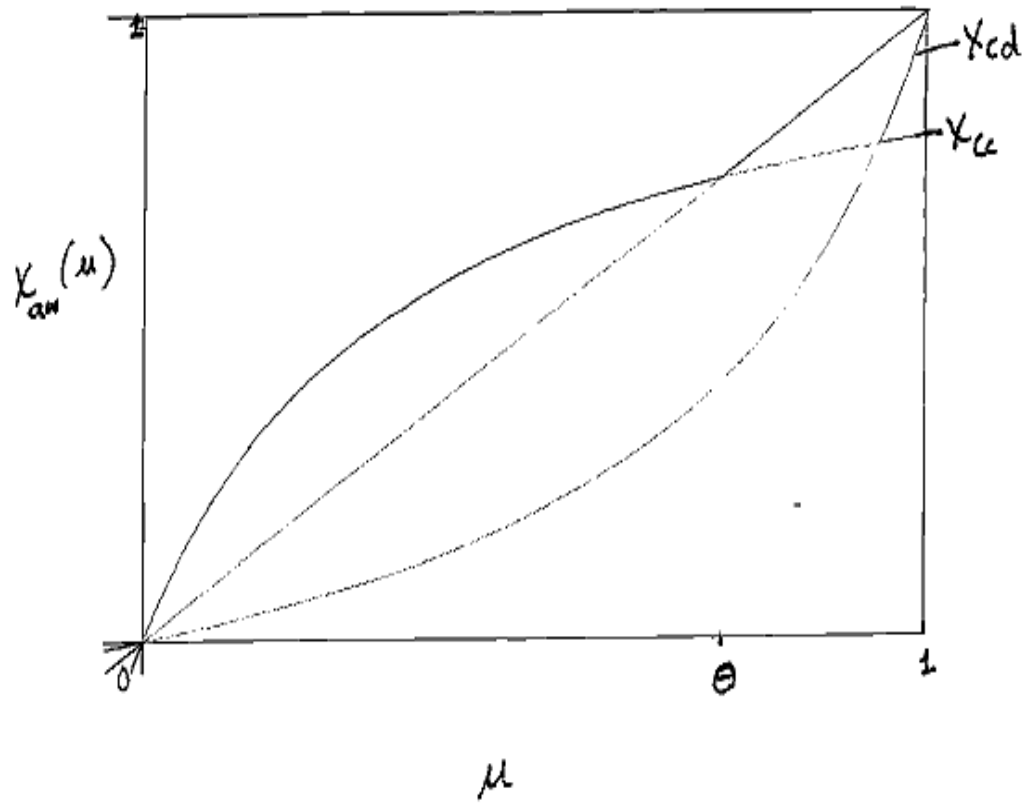


Fig. 1

The operator χ_{C_c} has an interior fixed point at θ , and $\chi_{C_c}(\mu) \leq \mu$ as $\mu \geq \theta$ (cf. Fig. 1). The value of θ depends upon (ε, ξ) in the following way

$$\theta(\varepsilon, \xi) = \frac{1 - 3\varepsilon - 2\xi}{1 - 2\varepsilon - 2\xi} \quad (6)$$

We shall assume that $\max\{\varepsilon, \xi\} < 1 - \varepsilon - \xi$, which in turn entails that

$\theta > \frac{1}{2}$.

We claim that the graph of each of the other three operators ($\chi_{Cd}, \chi_{Dc}, \chi_{Dd}$) lies below the 45⁰ line on Fig 1, provided that $\max\{\varepsilon, \xi\} < 1 - \varepsilon - \xi$. To verify this, take any typical expression from 3-5, and divide by μ . This yields ε (or ξ) in the numerator, while the denominator is strictly larger since it is a convex combination of $(\varepsilon + \xi)$ and $(1 - \varepsilon - \xi)$. Hence $\chi_{a\omega}(\mu) < \mu$ for each of these three operators.

Since $\mu < \theta \Rightarrow \chi_{a\omega}(\mu) < \mu$ for any belief operator, this immediately implies that if $p < \theta$, $\Xi(p) \subseteq [0, \theta]$. This follows from the fact that the initial belief p is strictly less than θ , and since we have demonstrated that no point $\mu' > \theta$ is the image of any $\mu \leq \theta$ under *any* belief operator.

Hence, provided that initial beliefs are given by $p < \theta$, it suffices to define our belief based strategy for beliefs in the set $[0, \theta]$. Let $\rho : [0, \theta] \rightarrow \{\sigma_C, \sigma_D, \sigma\}$ be defined as follows: $\rho(\mu) = \sigma_C$ if $\mu \in (p, \theta]$ and $\rho(\mu) = \sigma_D$ if $\mu \in [0, p)$. If $\mu = p$, $\rho(\mu) = \sigma$, i.e. ρ plays σ_C with probability p and σ_D with probability $1 - p$. Hence the pair (ρ, p) , i.e. ρ in conjunction with an initial belief p , specifies a strategy at every possible belief.

The advantage of this specification is that a player's continuation strategy is specified even at information sets which arise due to a player's deviating from ρ in the past. However, the definition of ρ is potentially problematic, in the sense that it may be inconsistent, i.e. for example, at some μ , $\rho(\mu)$ could prescribe σ_C , but $\rho(\chi_{Cc}(\mu))$ could prescribe σ_D . We now show that this problem does not arise.

Definition 1 $(\rho, \hat{\mu})$ is consistent if and only if $\forall \mu \in \Xi(\hat{\mu}), \rho(\mu) \in \{\sigma_C, \sigma\} \Rightarrow [\rho(\chi_{Cc}(\mu)) = \sigma_C \text{ and } \rho(\chi_{Cd}(\mu)) = \sigma_D]$ and $\rho(\mu) \in \{\sigma_D, \sigma\} \Rightarrow [\rho(\chi_{Dc}(\mu)) = \sigma_D \text{ and } \rho(\chi_{Dd}(\mu)) = \sigma_D]$.

Lemma 2 If $\frac{1}{2} < p < \theta(\varepsilon, \xi)$, (ρ, p) is consistent.

Proof. To establish that (ρ, p) is consistent we need to verify the following:

1. $\mu \in [p, \theta] \Rightarrow \chi_{Cc}(\mu) > p$.
2. $\mu \in [p, \theta] \Rightarrow \chi_{Cd}(\mu) < p$.
3. $\mu \leq p \Rightarrow \chi_{Dc}(\mu) < p$.
4. $\mu \leq p \Rightarrow \chi_{Dd}(\mu) < p$.

To verify 1, recall that $\chi_{Cc}(\mu) > \mu$ if $\mu < \theta$, so that $\chi_{Cc}^k(p) > p$ for any k . 3 and 4 follow from the fact already established that χ_{Dc} and χ_{Dd} lie below the 45⁰ line, and the fact that these are increasing functions. To verify 2, it suffices to verify that $\chi_{Cd}(\theta) \leq p$, since χ_{Cd} is strictly increasing.

$$\chi_{Cd}(\theta) = \frac{\varepsilon(1 - 3\varepsilon - 2\xi)}{(\varepsilon + \xi)(1 - 3\varepsilon - 2\xi) + (1 - \varepsilon - \xi)\varepsilon} < \frac{\varepsilon}{2\varepsilon} = \frac{1}{2} < \theta \quad (7)$$

■

Note that if ε and ξ are sufficiently small, we can select $\delta > \frac{g}{1+g}$ so that $p(\delta, \varepsilon, \xi) \in (\frac{1}{2}, \theta)$ — this follows from the fact that $p(\delta, \varepsilon, \xi) \rightarrow 1$ as $\delta \rightarrow \frac{g}{1+g}$ and $(\varepsilon + \xi) \rightarrow 0$, while $p(\delta, \varepsilon, \xi) \rightarrow 0$ if $\delta \rightarrow 1$ and $(\varepsilon + \xi) \rightarrow 0$. Henceforth we shall assume that δ is such that $p \in (\frac{1}{2}, \theta)$ so that (ρ, p) is consistent.

We have therefore established that the pair (ρ, p) defines a full strategy which is behaviorally equivalent to σ .

Proposition 3 *If $\frac{1}{2} < p < \theta(\varepsilon, \xi)$, the strategy profile where each player plays (ρ, p) is a sequential equilibrium.*

Proof. Note first that if $\mu = p$, a player is indifferent between playing σ_C and σ_D , and hence a one-step deviation from playing ρ is not profitable. Since the payoffs from playing σ, σ_C and σ_D are equal at belief p , one may also, for the purposes of computing payoffs, use σ_C or σ_D as is computationally convenient in the event of belief p .

Consider first the case when $\mu > p$. A one-step deviation from ρ is to play D , and to continue with ρ in the next period. The following sub-cases arise:

a) Suppose that $\chi_{Dc}(\mu) \leq p$ and $\chi_{Dd}(\mu) \leq p$. In this case, a one-step deviation from ρ is to play σ_D , whereas $\rho(\mu) = \sigma_C$. However, (1) establishes that in this case σ_C is preferable to σ_D , and hence a one-step deviation from ρ is unprofitable.

b) Suppose that $\chi_{Dc}(\mu) \leq p$ and $\chi_{Dd}(\mu) > p$, so that the one-step deviation is play D today and continue with σ_D if Dc is reached, and to continue with σ_C if Dd is reached. Let $\Delta\tilde{V}(\mu)$ be payoff difference between the equilibrium strategy and the one-step deviation.. Note that the one step deviation differs from σ_D only at the information set Dd ; at this information it continues by playing σ_C whereas σ_D continues with σ_D . Hence we can write $\Delta\tilde{V}(\mu)$ as the payoff difference between σ_C and σ_D minus the payoff difference between σ_C and σ_D conditional on Dd being reached, as follows:

$$\Delta\tilde{V}(\mu) = \Delta V(\mu) - \delta[\mu(\varepsilon + \xi) + (1 - \mu)(1 - \varepsilon - \xi)][\Delta V(\chi_{Dd}(\mu))] \quad (8)$$

Note that $\chi_{Dd}(\mu) < \mu$. Equation (1) shows that this implies that $\Delta V(\mu) > \Delta V(\chi_{Dd}(\mu))$. Since the coefficient multiplying $\Delta V(\chi_{Dd}(\mu))$ is strictly less

than one, this implies that $\Delta\tilde{V}(\mu) > 0$. Hence if $\mu > p$, a one-step deviation is unprofitable.

c) Finally, we establish that $\chi_{Dc}(\mu) < p \forall \mu \leq \theta$, so that no other sub-case need be considered. Evaluating χ_{Dc} at the upper bound θ , we have

$$\chi_{Dc}(\theta) = \frac{\varepsilon(1 - 3\varepsilon - 2\xi)}{(\varepsilon + \xi)(1 - 3\varepsilon - 2\xi) + (1 - \varepsilon - \xi)\varepsilon} < \frac{1}{2} \quad (9)$$

Consider now the case when $\mu < p$. In this case, a one-step deviation from ρ is to play C today, and to continue with σ_C if $\chi_{Cc}(\mu) \geq p$, but to continue with σ_D if $\chi_{Cc}(\mu) < p$. (Note that $\mu < p \Rightarrow \chi_{Cd}(\mu) < p$, so the continuation strategies do not differ in this event.) In the first sub-case, the one-step deviation from ρ corresponds to playing σ_C , and (1) establishes that in this case σ_D is preferable to σ_C , and hence a one-step deviation from ρ is unprofitable. In the second sub-case, the one-step deviation differs from σ_C only at the information set Cc — it plays σ_D at this information set rather than σ_C . Let $\Delta\hat{V}(\mu)$ denote the payoff difference between the one-step deviation and the equilibrium strategy σ_D . We have

$$\Delta\hat{V}(\mu) = \Delta V(\mu) - \delta[\mu(1 - \varepsilon - \xi) + (1 - \mu)(\varepsilon + \xi)][\Delta V(\chi_{Cc}(\mu))] \quad (10)$$

Since $p > \chi_{Cc}(\mu) > \mu$, $\Delta V(\mu) < \Delta V(\chi_{Cc}(\mu)) < 0$. Also, the coefficient multiplying $\Delta V(\chi_{Cc}(\mu))$ is less than 1 which establishes that $\Delta\hat{V}(\mu) > 0$.

We have therefore established that if a player's opponent j plays the strategy σ (which randomizes between σ_C and σ_D), it is optimal for player i to play ρ , with initial belief p . However, (ρ, p) is consistent and behaviorally equivalent to the strategy σ . Hence the profile where both players play (ρ, p) is a sequential equilibrium. ■

It may be of interest to ask, under what conditions is there a pure strategy sequential equilibrium where both players begin in period one by playing σ_C with probability one. The above analysis also permits an answer to this question, with the difference that the initial belief $\hat{\mu} = 1$ rather than p . Only consistency conditions 1 and 2 are affected, and must be modified as follows:

- 1'. $\chi_{Cc}^k(1) > p \forall k$.
- 2'. $\mu = 1$ or $\mu = \chi_{Cc}^k(1)$ for any $k \Rightarrow \chi_{Cd}(\mu) < p$.

Clearly, 1' is always satisfied as long as $p \leq \theta$. However, for 2' to be satisfied, it is necessary and sufficient that

$$\chi_{Cd}(1) = \frac{\varepsilon}{\varepsilon + \xi} \leq p < \theta \quad (11)$$

In Fig. 1, this requires that the χ_{Cd} function always lies below θ , which requires the inequality

$$\varepsilon^2 < \xi(1 - 3\varepsilon - 2\xi) \quad (12)$$

This inequality will be satisfied if ε is sufficiently small relative to ξ , i.e. if signals are sufficiently positively correlated. It is easily verified that this inequality cannot be satisfied if signals are independent or negatively correlated so that the equilibrium must be in mixed strategies.

Note that p plays a dual role in the construction of the mixed strategy equilibrium. On the one hand it is the randomization probability in the first period, and on the other hand, it is simply a number which defines the threshold at which behavior changes. These roles are obviously distinct, as is apparent from our discussion of the pure strategy equilibrium. This distinction is particularly relevant when we discuss the folk theorem in the following section.

With the construction of the mixed equilibrium, it is easy to show that one can use public randomization to approximate full cooperation.

Lemma 4 *If (v_1, v_2) is a sequential equilibrium payoff for some $\delta \in (0, 1)$, it is also an equilibrium payoff for any $\delta' > \delta$ if a public randomization device is available.*

Proof. Let τ be the strategy profile giving the required payoff given δ . Given δ' , let $m = \frac{\delta}{\delta'}$. Players play a sequence of games: they begin with the strategy profile τ . If the sunspot in any period $\phi > m$, they play a new game and re-start with τ . ■

Proposition 5 *If $\delta > \frac{g}{1+g}$, any payoff $x < 1$ is a symmetric equilibrium payoff if ε and ξ are sufficiently small.*

Proof. For any (ε, ξ) select $\delta(\varepsilon, \xi)$ so that $\frac{1}{2} < p < \theta(\varepsilon, \xi)$. We have verified that under these conditions ρ is a sequential equilibrium. Let $(\varepsilon, \xi) \rightarrow (0, 0)$ and $\delta(\varepsilon, \xi) \rightarrow \frac{g}{(1+g)}$, so that $p \rightarrow 1$. The equilibrium payoff tends to one. Lemma 4 ensures that this result holds for all $\delta > \frac{g}{1+g}$. ■

Note that in order to approximate the payoff $(1, 1)$, we require that the noise vanishes but do not require that discounting vanishes. In this sense the result proved here is stronger than that proved in Piccione [9], who requires that both discounting and the noise vanish in order to approximate the efficient payoff.

The equilibrium we have constructed is a very robust one. Equilibria in repeated games with imperfect monitoring are usually robust to refinements involving strategy perturbations. Since the monitoring structure has full support, a deviation by player i from his equilibrium strategy cannot be observed by the other player.² However, equilibria which have been constructed for repeated games with imperfect *private* monitoring are often not robust if there is a small amount of incomplete information about each player's payoffs, as in Harsanyi [6]. This issue is discussed in greater detail in Bhaskar [1], who proposes a refinement for repeated game equilibria based on robustness to payoff perturbations.³ This refinement requires the following condition: consider two information sets for a player at date t , h and h' , and where the player's continuation strategy is γ at h and γ' at h' . If it is also optimal for the player to play γ at h' and γ' at h , then the equilibrium is *robust* only if $\gamma = \gamma'$. To verify that the equilibrium (ρ, p) is robust, note that a player is indifferent between his continuation strategies only at the belief $\mu = p$, and at any information set where this belief arises, he always plays the same strategy σ . This robustness also applies to the private monitoring equilibria constructed by several others including Bhaskar and van Damme [3] Mailath and Morris [8], and Sekiguchi [10]. The equilibria constructed in these papers have the common feature that a player's beliefs (regarding his opponent's future behavior) vary with his private information, and his continuation strategy also varies with his beliefs. This contrasts with a different approach to constructing equilibria, where a player plays different continuation strategies γ at h and γ' at h' and is indifferent between the two strategies at both these information sets. This is the approach taken in Piccione [9], and also by Compte [4] and Kandori and Matsushima [7] in their analysis of private monitoring with independent signals.⁴ Piccione uses a strategy whereby a player is made indifferent between playing C and D at each period in which he is supposed to play C , and plays C when he observes a good signal and D when he observes a bad signal. These equilibria are not robust to payoff perturbations since players will ignore their payoff irrelevant information and only condition upon their payoff information.

²See Fudenberg, Levine and Maskin [5], p1025, for a discussion.

³See also [2] for a specific application of this refinement.

⁴The equilibria constructed by Compte and Kandori-Matsushima are robust in the case where signals are correlated; however, their two-player folk theorem results require independent signals.

3 Approximating Any Individually Rational Feasible Payoff

We now build on the construction of the previous section and show how to approximate any individually rational feasible payoff. The key to doing this is to approximate the payoff $(\frac{1+g+l}{1+l}, 0)$, which is player 1's maximal payoff within the set of individually rational and feasible payoffs in the usual prisoners' dilemma, where (C, C) is the symmetric efficient payoff. Since the payoff $(1, 1)$ has already been approximated in the previous section, one can then use public randomization to approximate any individually rational feasible payoff.

It might be useful to outline the basic construction and to explain the complications that arise. The basic idea of our construction is that play begins in the *asymmetric phase* where player 1 plays D and player 2 randomizes, playing C with a high probability, q_2 . This asymmetric phase continues or ends, depending upon the realization of a public randomization device. Thus player 1's per-period payoff in the asymmetric phase is approximately $1 + g$ while player 2's per-period payoff is approximately $-l$. Since the latter is less than the individually rational payoff for player 2, he must be rewarded for playing C . To ensure this, when the asymmetric phase ends, both player's continuation strategies depend upon their private information. Player 1 continues with σ_C if he has observed the signal c in the last period (i.e. if his information is Dc) and continues with σ_D if he has observed d (i.e. if his information is Dd). This ensures that player 2 is rewarded for playing C in the asymmetric phase. Similarly, player 2 continues with σ_C if his private information is Cd , the information set which is most likely when he plays C , and continues with σ_D if his private information is Dd . Hence, if the noise is small, player 2's continuation payoff when the asymmetric phase ends is approximately 1 if he has played C in the previous period and approximately zero if he has played D . Hence if δ is large relative to l ($\delta > \frac{l}{1+l}$), we can, by choosing the value of the sunspot appropriately, make player 2 indifferent between C and D in the asymmetric phase. The payoffs in this equilibrium converge to $(\frac{1+g+l}{1+l}, 0)$ as the noise vanishes.

However, one must also verify that the players find it optimal to play σ_C and σ_D , as appropriate, at each information set after the end of the asymmetric phase. A complication arises here, as compared to the previous section, since player 1 does not randomize in the asymmetric phase, i.e. he

plays D for sure. (Indeed, he cannot play C with positive probability, since in that case his payoff in the asymmetric phase is bounded above by 1 and hence cannot approximate $1 + g$).⁵ Hence when player 2 receives the signal c , he knows that there has been at least one error in signals, and his beliefs about player 1's continuation strategy depend upon the relative probability of one (ε) versus two errors (ξ). In other words, his continuation strategy at the information sets Cc and Dc depends upon the correlation structure of signals. Since player 2's continuation strategy depends upon this correlation structure, this implies that player 1's beliefs also depend upon the correlation structure.

We adopt two alternative approaches to handle this problem. First, we show that if signals are positively correlated, so that the probability of two errors is at least as large as the probability of one error, then one can approximate the asymmetric payoff, without any restriction upon payoffs. Second, we show that one does not need such positive correlation of signals provided that one can choose δ so that $p(\delta, \varepsilon, \xi)$ sufficiently close to one. This result applies to any prisoners' dilemma game where $g \geq l$ — in any such game one can approximate the asymmetric payoff arbitrarily closely. However, this second approach does not work if $l > g$, since in this case one cannot have $p(\delta, \varepsilon, \xi) \rightarrow 1$. The reason for this is the for p to be close to 1, we must have $\delta \rightarrow \frac{g}{1+g}$. However, in the asymmetric phase, player 2 incurs a loss of l by playing C , whereas his continuation payoff gain is no more than 1. Hence player 2 will be willing to play C in the asymmetric phase only if $\delta > \frac{l}{1+l}$. Hence if $l > g$, one cannot have p close to 1 since δ is bounded away from $\frac{g}{1+g}$.

We make the following assumption for this section:

Assumption A Either A1: $\xi \geq \varepsilon$ or A2: $g \geq l$ and $\xi(1-\xi)(1-2\varepsilon-\xi) > \varepsilon^3$.

Note that A1 is a relatively strong assumption that signals are positively correlated, but does not require any assumption on payoffs. On the other hand, A2 requires an assumption on payoffs but is a mild assumption about the relative probability of errors. It is always satisfied if signals are positively correlated, or independent. In the independent signal case, the left hand side is a term of order $(\varepsilon+\xi)^2$ whereas the right hand side is a term of order $(\varepsilon+\xi)^3$.

⁵This argument is more general and implies that one cannot have a folk theorem in completely mixed strategies in any repeated game. Let \hat{v}_1 be the supremum payoff of player 1 in any equilibrium where player 1 randomizes in every period at every information set. Since $\hat{v}_1 \leq (1-\delta) \min_{a_1} \{\max_{a_2} u_1(a_1, a_2)\} + \delta \hat{v}_1$, this implies $\hat{v}_1 \leq \min_{a_1} \{\max_{a_2} u_1(a_1, a_2)\}$.

Hence A2 is satisfied even if signals are negatively correlated provided that they are not too highly so.

We now define the players' strategies more precisely. In any period $t - 1$ in the asymmetric phase, player 1 plays D for sure, while player 2 randomizes between C and D , choosing C with a constant probability q_2 which is close to 1. At the end of period, players observe the realization, ϕ_{t-1} , of a sunspot which is uniformly distributed on $[0, 1]$. If $\phi_{t-1} > 1 - \lambda$, both players continue in the asymmetric phase for the next period. If $\phi_{t-1} \leq \lambda$, the asymmetric phase ends for both players, and is never reached again. In this case, each player's continuation strategy (i.e. his state) depends upon the realization of his private information, at date $t - 1$. (i.e. players ignore their private information from previous dates). Let ν_{t-1} denote the player's private information realization at date $t - 1$. Player 1 continues with σ_C if $\nu_{t-1} = Dc$; if $\nu_{t-1} = Dd$, he continues in period t with σ_D .⁶ Player 2's continues with σ_C if $\nu_{t-1} = Cd$, and continues with σ_D if $\nu_{t-1} = Dd$. If $\nu_{t-1} \in \{Cc, Dc\}$, player 2 continues with σ_C if $\mu_2(\nu_{t-1}) > p(\delta, \varepsilon, \xi)$ and with σ_D if $\mu_2(\nu_{t-1}) \leq p(\delta, \varepsilon, \xi)$.

Our analysis proceeds as follows. First, we show that player 2 is willing to randomize in the asymmetric phase provided that λ is appropriately chosen, and that the payoffs associated with this class of equilibria tend to $(\frac{1+g+l}{1+l}, 0)$ as the noise vanishes. Subsequently, we shall demonstrate that all players are choosing optimally at every information set.

Write $W_2(D)$ for the payoff of player 2 in the asymmetric phase given that he plays D , and $W_2(C)$ for the payoff in the asymmetric phase from playing C . Since $W_2(D) = W_2(C) = W_2$, we have

$$W_2(D) = \delta(1 - \lambda)W_2 + \delta\lambda V_2(D) \quad (13)$$

where $V_2(D)$ is the expected payoff to player 2 conditional on the fact that the asymmetric phase has ended and that he has played D . Similarly, letting $V_2(C)$ be the expected payoff to 2 conditional on the fact that the asymmetric phase has ended and that he has played C , we have

$$W_2(C) = (1 - \delta)(-l) + \delta(1 - \lambda)W_2 + \delta\lambda V_2(C) \quad (14)$$

⁶We show that any strategy which plays C in the asymmetric phase is dominated, and hence we need not define precisely the optimal continuation strategy after playing C . The existence of an optimal continuation strategy follows from the same argument as in Sekiguchi [10]. Since player 1 never plays C in the asymmetric phase, his continuation after his own deviation does not affect player 2's incentives.

Clearly, $V_2(D) \rightarrow 0$ as $(\varepsilon, \xi) \rightarrow (0, 0)$. We now show that $V_2(C) \rightarrow 1$ as $(\varepsilon, \xi) \rightarrow (0, 0)$. Let $V_2(Cd)$ (resp. $V_2(Cc)$) denote the continuation payoff at the end of the asymmetric phase, conditional on Cd (resp. Cc). Since player 1 plays D for sure in the asymmetric phase, we have

$$V_2(C) = (1 - \varepsilon - \xi)V_2(Cd) + (\varepsilon + \xi)V_2(Cc) \quad (15)$$

Hence it suffices to establish that $V_2(Cd) \rightarrow 1$ as $(\varepsilon, \xi) \rightarrow (0, 0)$. Write $\mu_2(Cd)$ for the probability that player 1's continuation strategy is σ_C , given that $\nu_{t-1} = Cd$. Since $\mu_2(Cd) \geq \frac{1-2\varepsilon-\xi}{1-\varepsilon-\xi}$, $\mu_2(Cd) \rightarrow 1$ as $\varepsilon \rightarrow 0$. Hence from equation (1) $V_2(Cd) \rightarrow V_{CC}$, where $V_{CC} \rightarrow 1$ as $(\varepsilon, \xi) \rightarrow (0, 0)$.

Hence if $\varepsilon + \xi$ is sufficiently small and $\delta > \frac{l}{1+l}$, there exists a value of λ which equates $W_2(C)$ and $W_2(D)$. Further, as $(\varepsilon + \xi) \rightarrow 0$, $\lambda \rightarrow \frac{(1-\delta)l}{\delta}$, and player 2's payoff converges to zero.

If $q_2 \rightarrow 1$, player 1's per-period payoff tends to $(1+g)$ in the asymmetric phase, and 1 in the cooperative phase. By substituting for the limiting value of λ which is $\frac{(1-\delta)l}{\delta}$, we see that player 1's payoff converges to $\frac{1+g+l}{1+l}$. (We shall establish later that $q_2 \rightarrow 1$).

We now verify that each player plays optimally at each information set in this equilibrium. In the asymmetric phase, this is so for player 2 by construction, since he is indifferent between C and D . It is easy to see that player 1 also plays optimally in the asymmetric phase, since he is choosing his one shot best response.⁷

Consider now the transition to the cooperative phase, i.e. the player's actions in the first period after the sunspot signals the end of the asymmetric phase. Since each player only conditions on his private information in the previous period, we may focus on this alone. Player 1 has two possible information sets, (Dc) and (Dd) , whereas player 2 has four possible information sets. Let $\mu_i(\nu)$ denote the probability assigned by player i to his opponent's continuation strategy being σ_C , given that i is at information set ν .

We shall assume that parameters are such that $\varepsilon + \xi < \frac{1}{3}$ and $p(\delta, \varepsilon, \xi) \in (\frac{1}{2}, \frac{1-2\varepsilon-\xi}{1-\varepsilon-\xi})$. It is easily verified that this assumption on p does not imply any

⁷It is possible that playing C in the asymmetric phase increases player 1's continuation payoff in the cooperative phase. However, it is easy to see that such an increase can never offset the loss from playing C . A simple proof is as follows. If playing C in the asymmetric phase is optimal for 1, then playing C in every period in the asymmetric phase is also optimal. The overall payoff of this strategy is approximately 1 if the noise is small, whereas the payoff of player 1 in the equilibrium tends to $\frac{1+g+l}{1+l}$, which is strictly greater.

restrictions upon g or l . However, if we invoke assumption A2, then we may also choose p to be arbitrarily close to its upper bound. We shall also assume that $q_2 \in (p, \gamma(\varepsilon, \xi))$, where $\gamma(\varepsilon, \xi) = \min\{\frac{1-2\varepsilon-2\xi}{1-2\varepsilon}, \frac{(\varepsilon+\xi)(1-2\varepsilon-\xi)}{\varepsilon(1-\varepsilon-\xi)}\}$. Note that $\gamma > \theta$, and hence $q_2 < \theta$. Also $\gamma \rightarrow 1$ as $(\varepsilon, \xi) \rightarrow (0, 0)$ and hence we can also have $q_2 \rightarrow 1$.

Consider first the beliefs of player 2. Let $\mu_2(\cdot)$ denote the probability assigned by 2 to the event that 1's continuation strategy is σ_C . Since 1 plays σ_C at Dc and σ_D at Dd , and since 1 does not play C in the asymmetric phase, we have

$$\mu_2(Cd) = \frac{1 - 2\varepsilon - \xi}{1 - \varepsilon - \xi} \quad (16)$$

Since we have assumed that $p < \frac{1-2\varepsilon-\xi}{1-\varepsilon-\xi}$, it is optimal to continue with σ_C today. Further, we have

$$\chi_{Cd}(\mu_2(Cd)) = \frac{(1 - 2\varepsilon - \xi)\varepsilon}{(1 - 2\varepsilon - \xi)(\varepsilon + \xi) + (1 - \varepsilon - \xi)\varepsilon} < \frac{1}{2} \quad (17)$$

Hence it is optimal for 2 to switch to the defection phase if he receives the signal Cd at any date in the future.

At Dd , we have

$$\mu_2(Dd) = \frac{\varepsilon}{1 - \varepsilon - \xi} \quad (18)$$

This is clearly less than p if $(\varepsilon + \xi) < \frac{1}{3}$, so that it is optimal to continue with σ_D .

Consider now the beliefs of player 2 at (Cc) and (Dc) , i.e. at the information sets where 2 knows that there has been at least one error in the signals.

$$\mu_2(Dc) = \frac{\xi}{\xi + \varepsilon} \quad (19)$$

$$\mu_2(Cc) = \frac{\varepsilon}{\xi + \varepsilon} \quad (20)$$

Recall that 2 plays σ_D at least at one of these information sets, since the above probabilities cannot be both greater than p , since this is greater than one-half. Hence there are three possibilities: either both $\mu_2(Dc)$ and $\mu_2(Cc)$ are less than p , or exactly one of these is greater than p . Now if $\mu_2(\cdot) < p$ at

any information set, it is optimal to continue with σ_D today, and at every future date. Hence it remains to verify the case when $\mu_2(\cdot) \geq p$.

Suppose that $\frac{\xi}{\xi+\varepsilon} > p$, so that 2 plays σ_C at Dc . If $\frac{\xi}{\xi+\varepsilon} \leq \theta$, lemma 1 verifies that it is optimal to continue with σ_C in this case. Hence consider the case where $\frac{\xi}{\xi+\varepsilon} > \theta$. We have that $\mu > \theta \Rightarrow \chi_{Cc}(\mu) < \mu$. Further, since χ_{Cd} is an increasing function, it suffices to verify that $\chi_{Cd}(\frac{\xi}{\xi+\varepsilon}) < p$, since this implies that $\chi_{Cd}(\mu) < p$ for $\mu = \chi_{Cc}^k(\frac{\xi}{\xi+\varepsilon})$ for any k .

$$\chi_{Cd}\left(\frac{\xi}{\varepsilon + \xi}\right) = \frac{\varepsilon\xi}{\xi(\varepsilon + \xi) + \varepsilon(1 - \varepsilon - \xi)} \quad (21)$$

This is less than $\frac{1}{2}$ if $\varepsilon + \xi$ is less than $\frac{1}{3}$. Hence player 2's continuation strategy is optimal at Dc .

Finally, we consider the case where that 2 plays σ_C at Cc , i.e. when $\frac{\varepsilon}{\varepsilon+\xi} > p$. Note that in this case A1 is violated. Hence we assume A2, which ensures that we can make p arbitrarily close to its upper bound $\frac{1-2\varepsilon-\xi}{1-\varepsilon-\xi}$, by selecting δ sufficiently close to $\frac{g}{1+g}$. We can find a value of p such that $\chi_{Cd}(\frac{\varepsilon}{\varepsilon+\xi}) < p$ provided that $\chi_{Cd}(\frac{\varepsilon}{\varepsilon+\xi})$ is less than the upper bound for p , i.e.

$$\chi_{Cd}\left(\frac{\varepsilon}{\varepsilon + \xi}\right) = \frac{\varepsilon^2}{\varepsilon^2 + \xi - \xi^2} < \frac{1 - 2\varepsilon - \xi}{1 - \varepsilon - \xi} \quad (22)$$

It is easily verified that the inequality above is ensured by condition A2.

Consider now the beliefs of player 1. His beliefs will depend upon player 2's strategy, which in turn depends upon the parameters of the signal distribution, and as we have seen, there are three possible cases.

Consider first the case where 2 plays σ_C only at information set Cd .

$$\mu_1(Dc) = \frac{q_2(1 - 2\varepsilon - \xi)}{q_2(1 - \varepsilon - \xi) + (1 - q_2)(\varepsilon + \xi)} \quad (23)$$

Note that the expression is such that $\mu_1(Dc) = \chi_{Cc}(q_2)$, where χ_{Cc} is the belief revision operator defined in the previous section. Hence it follows that if $q_2 \in [p(\delta, \varepsilon, \xi), \theta)$, it follows that $\chi_{Cc}^k(q_2) \in (p, \theta)$, $\forall k$, and hence it is optimal for player 1 to continue with σ_C at every information set.

Consider 1's beliefs at (Dd) . Once again, it is easy to verify that $\mu_1(Dd) = \chi_{Cd}(q_2)$, and since $q_2 < \theta$, it is optimal to continue with σ_D at this information set.

Consider next the case where $\sigma_2(Cd) = \sigma_2(Cc) = \sigma_C$ and $\sigma_2(Dd) = \sigma_2(Dc) = \sigma_D$. In this case, assumption A2 applies, so that we may choose p close to its upper bound. We have

$$\mu_1(Dc) = \frac{q_2(1 - \varepsilon - \xi)}{q_2(1 - \varepsilon - \xi) + (1 - q_2)(\varepsilon + \xi)} \quad (24)$$

If $q_2 > p$, then $\mu_1(Dc) > p$ so that it is optimal to start by playing σ_C in this case. To see that player 1 will find it optimal to switch to σ_D on receiving a bad signal, note that requires

$$\chi_{Cd}(\mu_1(Dc)) = \frac{q_2\varepsilon}{\varepsilon + \xi} < p \quad (25)$$

which yields the condition

$$q_2 < \frac{p(\varepsilon + \xi)}{\varepsilon} \quad (26)$$

Now if $\mu_1(Dc) \leq \theta$, lemma 1 has verified that σ_C is consistent in this case, i.e. the player will switch to σ_D on receiving signal Cd in any subsequent period. If $\mu_1(Dc) > \theta$, it suffices to verify that $\chi_{Cd}(\mu_1(Dc)) < \frac{1-2\varepsilon-\xi}{1-\varepsilon-\xi}$, which is the upper bound for p . This yields the condition

$$q_2 < \frac{(\varepsilon + \xi)(1 - 2\varepsilon - \xi)}{\varepsilon(1 - \varepsilon - \xi)} \quad (27)$$

The right hand side in (27) is precisely one of the arguments used in the definition of $\gamma(\varepsilon, \xi)$, and hence this condition poses no problem.

Finally, we consider the case where $\sigma_2(Cd) = \sigma_2(Dc) = \sigma_C$ and $\sigma_2(Dd) = \sigma_2(Cc) = \sigma_D$.

$$\mu_1(Dc) = \frac{q_2(1 - 2\varepsilon - \xi) + (1 - q)\xi}{q_2(1 - \varepsilon - \xi) + (1 - q_2)(\varepsilon + \xi)} < \frac{1 - 2\varepsilon - \xi}{1 - \varepsilon - \xi} \quad (28)$$

Hence it suffices to evaluate χ_{Cd} at the upper bound, which yields

$$\chi_{Cd}\left(\frac{1 - 2\varepsilon - \xi}{1 - \varepsilon - \xi}\right) = \frac{(1 - 2\varepsilon - \xi)\varepsilon}{(1 - 2\varepsilon - \xi)(\varepsilon + \xi) + (1 - \varepsilon - \xi)\varepsilon} < \frac{1}{2} \quad (29)$$

Hence $\chi_{Cd}(\mu_1(Dc)) < \frac{1}{2}$ for every value of q_2 .

We have therefore proved that the payoff $(\frac{1+g+l}{1+l}, 0)$ (and obviously the payoff $(0, \frac{1+g+l}{1+l})$) can be approximated under assumption A provided that

$\delta > \max\{\frac{g}{1+g}, \frac{l}{1+l}\}$ and provided that ε and ξ are sufficiently small. The payoff $(1, 1)$ has been approximated under a weaker set of assumptions ($\delta > \frac{g}{1+g}$) and ε and ξ sufficiently small), and the payoff $(0, 0)$ is a static Nash payoff. Since any payoff individually rational feasible payoff is a convex combination of these payoffs, and can be achieved via public randomization, we have proved the following theorem.

Theorem 6 *Assume that A is satisfied, then for any individually rational feasible payoff vector $u = (u_1, u_2)$ and any number $\zeta > 0$, there exist $\varepsilon(\zeta) > 0$, $\xi(\zeta) > 0$ such that there exists a sequential equilibrium with payoffs within ζ distance of u provided that $\varepsilon < \varepsilon(\zeta)$ and $\xi < \xi(\zeta)$ and $\delta > \max\{\frac{g}{1+g}, \frac{l}{1+l}\}$.*

Note that this result does not require vanishing discounting.

4 Concluding Comments

The main point of this paper has been to develop “belief-based” strategies as a way of constructing sequential equilibria in repeated games with private monitoring. This affords a major simplification as compared to the traditional method of analysis. While our construction has been restricted to the prisoners’ dilemma, and to a strategy profile which consists only of two continuation strategies, the idea underlying this simplification is obviously generalizable.

Our substantive results are most closely related to those in a recent paper by Piccione [9], who also analyzes the prisoners’ dilemma with imperfect private monitoring. Our results differ, both in terms of substance and in the techniques/strategies used. Piccione’s substantive results are that full cooperation can always be approximated, and further, any individually rational feasible payoff can be approximated in a class of prisoners’ dilemma games, i.e. for games where $l \geq g$. These approximations require that both the noise in monitoring and discounting vanish. Our results do not require vanishing discounting, and our “folk theorem” condition A, is in a sense, the opposite of Piccione’s condition. As we have already mentioned, the two papers use very different techniques.

References

- [1] V. Bhaskar, The Robustness of Repeated Game Equilibria to Incomplete Payoff Information, mimeo, 1999.
- [2] V. Bhaskar, Informational Constraints and the Overlapping Generations Model: Folk and Anti-Folk Theorems, *Review of Economic Studies*, 65 (1998) , 135-149.
- [3] Bhaskar, V., and E. van Damme, Moral Hazard and Private Monitoring, mimeo, 1997.
- [4] O. Compte, Communication in repeated games with imperfect private monitoring, *Econometrica*, 66 (1998),597-626.
- [5] D. Fudenberg, D. Levine and E. Maskin, The Folk Theorem with Imperfect Public Information, *Econometrica*, 62 (1994), 997-1040.
- [6] J. Harsanyi, Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-Strategy Equilibrium Points, *International Journal of Game Theory* 2 (1973), 1-23.
- [7] M. Kandori, and H. Matsushima, Private Observation, Communication and Collusion, *Econometrica*, 66 (1998), 627-652.
- [8] G. Mailath, and S. Morris, Repeated Games with Imperfect Private Monitoring: Notes on a Coordination Perspective, mimeo, 1997.
- [9] M. Piccione, The repeated prisoners' dilemma with imperfect private monitoring, mimeo, 1998.
- [10] T. Sekiguchi, Efficiency in the Prisoners' Dilemma with Private Monitoring, *Journal of Economic Theory*,76 (1997), 345-361.