Delayed Perfect Monitoring in Repeated Games

Kinateder, Markus
Universidad de Navarra

December 2009

Delayed Perfect Monitoring in Repeated Games

Markus Kinateder*
Universidad de Navarra†

4 December 2009

Abstract

Delayed perfect monitoring in an infinitely repeated discounted game is studied. A player perfectly observes any other player's action choice with a fixed, but finite delay. The observational delays between different pairs of players are heterogeneous and asymmetric. The Folk Theorem extends to this setup, although for a range of discount factors strictly below 1, the set of belief-free equilibria is reduced under certain conditions. This model applies to any situation in which there is a heterogeneous delay between information generation and the players' reaction to it.

JEL classification numbers: C72, C73
Keywords: Repeated Game, Delayed Perfect Monitoring, Folk Theorem

1 Introduction

Infinitely repeated discounted games capture dynamic strategic interaction between impatient economic agents. Additional equilibria arise compared to one-shot games and

*An earlier version of this paper was titled "Repeated Games Played in a Network". I am very grateful for the support received from my supervisor Jordi Massó. Additionally, I thank Toni Calvó-Armengol, Julio González-Díaz and Penélope Hernández for their advice and time. I benefited hugely from conversations with Elchanan Ben-Porath, Yann Bramoullé, Drew Fudenberg, Olivier Gossner, David Levine, Filippos Louis, Anna Papaccio, Parag Pathak, David Rahman, Jérôme Renault, Ariel Rubinstein, Rann Smorodinsky, Tristan Tomala, Marco van der Leij and Fernando Vega-Redondo, and from comments made by participants of this model's presentation in Valencia, Girona, Oviedo, Mannheim, Frankfurt, Aarhus, Palma, Pamplona, Vienna, at the RES Conference in Warwick, at the ESEM Meeting in Budapest, at the GAMES Conference in Chicago and at Universitat Autònoma de Barcelona (UAB). This paper forms part of my PhD thesis defended at UAB in September 2008. I thank the committee members for their generous advice.

†Departamento de Economía, Edificio de Biblioteca (Entrada Este), Universidad de Navarra, 31080 Pamplona, Spain; email: mkinateder@unav.es

the associated payoff vectors can be Pareto superior to those achieved in any stage game equilibrium. The well-known Folk Theorem states this result. For infinitely repeated discounted games, it is obtained by Fudenberg, Levine and Takahashi (2007), thereafter FLT. Frequently, a player is assumed to observe his opponents' behavior immediately and perfectly, referred to as perfect monitoring. This assumption is relaxed in the imperfect monitoring literature, in which each player receives an imperfect private or public signal of every action profile played.[1]

In this paper, monitoring is delayed since each player obtains a private signal about the action chosen by another player with a fixed, but finite delay. These signals are perfect, and thus, a repeated game with delayed perfect monitoring is studied. Formally, for each pair of players that participate in an infinitely repeated discounted game there exists a delay with which they observe each other's action choice—this delay might be asymmetric and is allowed to be heterogeneous for different pairs of players. In each period, a player observes the actions chosen by a subset of players, including himself, at different points of time in the past. The players take decisions under imperfect information in any but the first period. However, since players do not take into account beliefs about unobserved action choices in the past the concept of belief-free equilibrium, a sequential equilibrium with a simple belief system, is used.

The Folk Theorem extends to the delayed perfect monitoring model, that is, any feasible and strictly individually rational payoff vector is supported by a belief-free equilibrium strategy profile when the players are sufficiently patient. Then, they do not mind to receive the repeated game's history of action profiles gradually over time. However, for a range of discount factors strictly below 1, the delay in obtaining information, under certain conditions, triggers a player's deviation from some previously agreed sequence of play. In this setup, for impatient players, the set of belief-free equilibria is reduced in comparison to the perfect monitoring case under certain conditions.

The related literature considers different setups. In one, all players play the same repeated game and a player observes an imperfect private or public signal of each action profile (see footnote 1). Other models of imperfect monitoring are surveyed in Mailath and Samuelson (2006).

The next section introduces notation and definitions. In section 3, the model is illustrated for the Prisoner's Dilemma. In section 4, information spreading and punishment reward are defined. Both are prerequisites for the Folk Theorem, which is stated in section 5, along with conditions under which impatient players deviate from a given sequence of

[1] Fudenberg, Levine and Maskin (1994), for example, obtain a Folk Theorem under imperfect public monitoring, and Kandori (2002) surveys the imperfect private monitoring literature.

action profiles. Moreover, a comparative static result is provided. The model is presented in unobservable mixed actions. Before concluding, possible extensions are discussed.

2 Preliminaries

2.1 Stage Game and Observation Structure

Each player i in the finite set of players $I = \{1, ..., n\}$ has a finite set of pure actions A_i. Pure action a_i is an element of this set. The stage game's pure action space is $A = \times_{i \in I} A_i$, with generic element a, called pure action profile. To emphasize player i's role, it is written as (a_i, a_{-i}). For any subset of players $S \subset I$, let $A_S = \times_{i \in S} A_i$, and denote by a_S an element of this set. Player i's payoff function is a mapping $h_i : A \to \mathbb{R}$, and the payoff function $h : A \to \mathbb{R}^n$ assigns a payoff vector to each pure action profile. The stage game in normal form is then the tuple $G \equiv (I, (A_i)_{i \in I}, (h_i)_{i \in I})$. Define the convex hull of the finite set of payoff vectors corresponding to pure action profiles in G as $co(G) = co\{x \in \mathbb{R}^n \mid \exists\ a \in A : h(a) = x\}$. Define the mixed extension of G by $G^* \equiv (I, (\Sigma_i)_{i \in I}, (H_i)_{i \in I})$, where $\Sigma_i = \{\sigma_i : A_i \to [0, 1] \mid \sum_{a_i \in A_i} \sigma_i(a_i) = 1\}$ is player i's mixed action space and $H_i : \Sigma \to \mathbb{R}$ his payoff function for $\Sigma = \times_{i \in I} \Sigma_i$. Let $\sigma \in \Sigma$ be a mixed action profile. To emphasize player i's role, it is written as (σ_i, σ_{-i}). The function $H : \Sigma \to \mathbb{R}^n$ assigns a payoff vector to each mixed action profile. Note that a mixed action consists of a player's randomization experiment and the pure action he chooses. It is assumed that the randomization experiment is not observable, but only the pure action chosen. This is referred to as unobservable mixed actions.

Denote the delay with which player i observes player j's action choice by d_{ij}. It is a finite positive integer for all $i, j \in I$. The maximal delay between player i and any other player is defined by $d_i = \max_{j \in I} d_{ij}$, and the maximal delay between any pair of players is defined as $d = \max_{i \in I} d_i$. For each player i, partition the set of players with respect to the delay with which i observes their action choices: all players he immediately observes including himself are in $i(1) = \{j \in I \mid d_{ij} = 1\}$, and for any $2 \leq m \leq d_i$, define $i(m) = \{j \in I \mid d_{ij} = m\}$. Each of these sets might be empty, except of $i(d_i)$, by definition, and of $i(1)$ since it contains at least i. Denote this *observation structure* by OS. It can be represented in an $n \times n$ matrix: the ijth entry specifies the delay with which player i observes player j's action choice. This matrix need not be symmetric, that is, for any $i \neq j$, d_{ij} need not coincide with d_{ji}.

When the stage game is played repeatedly, in each period, a player first chooses an action, in a way specified below, and then makes observations. Since d_i is player i's maximal delay, with a lag of $d_i - 1$ periods, he observes the repeated game's entire history.[2] Additionally, a player has perfect recall. Hence, for any player $i \in I$ at any time period $t \geq 1$, there is a *set of observations*, denoted by Ob_i^t, that includes all histories of observations that i may have made at the end of period t. It is defined recursively as

$$\begin{aligned} Ob_i^1 &= A_{i(1)}, \\ Ob_i^2 &= A_{i(1)}^2 \times A_{i(2)}, \\ &\vdots \\ Ob_i^t &= A_{i(1)}^t \times A_{i(2)}^{t-1} \times \cdots \times A_{i(d_i)}^{t-d_i+1} \end{aligned}$$

for all $t \geq d_i$, where for any $1 \leq m \leq d_i$ and any $t \geq 1$, $A_{i(m)}^t = (\times_{j \in i(m)} A_j)^t$. Note that $A_{i(m)}^t = \emptyset$ if, and only if, $i(m) = \emptyset$, and that, by definition, only pure actions are observable.

Player i's observation at t is denoted by $ob_i^t \in Ob_i^t$. Given G^*, a sequence of mixed action profiles $\{\sigma^t\}_{t=1}^{\infty}$, where $\sigma^t \in \Sigma$ for all $t \geq 1$, generates a sequence of observations for player i,

$$\begin{aligned} ob_i^1 &= (a_i^1, a_{i(1)}^1), \\ ob_i^2 &= (a_i^1, a_{i(1)}^1, a_{i(2)}^1, a_i^2, a_{i(1)}^2), \\ &\vdots \\ ob_i^t &= (\{a_i^s\}_{s=1}^t, \{a_{i(1)}^s\}_{s=1}^t, \{a_{i(2)}^s\}_{s=1}^{t-1}, ..., \{a_{i(d_i)}^s\}_{s=1}^{t-d_i+1}) \end{aligned}$$

for all $t \geq d_i$. At any $t < d_i$, player i did not yet observe the behavior of at least one other player in period 1. At $t = d_i$, $ob_i^{d_i}$ contains the actions chosen by all players at $t = 1$.[3] Abusing notation, this is referred to as $a^1 \in ob_i^{d_i}$ (since a^1 belongs to A). At any $t > d_i$, action profiles $a^1, ..., a^{t-d_i+1}$ are identified by player i, and hence, in an abuse of terminology, said to be elements of ob_i^t. Thus, at any $t \geq 1$, the sequence of mixed action profiles generates an observation profile $ob^t \in Ob^t$, where $Ob^t = \times_{i \in I} Ob_i^t$. Given an observation structure OS, the players play an infinitely repeated discounted game.

2.2 Repeated Game with Delayed Perfect Monitoring

In the infinitely repeated discounted game with delayed perfect monitoring, at each point in discrete time, $t = 1, 2, ...$, stage game G^* is played.

[2] At the end of any $t \geq d_i$, player i knows the actions played at t by all players in $i(1)$, those played by all players in $i(1)$ and $i(2)$ at $t-1, ...,$ and finally the ones played by all players at $t - d_i + 1$ and before.

[3] This setup is equivalent to the following: each mixed action profile σ^t generates a public signal with a delay of $d - 1$ periods and certain private signals in all periods s, where $t \leq s < t + d - 1$.

Let player i's set of behavior strategies be $F_i = \{\{f_i^t\}_{t=1}^{\infty} \mid f_i^1 \in \Sigma_i$, and for all $t > 1$, $f_i^t : Ob_i^{t-1} \to \Sigma_i\}$. At any $t \geq 1$, player i's behavior strategy $f_i = \{f_i^t\}_{t=1}^{\infty}$ prescribes him to choose a mixed action. For $t > 1$, it maps his *set of observations* to his mixed action set. Let $F = \times_{i \in I} F_i$ be the behavior strategy space of the repeated game with delayed perfect monitoring and let behavior strategy profile $f = (f_1, ..., f_n)$ be an element of F. To emphasize player i's role, it is written as (f_i, f_{-i}). At any $t \geq 1$, each $f \in F$ recursively generates an action profile $\sigma^t(f) = (\sigma_1^t(f), ..., \sigma_n^t(f))$ and a corresponding observation profile $ob^t(f) = (ob_1^t(f), ..., ob_n^t(f))$.[4] Each $f \in F$ thus generates a sequence of action profiles $\{\sigma^t(f)\}_{t=1}^{\infty}$ and a sequence of observation profiles $\{ob^t(f)\}_{t=1}^{\infty}$.

Given a common discount factor $\delta \in [0, 1)$,[5] the function $H^{\delta} : F \to \mathbb{R}^n$ assigns a payoff vector to each behavior strategy profile. Given $f \in F$, player i's payoff, $H_i^{\delta}(f) = (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} H_i(\sigma^t(f))$, is the $(1-\delta)$-normalized discounted sum of stage game payoffs. The repeated game with delayed perfect monitoring associated with stage game G^*, discount factor δ and observation structure OS is then defined as the normal form game $G^{OS,\delta} \equiv (I, (F_i)_{i \in I}, (H_i^{\delta})_{i \in I})$, where the star superscript is suppressed.

If $i(1) = I$ for all $i \in I$, then $G^{OS,\delta}$ is identical to the infinitely repeated discounted game, referred to as G^{δ}. In this case f_i simplifies: at any $t > 1$ it maps $A^{t-1} = (\times_{i \in I} A_i)^{t-1}$ to Σ_i, that is, each player conditions his action choice on the history of observable action profiles chosen by all players between periods 1 and $t - 1$.

Finally, the players commonly know the game played, the observation structure and the strategy choices available to all players, and are assumed to observe their payoff with a delay of d periods.[6]

2.3 Payoff vectors generated by Belief-free Equilibria

A player's individually rational payoff is the lowest to which he can be forced in a stage game. It obtains when he maximizes his payoff while all other players minimize it and is called *minmax* payoff. For any $i \in I$, define his *minmax* payoff in mixed actions by

$$\nu_i \equiv \min_{\sigma_{-i} \in \Sigma_{-i}} \max_{a_i \in A_i} H_i(a_i, \sigma_{-i}). \tag{1}$$

[4] For any player i, let $\sigma_i^1(f) = f_i^1$ and $ob_i^1(f) = (a_i^1(f), a_{i(1)}^1(f))$, and for $t > 1$, given $ob_i^{t-1}(f) \in Ob_i^{t-1}$, $\sigma_i^t(f) = f_i^t(ob_i^{t-1}(f))$ and $ob_i^t(f)$ is defined accordingly. If the prescribed mixed action at t is degenerate, player i is asked to choose a pure action and this, abusing notation, is referred to as $a_i^t(f) = f_i^t(ob_i^{t-1}(f))$.

[5] It may be interpreted as the probability with which the game is played again in the next period. The probability that the repeated game ended by period T then converges to 1 as T goes to infinity.

[6] After $d_i - 1$ periods, player i observed the action profiles played between periods 1 and $t - d_i + 1$, and can calculate or equivalently observe his payoff for all these periods; and $d = \max_{i \in I} d_i$.

The *minmax* payoff is a player's individually rational payoff in any repeated game, in which the dimension of the payoff space is equal to the number of players.[7] Denote the vector of *minmax* payoffs in mixed actions by ν, and the mixed action profile forcing player i to his *minmax* payoff by $\bar{\sigma}^i$. It is one solution to the optimization problem on the right-hand-side of (1), on which the players agreed. Without loss of generality any player's *minmax* payoff is normalized to 0, that is, for all $i \in I$, $H_i(\bar{\sigma}^i) \equiv 0$.

The *set of feasible payoff vectors* of the repeated game with delayed perfect monitoring is defined as[8]

$$\mathcal{F} = \{x \in \mathbb{R}^n \mid \exists \; \{a^t\}_{t=1}^{\infty} : \; \forall \; t \geq 1, \; a^t \in A, \text{ and } \forall \; i \in I, \; x_i = (1-\delta) \sum_{t=1}^{\infty} \delta^{t-1} h_i(a^t)\}.$$

Any feasible payoff vector is achievable by a sequence of pure action profiles. Mixed actions need not be used, apart from the *minmax* punishment of a deviator.

The *set of feasible and strictly individually rational payoff vectors* is denoted by $\mathcal{F}^*$. It contains all feasible payoff vectors that are larger than $\nu = (0, ..., 0)$ and is defined as

$$\mathcal{F}^* \;\; = \;\; \{x \in \mathcal{F} \mid x > \nu\}.$$

Any payoff vector in this set is a candidate to be supported by a belief-free equilibrium.

In a belief-free equilibrium, each player conditions his action choices only on his observations and a strategy profile is sequentially rational for any consistent belief a player may have about the yet unobserved actions chosen by all other players (in the most recent periods).[9] Hence, beliefs are not modelled formally.

Definition 1. *A behavior strategy profile $f^* \in F$ is a* belief-free equilibrium (BFE) *of $G^{OS,\delta}$, if for all $t \geq 1$ and given any $ob^t \in Ob^t$, $\{f^{*\tau}(ob^{\tau-1})\}_{\tau=t+1}^{\infty}$ is such that for all $i \in I$ and all $f_i \in F_i$,*

$$(1-\delta) \sum_{s=t+1}^{\infty} \delta^{s-1} H_i(\sigma^s(f^*)) \;\; \geq \;\; (1-\delta) \sum_{s=t+1}^{\infty} \delta^{s-1} H_i(\sigma^s(f_i, f^*_{-i})).$$

When $i(1) = I$ for all i, then this definition includes G^{δ} and the concepts of belief-free and subgame-perfect equilibrium coincide. However, equilibria of $G^{OS,\delta}$ and G^{δ} are called belief-free when Definition 1 is satisfied, and the corresponding sets of BFE strategy

[7] The repeated game with delayed perfect monitoring extends to stage games with less than full-dimensional payoff space as is remarked in the conclusion.

[8] Any payoff vector in $co(G)$ is feasible for $\delta \in (1-\frac{1}{z}, 1)$, where z is the number of vertices of $co(G)$. For any discount factor in this range, sets $\mathcal{F}$ and $co(G)$ coincide; see Fudenberg, Levine and Maskin (1994).

[9] A player's belief for all observed action choices is uniquely determined. His strategy is only conditioned on observed actions, while his belief about unobserved actions is irrelevant for his choices.

profiles are denoted by $BFE(G^{OS,\delta})$ and $BFE(G^{\delta})$, respectively. A behavior strategy profile is a BFE if, and only if, no player's finite unilateral deviation is profitable at any point in time.[10]

3 The Observation Structure makes a difference

The following example illustrates how imposing an observation structure on a repeated game affects its set of BFE. Let $\hat{G} = (I, A, h)$ be a generalized Prisoner's Dilemma game, where $n > 2$. At each point in time, a player chooses either C (*cooperate*) or D (*defect*). The payoff function of any player $i \in I$ is defined as follows: for each $a \in A$,

$$h_i(a) = \begin{cases} 3 & \text{if } a_j = C, \ \forall\, j \in I \\ 0 & \text{if } a_i = C \text{ and } \exists\, j \in I \setminus \{i\} \text{ s.t. } a_j = D \\ 4 & \text{if } a_i = D \text{ and } a_j = C, \ \forall\, j \in I \setminus \{i\} \\ 2 & \text{if } a_i = D, \ \exists\, j \in I \setminus \{i\} \text{ s.t. } a_j = D \text{ and } \exists\, l \in I \setminus \{i, j\} \text{ s.t. } a_l = C \\ 1 & \text{if } a_j = D, \ \forall\, j \in I. \end{cases}$$

In the unique Nash Equilibrium of stage game $\hat{G}$ all players choose D, since it is a strictly dominant action. In the repeated Prisoner's Dilemma, strategy profiles that yield all players a higher payoff are sustained as BFE under certain conditions, such as the trigger strategy profile. It prescribes each player to cooperate as long as all players cooperate and to defect forever if any player defected. Player i's trigger strategy, denoted by $\hat{f}_i \in F_i$, is defined as follows: $\hat{f}_i^1 = C$, and for $t \geq 1$, given $ob_i^t \in Ob_i^t$,

$$\hat{f}_i^{t+1}(ob_i^t) = \begin{cases} D & \text{if } \exists\, 1 \leq \tau \leq t \text{ such that for } a^{\tau} \in ob_i^t, \ a_j^{\tau} = D, \text{ while } a_{-j}^{\tau} = C \\ C & \text{otherwise.} \end{cases}$$

Given $\hat{f} \in F$, observe that for all $i \in I$ and all $t \geq 1$, first $a_i^t(\hat{f}) = C$, and second, $ob_i^t(\hat{f})$ is such that for all $a_j^{\tau} \in ob_i^t(\hat{f})$, $a_j^{\tau} = C$ as well for all $1 \leq \tau \leq t$ and all $j \in I$. Hence, for all $i \in I$, $H_i^{\delta}(\hat{f}) = (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} 3 = 3$.

3.1 A one-period delay between two players

Consider a generalized Prisoner's Dilemma game with $n = 3$, as represented in Figure 1, where player 1 chooses rows, player 2 columns and player 3 matrices. Let the following

[10] Since $\delta < 1$, a player's gain from an infinite deviation can be approximated by that of a finite one. Thus, unilateral deviations of finite length from a behavior strategy profile are not profitable if, and only if, it is a BFE of the repeated game with delayed perfect monitoring; see Mailath and Samuelson (2006).

3

C

1-2	C	D
C	3, 3, 3	0, 4, 0
D	4, 0, 0	2, 2, 0

D

1-2	C	D
C	0, 0, 4	0, 2, 2
D	2, 0, 2	1, 1, 1

Figure 1: Prisoner's Dilemma for three players

symmetric observation structure OS be given: player 2 observes players 1 and 3, and both of them player 2 perfectly. However, players 1 and 3 observe each other's action choice with a delay of one period. The trigger strategy profile is a BFE of $\hat{G}^{OS,\delta}$ if, and only if, all players are patient enough, that is, δ is higher than some threshold value. Then, none of them ever deviates. Corresponding conditions on δ must hold for the truncation of the repeated Prisoner's Dilemma with delayed perfect monitoring at any point in time, that is, given any observation profile. A BFE does not impose restrictions on play after a multilateral deviation by two or more players. Any unilateral deviation that may arise can be uniquely allocated to one of the following three classes:

1) initial unilateral deviations,
2) subsequent unilateral deviations (before the initial is known by all players), and
3) unilateral deviations while the punishment takes place.

Obviously, unilateral deviations during the punishment are not profitable since all players choose D. This action profile is the stage game Nash Equilibrium in strictly dominant actions. Hence, every player best-replies independently of δ. For the same reason, no player can deviate profitably from the trigger strategy profile in class 2. After a player's initial deviation, he and any player who knows about it are best-off to play D forever (rather than to deviate and to choose C at any point in time).

It remains to show that no player can profitably deviate from the trigger strategy profile when all players should play C. Given δ, player 2 (who is perfectly observed by 1 and 3) does not deviate in any period τ if, and only if,

$$(1-\delta)\sum_{t=1}^{\infty} 3\delta^{t-1} \geq (1-\delta)\sum_{t=1}^{\tau-1} 3\delta^{t-1} + 4(1-\delta)\delta^{\tau-1} + (1-\delta)\sum_{t=\tau+1}^{\infty} 1\delta^{t-1},$$

$$(1-\delta)\sum_{t=\tau+1}^{\infty} 2\delta^{t-1} \geq (1-\delta)\delta^{\tau-1},$$

$$2\delta^{\tau+1} \geq (1-\delta)\delta^{\tau},$$

$$\delta \geq \tfrac{1}{3}.$$

The value of $\frac{1}{3}$ is not only the threshold value for player 2 in this example but also that for all players in a repeated Prisoner's Dilemma with perfect monitoring. The observation structure affects, however, the threshold value of the remaining two players in this example. Given δ, player 1 (and similarly 3) does not deviate from the trigger strategy profile in any period τ if, and only if,

$$(1-\delta)\sum_{t=1}^{\infty} 3\delta^{t-1} \geq (1-\delta)\sum_{t=1}^{\tau-1} 3\delta^{t-1} + 4(1-\delta)\delta^{\tau-1} + 2(1-\delta)\delta^{\tau} + (1-\delta)\sum_{t=\tau+2}^{\infty} 1\delta^{t-1},$$

$$(1-\delta)\delta^{\tau} + (1-\delta)\sum_{t=\tau+2}^{\infty} 2\delta^{t-1} \geq (1-\delta)\delta^{\tau-1},$$

which can be simplified to $2\delta + \delta^2 - 1 \geq 0$. The only positive solution to this quadratic equation is $\delta \approx 0.414$. Hence, in class 1 of the BFE conditions the requirement on δ, or the players' patience, is higher in this example than in a perfect monitoring model, due to the one period lag with which players 1 and 3 observe each other's action choice.

This example extends to any set of players where $n > 3$ as long as every player is observed by at least one other player immediately.

3.2 The Prisoner's Dilemma with any Observation Structure

A similar result holds for any observation structure in the repeated Prisoner's Dilemma in which all players follow the trigger strategy and every player is observed by at least one other player immediately. In the above example it takes 2 periods until full punishment sets in. Given any observation structure, it takes d_i periods until all other players punish player i. Until then the deviator's payoff is 2 since at least one player still chooses C. Thereafter, it is 1 forever.

Since d is the maximal delay between any pair of players, there is a discount factor δ^* that solves $2\delta^* + \delta^{*d} - 1 \geq 0$ such that no player deviates from the trigger strategy profile. Hence, for this strategy profile all repeated Prisoner's Dilemma games can be classified according to their observation structure. The threshold value of the discount factor δ^*, for which no player deviates from the trigger strategy profile, that is, the level of patience required to sustain cooperation is non-decreasing in d; since a higher delay implies that at least one pair of players observes each other after a larger time lag.

Although the expression $2\delta^* + \delta^{*d} - 1 \geq 0$ depends on d, even for very large values of d the threshold value for δ^* is bounded above by $\frac{1}{2}$. To see this, take the limit of the inequality when d converges to infinity. Since $\delta < 1$, the term δ^{*d} converges to 0 and the inequality simplifies to $2\delta^* - 1 \geq 0$ or $\delta^* \geq \frac{1}{2}$. Hence, for "moderately patient" players, the trigger strategy profile is a BFE in any repeated Prisoner's Dilemma with delayed perfect monitoring as long as every player is observed by at least one other player immediately.

The observation structure may thus reduce the set of discount factors for which a strategy profile is a BFE. Moreover, for a given discount factor, the set of BFE strategy profiles and the corresponding set of payoff vectors may be strictly smaller in the repeated game with delayed perfect monitoring than in the version with perfect monitoring.[11]

4 Information Spreading and Punishment Reward

The general conditions for a BFE are not as simple as in the previous section since the *minmax* action profile in most stage games is no Nash Equilibrium in strictly dominant and pure actions. Hence, punishment is asymmetric and costly at least for some players. The part of the Folk Theorem behavior strategy profile after a deviation is outlined next.

Until all players know about a deviation, they follow the originally prescribed sequence of action profiles. While in the Prisoner's Dilemma for the trigger strategy profile all players punish player i from d_i periods after his deviation on, in general, all players start to punish simultaneously any unilateral deviator after d periods. Only then the deviation is commonly known. The phase during which the information about a deviation spreads throughout the set of players is called Information Spreading Process (ISP). Note that the ISP-payoff is not normalized by $(1 - \delta)$.

Definition 2. *Given $f \in F$, the* Information Spreading Process *payoff of player i following an initial deviation in period t' only is defined as*

$$ISP_i^{t'} \; = \; H_i(\sigma^{t'+1}(f)) \; + \; ... \; + \; \delta^{d-2} H_i(\sigma^{t'+d-1}(f)).$$

The ISP extends easily to any player's deviation of finite length. Any subsequent deviator starts a new ISP which may overlap with the ongoing one. Once every player identified the last deviator, he is forced to his *minmax* payoff at least until his entire gain from deviating is taken away or another subsequent deviator is punished. During

[11] The reduction in the equilibrium payoff space for $\delta \in [\frac{1}{3}, \frac{1}{2})$, for example, is the point $(3, 3, 3)$, since the trigger strategy profile is no BFE if at least one pair of players obtains information about each other with a delay, and no other BFE strategy profile supports this payoff vector.

punishment some players incur a loss in their payoff. Hence, punishment starts once all players know about the deviation and is restricted to a minimal amount of time. Thereafter, a punishment reward phase is played in order to induce the punishers to randomize over the pure actions in the support of the mixed *minmax* action and to reward them for their temporary payoff loss, obviously, without benefiting the deviator.

Given any feasible and strictly individually rational target payoff vector $x \in \mathcal{F}^*$, there are player-specific punishment reward payoff vectors denoted by $\omega^1, ..., \omega^n$. They are achieved by sequences of pure action profiles and have the following properties. For any player i, $x_i > \omega_i^i > 0$, and for two distinct players $i \neq j$, $\omega_i^i < \omega_i^j$, that is, the i-th component of vector i is strictly smaller than that of any other one. In this way the punishers are rewarded but not the punished player i.

In order to induce the players to randomize in their punishment against some deviator i, who deviated at t', the sequence of pure action profiles that yields ω^i depends on the realized action profiles during punishment. Formally, define by

$$dif_j^{i,t'} \equiv (1-\delta)[\sum_{t=t'+d}^{t'+d+\bar{T}} \delta^{t-t'-d}(h_j(a^t) - H_j(\bar{\sigma}^i))]$$

the difference between any player j's realized payoff during the punishment against player i and his expected payoff given the mixed action profile $\bar{\sigma}^i$ that yields ν_i, where $\bar{T}$ is the endogenously determined last time period of i's punishment which is a positive integer. At period $\bar{T}+1$ player i's punishment reward phase starts. Denote the sequence of pure action profiles that is played during this phase by $\{c^s\}_{s=1}^{\infty}$. It is determined together with a positive integer $\tilde{T}$ such that for every player $j \in I$,

$$\omega_j^i \quad = \quad (1-\delta)[\sum_{t=\bar{T}+1}^{\tilde{T}} \delta^{t-1} h_j(\tilde{c}^t) + \sum_{t=\tilde{T}+1}^{\infty} \delta^{t-1} h_j(c^t)] + dif_j^{i,t'}.$$

Intuitively, d periods after the end of player i's *minmax* punishment, the realizations of all mixed actions chosen by the players during punishment are commonly known. The action profiles played during the periods after i's punishment are made conditional on these realizations such that each player j receives exactly ω_j^i in the punishment reward phase. Hence, a player whose randomization made him obtain a lower payoff during the punishment phase than his expected payoff from action profile $\bar{\sigma}^i$ receives a compensation while a player whose payoff during this phase is larger than the expected one from $\bar{\sigma}^i$ receives a penalty. The existence of this conditional punishment reward phase for high discount factors and given any $x \in \mathcal{F}^*$ follows, for example, from FLT. They show that this compensation phase ends in finite time. Its last time period is denoted by $\tilde{T}$. From

$\tilde{T}+1$ on, the sequence of action profiles played depends only on the name of the deviator, but not on his punishment phase. Together with the one played from $\bar{T}+1$ until $\tilde{T}$ this yields $\{c^s\}_{s=1}^{\infty}$. In this way, all players are made indifferent between randomizing over the pure actions in the support of the mixed *minmax* action since their payoff is the same independently of the realized action, and they actually randomize, although deviations within the support of the mixed *minmax* action would not be observable.

5 The Results

A behavior strategy profile can be constructed for which, given any observation profile, no player's unilateral deviation is profitable, provided that the players are patient enough. It is a BFE of the repeated game with delayed perfect monitoring and a Folk Theorem obtains. The proof of the Folk Theorem is relegated to Appendix A. Its basic idea is in line with Abreu, Dutta and Smith (1994).

Theorem 1. *Let G^* and OS be given. Then, for all $x \in \mathcal{F}^*$, there is $\tilde{\delta} < 1$ such that for each $\delta \in (\tilde{\delta}, 1)$, there is a corresponding $\tilde{f} \in F$ such that $\tilde{f} \in BFE(G^{OS,\delta})$ and $H^{\delta}(\tilde{f}) = x$.*

Various sequences of pure action profiles yield the same payoff vector $x \in \mathcal{F}^*$. Behavior strategy profile $\tilde{f}$ gives the structure to support any of them. It prescribes the players to follow a given sequence of pure action profiles and to punish any unilateral deviator from d periods after his deviation on until his entire gain is taken away or some other player is punished. Thereafter, his punishment reward phase is played. Mixed actions are only used for punishment. Each observation profile that may arise belongs to one of a small number of classes of observation profiles. For each it is shown that no player can deviate profitably. The objective of the Folk Theorem is not to find the most efficient strategy profile and it obtains as well for other possibly more efficient strategy profiles.

Patient enough players do not mind to receive the repeated game's history gradually over time. That punishment is not immediate but sets in after a finite delay is strong enough a threat for them. In the limit, the effects of the delay in observations disappear and the same set of payoff vectors is generated by BFE in the repeated game and in its version with delayed perfect monitoring.

Corollary 1. *Let G^* and OS be given. Then, there is $\bar{\delta} < 1$ such that for all $\delta \in (\bar{\delta}, 1)$ and all $x \in \mathcal{F}^*$, there are $f \in BFE(G^{OS,\delta})$ and $\bar{f} \in BFE(G^{\delta})$ such that $\{a^t(f)\}_{t=1}^{\infty} \equiv \{a^t(\bar{f})\}_{t=1}^{\infty}$, and $H^{\delta}(f) = H^{\delta}(\bar{f}) = x$.*

For impatient players, or in other words, for a range of discount factors strictly below 1, the delay in observation makes a difference, as already shown for the Prisoner's Dilemma in section 3. A similar result can be derived for any stage game.

A lower bound of the discount factor $\underline{\delta}$ is identified such that for all $\delta \in [0, \underline{\delta}]$, only sequences of action profiles that prescribe the infinite repetition of stage game Nash Equilibria are supported by BFE in both games. Together with Corollary 1, the reduction in the set of sequences of action profiles that are supported by BFE in a repeated game with perfect monitoring but not with delayed perfect monitoring is then stated formally.

Corollary 2. *Let G^* and OS be given. Then, there are $0 < \underline{\delta} \leq \bar{\delta} < 1$ such that for all $\delta \in (\underline{\delta}, \bar{\delta}]$, $\{\{a^t(f)\}_{t=1}^{\infty} \mid f \in BFE(G^{OS,\delta})\} \subset \{\{a^t(\bar{f})\}_{t=1}^{\infty} \mid \bar{f} \in BFE(G^{\delta})\}$.*

For a range of intermediate discount factors, the observation structure reduces the set of sequences of action profiles that are generated by BFE strategy profiles. In special cases, the lower and upper bound of δ coincide and the corollary is trivially true.[12]

Finally, formal conditions are given under which an observation structure reduces the set of BFE strategy profiles for impatient players. Given G^*, OS and δ, assume that $\bar{f} \in BFE(G^{\delta})$ and let $\{\dot{a}^t\}_{t=1}^{\infty} \equiv \{a^t(\bar{f})\}_{t=1}^{\infty}$. Say that *the delay in observations has an impact with respect to $\tilde{f}$*, as defined in Theorem 1, if $\tilde{f}$ does not support $\{\dot{a}^t\}_{t=1}^{\infty}$ as a BFE of $G^{OS,\delta}$.[13] Suppose that player i gains

$$\beta_i^{\tau} \quad \equiv \quad \sum_{t=\tau}^{\tau+d-1} \delta^{t-\tau}[\max_{a_i \in A_i} h_i(a_i, \dot{a}_{-i}^t) - h_i(\dot{a}^t)]$$

by a deviation of length $d-1$ from $\{\dot{a}^t\}_{t=1}^{\infty}$ that starts at τ. Let

$$\lambda_i^{\tau}(T) \quad \equiv \quad \sum_{t=\tau+d}^{\infty} \delta^{t-\tau-1} h_i(\dot{a}^t) - (1-\delta)^{-1}\delta^T \omega_i^i$$

for $T \geq 2d-2$. It takes $d-1$ periods until all players know about i's deviation, and $2d-2$ periods after it, all of them know if i deviated again one period before his punishment started. Then, Proposition 1 identifies conditions under which *the delay in observations has an impact with respect to $\tilde{f}$*.

Proposition 1. *Let G^*, OS and $\delta < 1$ be given. Suppose there is $\bar{f} \in BFE(G^{\delta})$, $i \in I$ and $\tau \geq 1$, such that for all positive integers $T \geq 2d-2$, $\beta_i^{\tau} > \lambda_i^{\tau}(T)$. Then, the delay in observations has an impact with respect to $\tilde{f}$.*

[12] It is taken into account that other behavior strategy profiles than $\tilde{f}$ may yield the Folk Theorem for discount factors below $\tilde{\delta}$, identified in Theorem 1.

[13] Note, however, that this does not rule out that there is some other behavior strategy profile $f \neq \tilde{f}$ such that $f \in BFE(G^{OS,\delta})$ and $\{a^t(f)\}_{t=1}^{\infty} = \{\dot{a}^t\}_{t=1}^{\infty}$.

Appendix B contains the proof of Proposition 1. Intuitively, player i deviates from $\{\dot{a}^t\}_{t=1}^{\infty}$, if the punishment threat of behavior strategy profile $\tilde{f}$ is discounted by too much, and hence, is not strong enough to prevent i's deviation. Whereas the initially prescribed sequences of action profiles under $\bar{f}$ and $\tilde{f}$ are identical, punishment is immediate under $\bar{f}$ but sets in after a lag of d periods under $\tilde{f}$. Thus, the behavior strategy profile defined in Theorem 1 does not support the sequence of action profiles $\{\dot{a}^t\}_{t=1}^{\infty}$ as a BFE of $G^{OS,\delta}$, and *the delay in observations has an impact with respect to* $\tilde{f}$.

Another comparative static result is straightforward given the previous statements. To simplify notation, given some observation structure OS, denote the maximal delay among any pair of players by $d(OS)$.

Corollary 3. *Let* G^*, OS *and* $f \in F$ *with the same structure as* $\tilde{f}$ *be given. Assume that* $f \in BFE(G^{OS,\delta})$ *for all* $\delta \in (\hat{\delta}, 1)$. *Then, for any other observation structure* OS' *represented by* $d(OS')$ *and all* $\delta \in (\hat{\delta}, 1)$, $f \in BFE(G^{OS',\delta})$ *and* $H^{\delta}(f) > 0$, *if, and only if,* $d(OS') \leq d(OS)$.

This result requires that punishment starts d periods after a unilateral deviation, that is, it holds for a behavior strategy profile of the same structure as $\tilde{f}$.

6 Final Remarks

6.1 Less than full-dimensional payoff space and network

In Kinateder (2008), it is shown how this model extends to repeated games with any dimension of the payoff space. The proof of the Folk Theorem and several other results obtain, though the model is significantly more complex. Therefore, it is presented in pure actions. The setup identified there, however, can be extended to mixed actions using the same idea as in FLT.

Kinateder (2008) also identifies a possible application of the model. Suppose that all players that play a repeated game are allocated to a connected network. The distance between any pair of players along shortest paths gives the delay with which both of them observe each other. Then, a Folk Theorem obtains though the network reduces the set of BFE for a certain range of discount factors and under certain conditions.

One way to interpret the network is as a communication network. This is done in Kinateder (2009) who studies the repeated Prisoner's Dilemma in a network. Two players that are linked communicate with each other. Strategic communication is studied and it is shown that for a range of discount factors the set of BFE in this setup does intersect but not coincide with the one in a perfect monitoring repeated Prisoner's Dilemma in which truthful communication is imposed exogenously. New BFE with richer than truthful communication arise while other strategy profiles fail to remain BFE since some player's lie is profitable.

6.2 Conclusion

In this paper, delayed perfect monitoring in an infinitely repeated discounted game is modelled. Each player receives a perfect signal of every other player's action choice with a fixed and finite delay. Two players may observe each other with an asymmetric delay and the delay among different pairs of players is heterogeneous. A Folk Theorem obtains since patient players do not mind to receive the repeated game's history gradually over time. For impatient players the observation structure makes a difference, as shown for the Prisoner's Dilemma. Due to the observation structure, the set of equilibrium payoff vectors is reduced for a range of discount factors and a behavior strategy profile is a BFE over a smaller range of discount factors, both compared with a repeated game with perfect and immediate monitoring.

There are several possibilities to extend the results obtained here, as for example in Kinateder (2008 and 2009). Other extension are extremely involved and therefore left for future research. To identify efficient strategy profiles requires to pick the most efficient one from an infinite number of possible ones. To show the results presented here for a fixed payoff vector requires a similar exercise since frequently an infinite number of sequences of pure action profiles yields the same payoff vector.

References

Abreu, D., P. Dutta and L. Smith (1994), "The Folk Theorem for Repeated Games: A Neu Condition," *Econometrica* 62, 939-948.

Fudenberg, D., D. Levine and E. Maskin (1994), "The Folk Theorem with Imperfect Public Information," *Econometrica* 62, 997-1039.

Fudenberg, D., D. Levine and S. Takahashi (2007), "Perfect Public Equilibrium when Players are Patient," *Games and Economic Behavior* 61, 27-49.

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

Kinateder, M. (2008), “Repeated Games Played in a Network,” *mimeo.*

Kinateder, M. (2009), “The Repeated Prisoner’s Dilemma in a Network,” *mimeo.*

Mailath, G. and L. Samuelson (2006), *Long-Run Relationships,* Oxford University Press.

Appendix A Proof of Theorem 1

Given G^* and OS, fix $x \in \mathcal{F}^*$ such that x is feasible (see footnote 8). Behavior strategy profile $\tilde{f} \in F$, which after being defined is shown to be a BFE of $G^{OS,\delta}$ for any $\delta \in (\tilde{\delta}, 1)$, prescribes a different sequence of pure action profiles $\{a^t\}_{t=1}^{\infty}$ to yield x for each δ, although its structure is unchanged. For any $j \in I$, define $\tilde{f}_j \in F_j$ as follows:
$\tilde{f}_j^1 = a_j^1$, and for $t > 1$, given $ob_j^{t-1} \in Ob_j^{t-1}$, in a slight abuse of notation, let $\tilde{f}_j^t(ob_j^{t-1}) =$

1) a_j^t, unless there is $1 \leq t' < t$ such that for $\hat{a}^{t'} \in ob_j^{t-1}$, $\hat{a}_i^{t'} \neq a_i^{t'}$, while $\hat{a}_{-i}^{t'} = a_{-i}^{t'}$. In this case, switch to phase 2 at $t' + d_j$ and let $\tilde{\sigma}_j^t = a_j^t$, for all $t \geq 1$.

2) $\tilde{\sigma}_j^t$, if $t' + d_j \leq t < t' + d$, unless player l, where $l \neq i$ deviates at any t'', where $t' < t'' < t' + d$. Then, restart phase 2, set $t' = t''$ and choose $\tilde{\sigma}_j^t$ accordingly. Otherwise, switch to phase 3 at $t' + d$.

3) $\bar{\sigma}_j^i$, if $t' + d \leq t \leq t' + T$, where T is determined below. If any player l deviates at any $\bar{t}$, where $t' + T \geq \bar{t} \geq t' + d$, restart phase 2, set $t' = \bar{t}$ and choose $\tilde{\sigma}_j^t$ accordingly. Otherwise, switch to phase 4 at $t' + T + 1$.

4) c_j^s, if $t \geq t' + T + s$, where $\{c^s\}_{s=1}^{\infty}$ is the sequence of action profiles that yields ω^i. If any player l deviates at any $\tau > t' + T$, restart phase 2, set $t' = \tau$ and choose $\tilde{\sigma}_j^t$ accordingly.

Phase 2 corresponds to the ISP, phase 3 to the *minmax* punishment of the last deviator, and phase 4 to the punishment reward phase. After any subsequent unilateral deviation, the phase in which the game is at the time of the deviation prescribes the play of the following $d - 1$ periods—in general, phase 2 is restarted. Then, the new deviator is punished. If the same player deviates again in phase 2 (and no other does), however, phase 2 is not restarted, but his punishment begins d periods after his first deviation. He is forced to his *minmax* payoff for at least $d - 1$ periods. Then, all players know if he deviated again in the period before punishment started, and hence, for how long it has to last in order to eliminate his entire gain.

By construction, the players can ignore multilateral deviations from $\tilde{f}$. Given any observation profile, behavior strategy profile $\tilde{f}$ prescribes a continuation play from which no player can deviate profitably for large enough δ. The result for phase 2 is shown first since it introduces arguments used thereafter to prove the results of phases 4, 1 and 3.

PHASE 2

Figure 2 illustrates the order of time periods in phase 2. Suppose player i deviates at t'. During the ISP player $j \neq i$ receives $ISP_j^{t'}$. By deviating at t'', where $t' < t'' < t' + d$, he can maximally gain $b_j = \max_{\sigma \in \Sigma}[\max_{a_j \in A_j} H_j(a_j, \sigma_{-j}) - H_j(\sigma)]$, since his remaining ISP-payoff is unchanged. However, from period $t'' + d$ on, he is forced to his *minmax* payoff of 0, and then, his punishment reward phase is played. Player j's deviation at t'' is not profitable if for some positive integer $\hat{T}_2$, where $t'' + d \leq t' + \hat{T}_2$,

$$(1-\delta)b_j + \delta^{\hat{T}_2}\omega_j^j - (1-\delta)\sum_{t=t''+d}^{t'+\hat{T}_2} \delta^{t-t''-1} H_j(\bar{\sigma}^i) - \delta^{t'+\hat{T}_2-t''}\omega_j^i \quad < \quad 0,$$

$$(1-\delta)b_j - (1-\delta)\sum_{t=t''+d}^{t'+\hat{T}_2} \delta^{t-t''-1} H_j(\bar{\sigma}^i) \quad < \quad \delta^{t'+\hat{T}_2-t''}\omega_j^i - \delta^{\hat{T}_2}\omega_j^j. \tag{2}$$

Substituting $\delta^{t'+\hat{T}_2-t''}$ with $\delta^{\hat{T}_2}$ makes the right-hand-side of (2) smaller (since $t'' > t'$, $\delta^{t'+\hat{T}_2-t''} > \delta^{\hat{T}_2}$ holds for all $\delta < 1$.) Hence, (3) implies (2) and it suffices to show (3).

$$(1-\delta)b_j - (1-\delta)\sum_{t=t''+d}^{t'+\hat{T}_2} \delta^{t-t''-1} H_j(\bar{\sigma}^i) \quad < \quad \delta^{\hat{T}_2}[\omega_j^i - \omega_j^j] \tag{3}$$

As δ converges to 1, (3) is fulfilled: its left-hand-side converges to zero while its right-hand-side is strictly positive since $\omega_j^i > \omega_j^j$. This may hold for several distinct pairs of discount factor and strictly positive integer. (The last inequality is fulfilled trivially when player j's gain from punishing player i is larger than b_j.) The case $t'' + d > t' + \hat{T}_2$ is simpler since the sum on the left-hand-side of (3) and j's payoff in the first period(s) of i's punishment reward phase both drop out, which for δ close to 1 is negligible.

For $j = i$ after player i's deviation at any t'', where $t' < t'' < t' + d$, the ISP about i's first deviation continues. Once all players know about i's deviation, $\bar{\sigma}^i$ is played for at least $d - 1$ periods, that is, at least until period $t' + 2d - 2$, and at most until his entire gain from all his deviations is eliminated. Thereafter, player i's punishment reward phase is played. Finally, select a large enough, strictly positive integer T_2 such that no player can deviate profitably in phase 2.

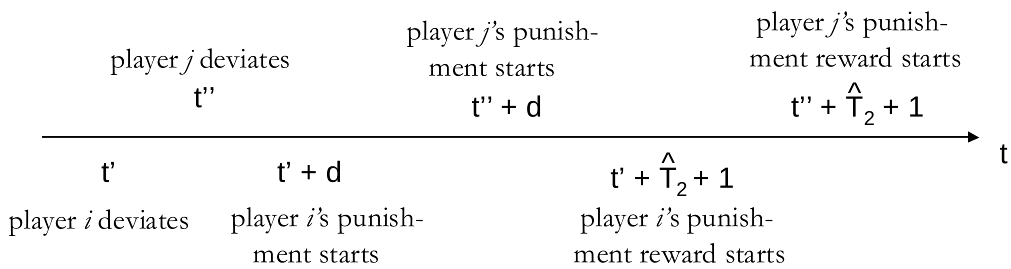

Figure 2: Order of time periods in phase 2

PHASE 4 and PHASE 1

The result for phase 4 is stated first since it implies the result for phase 1. Suppose that player $j \neq i$, and that i is the last deviator. Player j does not deviate at τ, the first period of i's punishment reward phase, if for some positive integer $\hat{T}_4$,

$$(1-\delta)\max_{a_j \in A_j} h_j(a_j, c^1_{-j}) + \delta(1-\delta)ISP^{\tau}_j + \delta^{\hat{T}_4}\omega^j_j - \omega^i_j \quad < \quad 0,$$

$$(1-\delta)\max_{a_j \in A_j} h_j(a_j, c^1_{-j}) + \delta(1-\delta)ISP^{\tau}_j \quad < \quad \omega^i_j - \delta^{\hat{T}_4}\omega^j_j.$$

When δ converges to 1, the left-hand-side of the last inequality converges to zero whereas the right-hand-side is strictly positive (since $\omega^i_j > \omega^j_j$, and for any $\delta < 1$, $\delta^{\hat{T}_4} < 1$). The same argument holds when player j deviates in any other than the first period of player i's punishment reward phase since for δ close to 1, the payoff obtained at the beginning of any punishment reward phase is negligible.

If $j = i$, player i cannot deviate profitably in the first period of his own punishment reward phase, if there is a positive integer $\dot{T}_4$ such that

$$(1-\delta)b_i + \delta(1-\delta)ISP^{\tau}_i + \delta^{\dot{T}_4}\omega^i_i - \omega^i_i \quad < \quad 0,$$

where $\tau \equiv t' + \dot{T}_4 + 1$. This simplifies to

$$(1-\delta)b_i + \delta(1-\delta)ISP^{\tau}_i \quad < \quad \omega^i_i - \delta^{\dot{T}_4}\omega^i_i \ ,$$

$$b_i + \delta ISP^{\tau}_i \quad < \quad \tfrac{(1-\delta^{\dot{T}_4})}{(1-\delta)}\omega^i_i \ . \tag{4}$$

When δ converges to 1, the left-hand-side of (4) is bounded above by a positive number and the right-hand-side, by l'Hospital, converges to $\dot{T}_4\omega^i_i > 0$. The same argument holds when player i deviates in any other than the first period of his own punishment reward phase since for δ close to 1, the payoff obtained at the beginning of any punishment reward phase is negligible. For $\dot{T}_4$ large enough, (4) holds. Hence, no player's unilateral deviation of finite length is profitable in phase 4. Finally, let T_4 be the smallest positive integer such that no player can deviate profitably in phase 4.

The result of phase 4 extends to phase 1 since by assumption any player's target payoff is strictly larger than his punishment reward payoff. Hence, neither any player's finite deviation nor subsequent ones by any player are profitable in phase 1. Again a discount factor $\delta < 1$ and a positive integer T_1 exist such that no player can deviate profitably from behavior strategy profile $\tilde{f}$ in phase 1.

PHASE 3

Suppose player i is forced to his *minmax* payoff because he deviated at t'. By definition, player i cannot deviate profitably in this phase. Neither can any player $j \neq i$ deviate profitably within the support of the mixed *minmax* action. Player j does not deviate by choosing any action outside of the support of the mixed *minmax* action at any $\bar{t}$, where $t' + d \leq \bar{t} \leq t' + T_3$, if

$$(1-\delta)b_j + \delta(1-\delta)ISP_j^{\bar{t}} + \delta^{T_3}\omega_j^j - (1-\delta)\sum_{t=\bar{t}}^{T_3}\delta^{t-\bar{t}}H_j(\bar{\sigma}^i) - \delta^{t'+T_3-\bar{t}}\omega_j^i \quad < \quad 0,$$

$$(1-\delta)b_j + \delta(1-\delta)ISP_j^{\bar{t}} - (1-\delta)\sum_{t=\bar{t}}^{T_3}\delta^{t-\bar{t}}H_j(\bar{\sigma}^i) \quad < \quad \delta^{t'+T_3-\bar{t}}\omega_j^i - \delta^{T_3}\omega_j^j. \qquad (5)$$

Proceeding as in phase 2, that is, substituting on (5)'s right-hand-side $\delta^{t'+T_3-\bar{t}}$ with δ^{T_3} (for any $\delta < 1$, $\delta^{T_3-(\bar{t}-t')} > \delta^{T_3}$ since $\bar{t} > t'$) and taking the limit of δ converging to 1, fulfills (5) for at least one pair of discount factor $\delta < 1$ and strictly positive integer T_3. An analogous argument holds for deviations, or a sequence of deviations by different players. Choose T_3 large enough to prevent any such deviation.

Let $T = \max\{T_1, T_2, T_3, T_4\}$, and let $\tilde{\delta}$ be the lowest discount factor, for which, given T, no player can deviate profitably in any phase. (If there are several pairs of T and δ for which the proof holds, the pair with the lowest discount factor is selected.) Then, for any $\delta \in (\tilde{\delta}, 1)$, $\tilde{f}$ is a BFE strategy profile of $G^{OS,\delta}$ and $H^{\delta}(\tilde{f}) = x$.

Appendix B Proof of Proposition 1

Let G^*, OS and $\delta < 1$ be given. Select $\bar{f} \in BFE(G^\delta)$ that generates the sequence of action profiles $\{a^t(\bar{f})\}_{t=1}^\infty \equiv \{\dot{a}^t\}_{t=1}^\infty$. Take a behavior strategy profile with the same structure as $\tilde{f}$, defined in Theorem 1, to support this sequence of action profiles as a BFE of $G^{OS,\delta}$. Then, *the delay in observations has an impact with respect to* $\tilde{f}$ if some player can deviate profitably. Suppose that for some player $i \in I$, some $\tau \geq 1$, and all positive integers $T \geq 2d - 2$,

$$(1-\delta)\sum_{t=\tau}^{\tau+d-1} \delta^{t-\tau} \max\nolimits_{a_i \in A_i} h_i(a_i, \dot{a}_{-i}^t) + \delta^T \omega_i^i \;\; > \;\; (1-\delta)\sum_{t=\tau}^{\infty} \delta^{t-\tau} h_i(\dot{a}^t),$$

$$\sum_{t=\tau}^{\tau+d-1} \delta^{t-\tau}[\max\nolimits_{a_i \in A_i} h_i(a_i, \dot{a}_{-i}^t) - h_i(\dot{a}^t)] + (1-\delta)^{-1}\delta^T \omega_i^i \;\; > \;\; \sum_{t=\tau+d}^{\infty} \delta^{t-\tau-1} h_i(\dot{a}^t).$$

Subtracting $(1-\delta)^{-1}\delta^T \omega_i^i$ from both sides yields $\beta_i^\tau > \lambda_i^\tau(T)$ and *the delay in observations has an impact with respect to* $\tilde{f}$.