A folk theorem in infinitely repeated prisoner’s dilemma with small observation cost

Hino, Yoshifumi

Vietnam-Japan University

10 December 2018

Online at https://mpra.ub.uni-muenchen.de/99972/
MPRA Paper No. 99972, posted 02 May 2020 10:13 UTC
A folk theorem in infinitely repeated prisoner’s dilemma with small observation cost

Yoshifumi Hino

Business Administration Program, Vietnam–Japan University

This version: April 30, 2020

Abstract

We consider an infinitely repeated prisoner’s dilemma under costly observation. Players choose whether to observe the opponent or not after they choose their actions. If a player observes the opponent, he pays a small observation cost and he can observe the action chosen by his opponent in that period. Otherwise, he receives no signal or an inaccurate private signal. First, we prove an efficiency result that players can achieve a symmetric nearly Pareto efficient outcome. Then, we extend the idea with an interim public randomization device, which is realized just after players choose actions. Players can decide their observational decision after they see the interim public randomization device. We present a folk theorem for a sufficiently small observation cost when players are sufficiently patient.

Keywords Costly observation; Efficiency; Folk theorem; Prisoner’s dilemma

JEL Classification: C72; C73; D82

1 Introduction

A standard insight in the theory of repeated games is that repetition enables players to obtain collusive and efficient outcomes. However, a common and important assumption behind such results is that the players in the repeated game can monitor each other’s past behavior without any cost. We analyze an infinitely repeated prisoner’s dilemma game where each player can only observe his opponent’s previous action at a (small) cost after they choose actions. We establish an approximate efficient result. Then, we introduce an interim public randomization device, which is realized just after they choose actions, and show an approximate folk theorem.

*This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

†Business Administration Program, Vietnam–Japan University, Luu Huu Phuoc Street, My Dinh 1 Ward, Nam Tu Liem District, Hanoi, 12015, Vietnam
Tel: +84 24 7306 6001
Email: hino@vnu.edu.vn
In our model, we consider costly observation as a monitoring structure. Each player chooses his action, and then he makes an observational decision. If a player chooses to observe his opponent, then he can observe the action chosen by the opponent. The observational decision itself is unobservable. The player receives extremely inaccurate private signal.

Furthermore, no player can statistically identify the observational decision of his opponent. That is, our monitoring structure is neither almost-public private monitoring (Hörner and Olszewski (2009); Mailath and Morris (2002, 2006); Mailath and Olszewski (2011)), nor almost perfect private monitoring (Bhaskar and Obara (2002); Chen (2010); Ely and Välimäki (2002); Ely et al. (2005); Hörner and Olszewski (2006); Sekiguchi (1997); Piccione (2002); Yamamoto (2007, 2009)).

We present two results. First, we show that the symmetric Pareto efficient payoff vector can be approximated by a sequential equilibrium under some assumptions regarding the payoff matrix when players are patient and the observation cost is small (efficiency). This first result shows that collusive outcomes can be approximated if it is symmetric. The second result is an approximate folk theorem. We introduce an interim public randomization device just after players choose actions. Players can see the public randomization before they choose their observational decisions. We present an approximate folk theorem under some assumptions regarding the payoff matrix when players are patient and the observation cost is small. We also show that a (standard) public randomization device which is realized at the end of stage game does not work instead of the interim public randomization device. This second result shows that any collusive outcomes can be approximated if an interim public randomization device is available.

The nature of our strategy is similar to the keep-them-guessing strategies in Chen (2010). In our strategy, each player $i$ chooses $C_i$ with certainty at the cooperation state, but randomizes the observational decision. Depending on the observation result, players change their actions from the next period. If the player plays $C_i$ and observes $C_j$, he remains in a cooperation state. However, in other cases (for example, the player does not observe his opponent), player $i$ moves out of the cooperation state and chooses $D_i$. From the perspective of player $j$, player $i$ plays the game as if he randomizes $C_i$ and $D_i$, even though player $i$ chooses pure actions in each state. Such randomized observations create uncertainty about the opponents’ state in each period and give an incentive to observe.

Our main contribution is the efficiency result and an approximate folk theorem in an infinitely repeated prisoner’s dilemma. Some previous studies show that the efficiency result holds if communication or private signals are available. For example, Miyagawa et al. (2008) assume that some noisy information is available even if players do not observe their opponent. We discuss previous studies in Section 2. Our efficiency result holds in the least stringent setting compared with other studies.

Another contribution of the paper is a new approach to the construction of a sequential equilibrium. We consider randomization of observation, whereas previous studies confine their attention to randomization of actions. In many cases, the observational decision is supposed to be unobservable in costly observation models. Therefore, even if a player observes his opponent, he cannot know whether the opponent observes him. If the continuation strategy of the opponent depends on the observational decision in the previous period, the opponent might randomize actions from the perspective of the player, even though the opponent chooses pure actions in each history. This new approach enables us to construct a nontrivial sequential equilibrium.

The rest of this paper is organized as follows. In Section 2, we discuss previous studies,
and in Section 2.1, we focus on some previous literature and explain some difficulties in constructing a cooperative relationship in an infinitely repeated prisoner’s dilemma under costly observation. Section 3 introduces a repeated prisoner’s dilemma model with costly observation. In Section 4, we present our efficiency result. For efficiency result, we do not utilize an interim public randomization device. After that, applying the efficiency result, we present a folk theorem with an interim public randomization device. Section 5 provides concluding remarks.

## 2 Literature Review

We review previous studies on repeated games under costly observation.

One of the greatest difficulties in costly observation is observing the observation activity of opponents, because observational behavior under costly observation is often assumed to be unobservable. Each player has to check this unobservable observation behavior to motivate the other player to observe. Some previous studies circumvent the difficulty by assuming that the observational decision is observable. Kandori and Obara (2004) and Lehrer and Solan (2018) assume that players can observe other players’ observational decisions.

Ben-Porath and Kahneman (2003) analyze an information acquisition model with communication. They show that players can share their information through explicit communication, and present a folk theorem for any level of observation cost. Ben-Porath and Kahneman (2003) consider randomizing actions on the equilibrium path. In their strategy, players report their observations to each other. Then, each player can check whether the other player observes him by the reports. Therefore, players can check the observation activities of other players.

Miyagawa et al. (2008) consider that communication is not allowed, but players can obtain imperfect private signals about the other player’s action even when they do not observe their opponent. They show that players can communicate with each other using private signals, and present a folk theorem for any level of observation cost.

Another approach is introduction of nonpublic randomization device to infinitely repeated prisoner’s dilemma. The nonpublic randomization device enables players to correlate their actions. Hino (2019) shows that if a nonpublic randomization device is available before players choose their actions and observational decisions, then players can achieve an efficiency result.

If these assumptions do not hold, that is, if costless information is unavailable, then cooperation is difficult. Two other papers present folk theorems without costless information. Flesch and Perea (2009) consider observation structures similar to our structure. In their model, players can purchase the information about the actions taken in the past if the players incur an additional cost. That is, some organization keeps track of all the sequence of the action profiles, and each player can purchase the information from the organization. Flesch and Perea (2009) present a folk theorem for an arbitrary observation cost. Miyagawa et al. (2003) consider less stringent models. They assume that no organization keeps track of all the sequence of the action profiles for players. Players can observe the opponent’s action in the current period, and cannot purchase the information about the actions in the past. Therefore, if a player wants to keep track of actions chosen by the opponent, he has to pay observation cost every period. This observation structure is the same as the one in the current paper. Miyagawa et al. (2003) present a folk theorem with a small observation cost.

The above two studies, Flesch and Perea (2009) and Miyagawa et al. (2003), consider
communication through mixed actions. To communicate with each other by mixed actions, the above two papers need more than two actions for each player. This means that their approach cannot be applied to infinitely repeated prisoner’s dilemma under costly. We discuss their implicit communication in Miyagawa et al. (2003); Flesch and Perea (2009) in Section 2.1 in more detail.

It is an open question of whether players can achieve an efficiency result and a folk theorem in two-action games, even though the observation cost is sufficiently small. We show an efficiency result without any randomization device using a mixed observation rather than mixed actions when observation cost is small. We will extend the efficiency result using public randomization, and present a folk theorem in infinitely repeated prisoner’s dilemma when observation cost is small.

2.1 Cooperation failure in the prisoner’s dilemma (Miyagawa et al. (2003))


<table>
<thead>
<tr>
<th></th>
<th>$C_2$</th>
<th>$D_2$</th>
<th>$E_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$</td>
<td>1, 1</td>
<td>-1, 2</td>
<td>-1, -1</td>
</tr>
<tr>
<td>$D_1$</td>
<td>2, -1</td>
<td>0, 0</td>
<td>-1, -1</td>
</tr>
<tr>
<td>$E_1$</td>
<td>-1, -1</td>
<td>-1, -1</td>
<td>0, 0</td>
</tr>
</tbody>
</table>

Table 1: Extended prisoner’s dilemma

Players choose whether he observes the opponent or not together with his action choice. Miyagawa et al. (2003) consider the following keep-them-guessing automaton strategy to approximate payoff vector (1, 1). There are three states: cooperation, punishment, and defection.

In the cooperation state, each player chooses $C_i$ with a sufficiently high probability and chooses $D_i$ with the remaining probability. Players observe their opponent irrespective of their actions in the cooperation state. If players observe $(C_1, C_2)$ or $(D_1, D_2)$, the state remains the same. When $(C_1, D_2)$ or $(D_1, C_2)$ is realized, the state moves to the punishment state. The state moves to the defection state if player $i$ chooses $E_i$ or observes $E_j$. In the punishment state, both players choose $E_i$ for some periods, and then the state moves back to a cooperation state. In the defection state, both players choose $E_i$, and the state remains the same. In both the punishment state and the defection state, the players do not observe their opponent.

Players have an incentive to observe their opponent because their opponent randomizes actions $C_j$ and $D_j$ in the cooperation state. If a player does not observe their opponent, the player cannot know in which state the opponent is in the next period. If the opponent is in the cooperation state, action $E_i$ is a suboptimal because the opponent never chooses action $E_j$. That is, choosing action $E_i$ has some opportunity cost because the opponent is in the cooperation state with a positive probability. However, if the opponent is the punishment state, then action $E_i$ is a unique optimal action. Choosing actions $C_i$ or $D_i$ also has opportunity costs because the opponent is in the punishment state with a positive probability. To avoid these opportunity costs, players have an incentive to observe.
These ideas do not hold in two-action games. Suppose that action $E_i$ is not available and consider the prisoner’s dilemma as an example. If players randomize $C_i$ and $D_i$ in the cooperation state, then one of the best response actions in the cooperation state is action $D_i$. The best response action in punishment and defection states is also $D_i$. As a result, irrespective of player $i$’s observation result, one of the optimal continuation strategies is choosing $D_i$ and not observe player $j$ every period. Therefore, Players don’t have an incentive to observe.

I consider the following automaton strategy. In the initial state, player $i$ randomizes actions and observe the opponent with a positive probability only when he chooses $C_i$. If he chooses $C_i$ and observes $C_j$, he moves to the cooperation state in the next period. Otherwise, he moves to the defection state.\(^1\) In the cooperation state, player $i$ chooses action $C_i$ with probability one, but randomizes observational decision. Only if player $i$ chooses $C_i$ and observes $C_j$, player $i$ can remain in the cooperation state. Otherwise, player $i$ moves to the defection state.

The reason why our strategy works is that the strategy prescribes pure action of $C_i$ and does not prescribe a mixed actions in the cooperation state. The repetition of $D_i$ from the cooperation state is not prescribed action. However, it causes another problem related to the observation incentive. As player $j$ does not randomize his action in the cooperation state, player $i$ can easily guess player $j$’s action if he knows that player $j$ is in the cooperation state. In such a situation, player $i$ loses the observation incentive again.

Our strategy can overcome this difficulty as well. Since player $j$ randomize his observational decisions in the cooperation state, player $i$ in the cooperation state cannot know whether player $j$ observed player $i$ or not. If player $j$ does not observes player $i$, player $j$ moves to the defection state and chooses $D_j$. Player $i$ cannot be certain that player $j$ is in the cooperation state even if he chooses $C_i$ and observes $C_j$ in the previous period. To obtain the latest information about player $j$’s state, player $i$ has an incentive to observe the opponent in the cooperation state. This is why player $i$ has an incentive to observe player $j$ given our strategy.

### 3 Model

The stage game is a symmetric prisoner’s dilemma, but it has two phases: the action phase and the observation phase. In the action phase, each player $i$ $(i = 1, 2)$ chooses an action, $C_i$ or $D_i$. Let $A_i \equiv \{C_i, D_i\}$ be the set of actions for player $i$. After both players chooses actions, each player $i$ receives a signal $z_i$ costlessly and privately. The set of private signal for player $i$ is finite set and denoted by $Z_i$. A signal profile $z = (z_1, z_2) \in Z \equiv Z_1 \times Z_2$ is realized with probability $\rho(z|a)$ given an action profile $a = (a_1, a_2) \in A \equiv A_1 \times A_2$.

#### Assumption 1

There exists some $\zeta > 0$ such that

$$\rho(z|a) > \zeta, \quad \forall z \in Z, \quad \forall a \in A.$$ 

We define the accuracy $\eta_i$ of the signal $z_i$ as follows.

$$\eta_i \equiv 1 - \min_{z_i \in Z_i, a, a' \in A} \frac{\rho(z_i|a')}{\rho(z_i|a)}.$$

\(^1\)For the formal proof, we need another state (transition state). Transition state is crucial only when we consider off the equilibrium path. Therefore, it is omitted here.
The base game payoff for player $i$ is given by $\pi_i(a_i, z_i)$. Given an action profile $a \in A$, an expected base game payoff for player $i$, $u_i(a) \equiv \sum_{z_i \in Z_i} P(z_i|a) \pi_i(a_i, z_i)$, is displayed in Table 2.

<table>
<thead>
<tr>
<th>Player 2</th>
<th>C2</th>
<th>D2</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>$1$, $1$</td>
<td>$-\ell$, $1 + g$</td>
</tr>
<tr>
<td>D1</td>
<td>$1 + g$, $-\ell$</td>
<td>$0$, $0$</td>
</tr>
</tbody>
</table>

Table 2: Prisoner’s dilemma

We make a usual assumption about the above payoff matrix.

Assumption 2. (i) $g > 0$ and $\ell > 0$; (ii) $g - \ell < 1$.

The first condition implies that action $C_i$ is dominated by action $D_i$ for each player $i$, and the second condition ensures that the payoff vector of action profile $(C_1, C_2)$ is Pareto efficient.

We impose an additional assumption.

Assumption 3. $g - \ell > 0$.

Assumption 3 is the same as Assumption 1 in Chen (2010).

Players simultaneously choose their observational decision in the observation phase after they choose their actions in the action phase. Let $m_i$ represent the observational decision for player $i$. Let $M_i \equiv \{0, 1\}$ be the set of observational decisions for player $i$, where $m_i = 1$ represents “to observe the opponent,” and $m_i = 0$ represents “not to observe the opponent.”

If player $i$ observes the opponent, he incurring an observation cost $\lambda > 0$, and receives complete information about the action chosen by the opponent at the end of the stage game. If player $i$ does not observe the opponent, he does not incur any observation cost and obtains no information about his opponent’s action. We assume that the observational decision for a player is unobservable.

A stage behavior for player $i$ is a pair of base game action $a_i$ for player $i$ and observational decision $m_i$ for player $i$ and is denoted by $b_i = (a_i, m_i)$. An outcome of the stage game is a pair of stage behaviors $b = (b_1, b_2)$. Let $B_i \equiv A_i \times M_i$ be the set of stage-behaviors for player $i$, and let $B \equiv B_1 \times B_2$ be the set of outcomes of the stage game. Given an outcome $b \in B$, the stage game payoff $\pi_i(b)$ for player $i$ is given by

$$U_i(b) \equiv u_i(a_1, a_2) - m_i \cdot \lambda.$$ 

For any observation cost $\lambda > 0$, the stage game has a unique stage game Nash equilibrium outcome, $b^* = ((D_1, 0), (D_2, 0))$.

Let $\delta \in (0, 1)$ be a common discount factor. Players maximize their expected average discounted stage game payoffs. Given a sequence of outcomes of the stage games $(b^t)_{t=1}^{\infty}$, player $i$’s payoff is given by average discounted stage game payoff:

$$(1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} U_i(b^t).$$

Player $i$’s nonaveraged payoff is given by:

$$\sum_{t=1}^{\infty} \delta^{t-1} U_i(b^t).$$
We assume that an interim public randomization device is available just before players choose their observational decisions. The random variable $X$ is uniformly distributed over $[0,1)$ independently of the action profile. Each player observes the realized public randomization without any cost.

Let $a_i \in A_j \cup \{\phi_i\}$ be an observation result for player $i$. Observation result $a_i = a_j \in A_j$ implies that player $i$ chose observational decision $m_i = 1$ and observed $a_j$. Observation result $a_i = \phi_i$ implies that player $i$ chose $m_i = 0$, that is, he obtains no information about the action chosen by the opponent.

Let $h^t_i$ be a (private) history of player $i$ at the beginning of the action phase in period $t \geq 2$. This history $h^t_i$ is a sequence of his own actions, realized public randomizations, observation results, and private signals up to period $t - 1$: $h^t_i = (a^k_i, x^k_i, o^k_i, z^k_i)_{k=1}^{t-1}$. We omit the observational decisions $m^k_i(k < t)$ from $h^t_i$ because observation result $a^k_i$ implies the observational decision $m^k_i$ for any $k < t$. Let $H^t_i$ denote the set of all the histories for player $i$ at the beginning of the action phase in period $t \geq 1$, where $H^1_i$ is an arbitrary singleton set. Similarly, a history $h^t_i$ at the beginning of the observation phase in period $t \geq 1$ is $(h^t_i, a^t_i, x^t_i)$.

An action strategy for player $i$ in the repeated game is a function of the history $h^t_i$ of player $i$ in the action phase to his (mixed) actions. An observation strategy for player $i$ in the repeated game is a function of a history $h^t_i$ in the observation phase to his (mixed) observational decision. A (behavior) strategy is a pair of action strategy and observation strategy.

The belief $\psi^t_i$ of player $i$ in period $t$ is a function of the history $h^t_i$ in period $t$ to a probability distribution over the set of histories for player $j$ in period $t$; $H^t_j$. Let $\psi_i \equiv (\psi^t_i)_{t=1}^{\infty}$ be a belief for player $i$, and $\psi = (\psi_1, \psi_2)$ denote a system of beliefs.

A strategy profile $\sigma$ is a pair of strategies $\sigma_1$ and $\sigma_2$. Given a strategy profile $\sigma$, a sequence of completely mixed behavior strategy profiles $(\sigma^n)_{n=1}^{\infty}$ that converges to $\sigma$ is called a tremble. Each completely mixed behavior strategy profile $\sigma^n$ induces a unique system of beliefs $\psi^n$.

The solution concept is a sequential equilibrium. We say that a system of beliefs $\psi$ is consistent with strategy profile $\sigma$ if a tremble $(\sigma^n)_{n=1}^{\infty}$ exists such that the corresponding sequence of systems of beliefs $(\psi^n)_{n=1}^{\infty}$ converges to $\psi$. Given the system of beliefs $\psi$, strategy profile $\sigma$ is sequentially rational if, for each player $i$, the continuation strategy from any history in each phase is optimal given his belief and the opponent’s strategy. It is defined that a strategy profile $\sigma$ is a sequential equilibrium if a consistent system of beliefs $\psi$ for which $\sigma$ is sequentially rational exists.

## 4 Results

In this section, we show our efficiency result. Then, applying the efficiency result, we present a folk theorem with an interim public randomization device.

To prove the desired propositions, first, we assume $\eta_1 = \eta_2 = 0$. It means that a player obtains no information about the action of the opponent if he does not observe the opponent. We present related propositions given $\eta_1 = \eta_2 = 0$. After that, we will show the desired propositions using the related propositions.
4.1 Efficiency

The following proposition shows that the symmetric efficient outcome is approximated by a sequential equilibrium if the observation cost $\lambda$ is small and the discount factor $\delta$ is moderately low.

Proposition 1. Suppose that $\eta_1 = \eta_2 = 0$, Assumptions 2 and 3 are satisfied. For any $\varepsilon > 0$, there exist $\tilde{\delta} \in \left(\frac{g}{1+g}, 1\right)$, $\bar{\delta} \in (\tilde{\delta}, 1)$, and $\bar{\lambda} > 0$ such that for any discount factor $\delta \in [\tilde{\delta}, \bar{\delta}]$ and for any observation cost $\lambda \in (0, \bar{\lambda})$, there exists a symmetric sequential equilibrium $\sigma^*$ whose payoff vector $(v_1^*, v_2^*)$ satisfies $v_i^* \geq 1 - \varepsilon$ for each $i = 1, 2$.

Proof. See Appendix A. □

We here present the main idea. The precise proof will be given in Appendix A.

Strategy

First, we define our strategy $\sigma^*$. Fix any $\varepsilon > 0$. We define $\varepsilon$, $\tilde{\delta}$, $\bar{\delta}$, and $\bar{\lambda}$ as follows.

$$\varepsilon \equiv \frac{\ell^2}{4(1+2g)^3} \frac{\varepsilon}{1+\varepsilon},$$

$$\tilde{\delta} \equiv \frac{g}{1+g} + \varepsilon,$$

$$\bar{\delta} \equiv \frac{g}{1+g} + 2\varepsilon < 1,$$

$$\bar{\lambda} \equiv \frac{1}{16} \frac{\ell}{(1+2g)^2 \varepsilon^2}.$$

We fix an arbitrary discount factor $\delta \in [\tilde{\delta}, \bar{\delta}]$ and an arbitrary observation cost $\lambda \in (0, \bar{\lambda})$.

Our strategy $\sigma^*$ is represented by an automaton independently of private signal $z$. Let us consider the following automaton who has four types of states: initial state $\omega_1^i$, cooperation state $(\omega_{i,t}^C)_{t=2}$, transition state $(\omega_{i,t}^E)_{t=2}$, and defection state $\omega_i^D$. In the initial state $\omega_1^i$, player $i$ chooses $D_i$ with probability $\beta_{i,1}$, and chooses $C_i$ with probability $1 - \beta_{i,1}$. When player $i$ chooses $C_i$, he observes the opponent with probability $1 - \beta_{i,1}$. Player $i$ never observes the opponent when player $i$ chooses $D_i$. In the cooperation state $\omega_{i,t}^C (t \geq 2)$, player $i$ chooses $C_i$. If player $i$ chooses $C_i$, he chooses $m_i = 1$ with probability $1 - \beta_{i,t+1}$. When player $i$ chooses $D_i$, he never observes the opponent. In the transition state $\omega_{i,t}^E (t \geq 2)$ and defection state $\omega_i^D$, player $i$ chooses $D_i$ and does not observe the opponent irrespective of his action. The prescribed actions and observational decisions are summarized in the table below.

<table>
<thead>
<tr>
<th>State</th>
<th>$\omega_1^i$</th>
<th>$\omega_{i,t}^C$</th>
<th>$\omega_{i,t}^E$</th>
<th>$\omega_i^D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action</td>
<td>$C_i$ w.p. $1 - \beta_{i,1}$</td>
<td>$C_i$</td>
<td>$D_i$</td>
<td>$D_i$</td>
</tr>
<tr>
<td></td>
<td>$D_i$ w.p. $\beta_{i,1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_i$ given $C_i$</td>
<td>$m_i = 1$ w.p. $1 - \beta_{i,2}$</td>
<td>$m_i = 1$ w.p. $1 - \beta_{i,t+1}$</td>
<td>$m_i = 0$</td>
<td>$m_i = 0$</td>
</tr>
<tr>
<td></td>
<td>$m_i = 0$ w.p. $\beta_{i,2}$</td>
<td>$m_i = 0$ w.p. $\beta_{i,t+1}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_i$ given $D_i$</td>
<td></td>
<td></td>
<td>$m_i = 0$</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Actions and observational decisions

---

2 The probability $\beta_{i,t} (t \geq 1)$ will be defined using (1) and (2) later.
The state transition function is defined as follows. In the initial state $\omega^1_i$, if player $i$ observes $(a^t_i, o^t_i) = (C_i, C_j)$, he moves to the cooperation state $\omega^C_{i,2}$. When player $i$ chooses $D_i$ or observes $D_j$, the state in the next period is $\omega^D_i$. Only when player $i$ observes $(a^t_i, o^t_i) = (C_i, \phi_i)$, the state moves to the transition state $\omega^{E,2}_i$. In the cooperation state and transition state in period $t$, player $i$ moves to the cooperation state $\omega^C_{i,t+1}$ if he observes $(a^t_i, o^t_i) = (C_i, C_j)$. If $(a^t_i, o^t_i) = (C_i, \phi_i)$, he moves to the transition state $\omega^{E,t+1}_i$. If player $i$ chooses $D_i$ or observes $D_j$, the state moves to the defection state $\omega^D_i$. Note that player $i$ moves back to the cooperation state $\omega^C_{i,t+1}$ from the transition state in period $t$ if he observes $(C_i, C_j)$, which is the event off the equilibrium path. The defection state $\omega^D_i$ is an absorbing state and player $i$ never moves to another state from the defection state $\omega^D_i$.

The state transition is summarized in Figure 1.

![State transition function](image)

**Figure 1:** State transition function

Using the above automaton, we fix randomization probabilities in each state. Let us define $\varepsilon' = \delta - \frac{q}{1+g}$. First, we fix a small probability $\beta_{i,1} \equiv \frac{1+g+\ell}{g+\ell} \varepsilon'$. We fix a probability $\beta_{i,2}$ so that player $j$ is indifferent between actions $C_j$ and $D_j$ in the initial state $\omega^1_j$. Hence, $\beta_{i,2}$ is determined as the solution of the following equality.

$$(1 - \beta_{i,1})(1 + g) = (1 - \beta_{i,1}) \cdot 1 - \beta_{i,1} \cdot \ell + \delta(1 - \beta_{i,1})(1 - \beta_{i,2})(1 + g).$$ (1)

The left-side is the nonaveraged payoff when player $j$ chooses $(a^1_j, m^1_j) = (D_j, 0)$ in the initial state $\omega^1_j$. The right-side is the one when player $j$ chooses $(a^1_j, m^1_j) = (C_j, 0)$.

Probability $\beta_{i,t+2}(t \geq 1)$ is determined to make player $j$ in state $\omega^C_{j,t}$ indifferent between $m_j = 1$ and $m_j = 0$ given his action $C_j$. Player $j$ believes that player $i$ is in the cooperation state with probability $1 - \beta_{i,t}$ because he observes $C_j$ in the previous period and he is sure that player $j$ was in the cooperation state $\omega^C_{j,t-1}$ in the previous period $t - 1$. Therefore, probability $\beta_{i,t+2}$ is a solution of the following equality.

$$\delta(1 - \beta_{i,t})(1 - \beta_{i,t+1})(1 + g)$$
$$= (1 - \beta_{i,t}) \cdot 1 - \beta_{i,t} \cdot \ell - \lambda$$
$$+ \delta(1 - \beta_{i,t}) \{(1 - \beta_{i,t+1}) \cdot 1 - \beta_{i,t+1} \cdot \ell + \delta(1 - \beta_{i,t+1})(1 - \beta_{i,t+2})(1 + g)\}$$ (2)
The left-side is the nonaveraged payoff when player $j$ chooses $m_j = 0$ in period $t$. The right-side is the one when player $j$ chooses $m_j = 1$ in period $t$ and chooses $(C_j, 0)$ if he is in the cooperation state $\omega_j^{C,t+1}$ in period $t + 1$.

Specifically, $\beta_{i,2}$ is defined by (1), and $\beta_{i,t+2}$ ($t \in \mathbb{N}$) is defined by (2), or

$$
\beta_{i,2} = \frac{(1 - \beta_{i,1}) \{\delta(1 + g) - g\} - \beta_{i,1} \ell}{\delta(1 - \beta_{i,1})(1 + g)} = \frac{g + g^2 - \ell^2 - (1 + g + \ell)(1 + g)\varepsilon'}{(g + \ell) \{g + (1 + g)\varepsilon\} \left(1 - \frac{1 + g + \ell \varepsilon'}{g + \ell}\right)} \varepsilon' \\
\beta_{i,t+2} = \frac{(1 - \beta_{i,t+1}) \{\delta(1 + g) - g\} - \beta_{i,t+1} \ell - \frac{\lambda}{\delta(1 - \beta_{i,t+1})(1 + g)}}{\lambda}, \quad \forall t \in \mathbb{N}.
$$

The following Lemma 1, which is proved in Appendix B, ensures that any $\beta_{i,t}$ is greater than zero and smaller than one.

**Lemma 1.** Suppose that Assumptions 2 and 3 are satisfied. Fix any discount factor $\delta \in [\underline{\delta}, \overline{\delta}]$ and observation cost $\lambda \in (0, \overline{\lambda})$. Then, it holds that

$$
\frac{1}{2} \frac{1 + g - \ell}{g + \ell} \varepsilon' < \beta_{i,2} < \beta_{i,4} < \beta_{i,6} \cdots < \beta_{i,5} < \beta_{i,3} < \beta_{i,1} = \frac{1 + g + \ell}{g + \ell} \varepsilon'.
$$

Strategy $\sigma^*$ is the strategy defined by the above automaton.

Next we define a consistent system of beliefs with strategy profile $\sigma^*$. We consider a sequence of behavioral strategy profiles $(\hat{\sigma}^n)_{n=1}^\infty$ such that each strategy profile attaches a positive probability to every move, but puts far greater weights on the trembles on $C_i$ in the defection state $\omega_i^D$ compared with other stage behaviors in the other states. These trembles induce a consistent system of beliefs that player $i$ at any defection state $\omega_i^D$ is sure that the state of their opponent is the defection state $\omega_j^D$ or transition state $\omega_j^{E,t}$ for some $t \geq 2$.

Let us confirm this property of the belief. There are two cases when player $i$ moves to the defection state $\omega_i^D$: (1) player $i$ observes $D_j$, (2) player $i$ chooses $D_i$. The property is obvious in the first case. In any state of player $j$, player $j$ moves to the defection state $\omega_j^D$ after he chooses $D_j$. Furthermore, the defection state $\omega_j^D$ is an absorbing state. Therefore, player $i$ is certain that player $j$ is in the defection state $\omega_j^D$ after player $i$ observes $D_j$. The property is not obvious in the second case; $a_i = D_i$. Let us consider the following history of player $i$ in period 3. Player $i$ chooses $a_i = D_i$ and $m_i = 0$ in period 1, and he chooses $C_i$ and $m_i = 1$ (by mistakes) and observes $C_j$ in period 2. We can consider the following two types of player $j$’s histories which are consistent with the history of player $i$. The first type of history is that player $j$ chooses $a_j = D_j$ in period 1, and he chooses $a_j = C_j$ (by mistake) at the defection state $\omega_j^D$ in period 2. The second type of history is that player $j$ chooses $a_j = C_j$ and $m_j = 0$ in period 1, and he chooses $a_j = C_j$ (by mistake) at the transition state $\omega_j^{E,2}$ in period 2. As we put far greater weights on the trembles on $C_j$ in the defection state $\omega_j^D$, player $i$ is sure that the first type of history is realized, and player $j$ is in the defection state $\omega_j^D$. A similar argument holds even if player $i$ observes $(a_i, o_i) = (C_i, C_j)$ many times after he chooses $D_i$. 

10 / 29
An illustration

We here explain that the strategy $\sigma^*$ is a sequential equilibrium whose payoff vector $(v_i^*, v_j^*)$ satisfies $v_i^* \geq 1 - \varepsilon$ for each $i = 1, 2$.

Let us consider sequential rationality in each state. First, we consider the defection state $\omega_i^D$. As we have considered above, player $i$ in the defection state $\omega_i^D$ is certain that player $j$ is in the defection state $\omega_j^D$. Therefore, action $D_i$ is optimal because player $i$ is sure that player $j$ does not observe player $i$. As player $i$ is certain that player $j$ is in the defection state $\omega_j^D$ and chooses $D_j$, observational decision $m_i = 0$ is also optimal.

Let us consider sequential rationality in the initial and cooperation states. By the definition of $\beta_{j,2}$ and $\beta_{j,3}$, player $i$ is indifferent among $(C_i, 1)$, $(C_i, 0)$, and $(D_i, 0)$. Furthermore, if player $i$ chooses $D_i$ in the initial state $\omega_i^1$, player $j$ moves to the transition state $\omega_i^{E,2}$ or defection state $\omega_i^D$. In either case, the continuation strategy of player $j$ is a repetition of $(a_j, m_j) = (D_j, 0)$. As the observation result has no effect on the conjecture over the continuation strategy, player $i$ has no incentive to choose $m_i = 0$ when he chooses action $D_i$.

Therefore, it is optimal for player $i$ to follow strategy $\sigma^*$ in the initial state $\omega_i^1$.

In the cooperation state $\omega_i^{C,t}(t \geq 2)$, player $i$ is indifferent to his observational decisions by the definition of $\beta_{j,t+2}$. It is also suboptimal to choose $(a_i^t, m_i^t) = (D_i, 1)$ as in the initial state $\omega_i^1$. Furthermore, the definition of $\beta_{j,t+1}$ ensures that player $i$ strictly prefers action $C_i$ to $D_i$ in the cooperation state $\omega_i^{C,t}$. Using (2) for $t - 1$, we obtain the following equation.

\[ (1 - \beta_{j,t}) - \beta_{j,t+1} \ell + \delta (1 - \beta_{j,t})(1 - \beta_{j,t+1})(1 + g) - \delta (1 + g) = \frac{\lambda}{\delta (1 - \beta_{j,t-1})} \] (3)

The first three terms on the right-hand side represent the nonaveraged payoff when player $i$ chooses $C_i$ and $m_i = 0$ in the cooperation state $\omega_i^{C,t}$. The last term on the right-hand side is the nonaveraged payoff when player $i$ chooses $(a_i, m_i) = (D_i, 0)$ in the cooperation state $\omega_i^{C,t}$. Therefore, (3) shows that choosing $D_i$ at the cooperation state $\omega_i^{C,t}$ is not optimal. Sequential rationality at the cooperation state $\omega_i^{C,t}$ is satisfied.

Another explanation is as follows. Suppose that player $i$ weakly prefers action $D_i$ at the cooperation state $\omega_i^{C,t}$ in period $t$. As player $j$ moves to the transition state $\omega_i^{E,t+1}$ or the defection state $\omega_i^D$ after player $i$ chooses $D_i$ in period $t$, the assumption implies that player $i$ weakly prefers $(a_i, m_i) = (D_i, 0)$ from period $t$ onwards. One of the optimal continuation strategies from the cooperation state $\omega_i^{C,t}$ coincides with the one from the defection state $\omega_i^D$.

Then, player $i$ has no incentive to observe player $j$ in the cooperation state $\omega_i^{t-1}$ because the repetition of $(a_i, m_i) = (D_i, 0)$ is one of his optimal continuation strategies irrespective of the observation result. It contradicts the definition of $\beta_{j,t+1}$. Therefore, player $i$ strictly prefers action $C_i$ in the cooperation state $\omega_i^{C,t}$.

Next, let us consider the transition state $\omega_i^{E,t}$. In the transition state $\omega_i^{E,t}$, player $i$ does not know the action chosen by the opponent in the previous period $t - 1$. If $(a_i, a_j) = (C_i, C_j)$ is realized in the previous period, player $i$ should be at the cooperation state $\omega_i^{C,t}$ and action $D_i$ is suboptimal.

Although action $D_i$ is suboptimal in the cooperation state $\omega_i^{C,t}$, the payoff when player $i$ chooses $D_i$ at $\omega_i^{C,t}$ is close enough to the one when he chooses $D_i$ at $\omega_i^{C,t}$ when the observation cost $\lambda$ is sufficiently small. If the payoffs are not close to each other, player $i$ strictly prefers $m_i = 1$ at the cooperation state $\omega_i^{C,t-1}$ to know which state he should move to because the observation cost is small.
The loss from choosing $D_i$ in the transition state $\omega^{E,t}_i$ is small. The loss from choosing $C_i$ is strictly positive. Player $j$ is in the transition state $\omega^{E,t}_j$ or defection state $\omega^D_j$ with probability at least $(1 - \beta_{j,t-1})(1 - \beta_{j,t})$, and then choosing $C_i$ makes a loss of $-\ell$. Therefore, choosing $C_i$ is suboptimal at the transition state $\omega^{E,t}_i$. We will prove this fact in Appendix A.

Next, let us consider the observation decision in the transition state $\omega^{E,t}_i$. It is straightforward that if player $i$ chooses $D_i$, then $m_i = 0$ is optimal. Assume that player $i$ chooses $C_i$. If player $j$ chooses $C_j$ in the previous period, then player $i$ should have been at the cooperation state $\omega^{C,t}_i$ and one of the optimal stage behaviors given action $C_i$ was $m_i = 0$. If player $j$ chooses $D_j$ in the previous period, then one of player $i$’s optimal stage behaviors was $m_i = 0$. In each case, $m_i = 0$ is optimal. Therefore, $m_i = 0$ is optimal in the transition state $\omega^{E,t}_i$.

Lastly, let us consider the payoff. As player 1 prefers action $D_i$ in the initial state $\omega^1_i$, his payoff is given by

$$v^*_i = (1 - \delta)(1 - \beta_{j,1})(1 + g)$$

$$= (1 - \delta)g(1 + g)$$

$$= (1 + g + \ell)\varepsilon'$$

$$> 1 - (1 + g + \ell)\varepsilon'$$

$$> 1 - \varepsilon.$$
Next, we extend Proposition 2 using Lemma 2.

**Lemma 2.** Fix any payoff vector $v$ and any $\varepsilon > 0$. Suppose that there exist $\delta \in \left(\frac{\eta}{1+g}, 1\right)$, $\delta \in (\delta, 1)$ such that for any discount factor $\delta \in [\delta, \delta]$, there exists a sequential equilibrium whose payoff vector $(v_1^*, v_2^*)$ satisfies $|v_i^* - v_i| \leq \varepsilon$ for each $i = 1, 2$. Then, there exists $\delta^* \in \left(\frac{\eta}{1+g}, 1\right)$ such that for any discount factor $\delta \in [\delta^*, 1)$, there exists a sequential equilibrium whose payoff vector $(v_1^*, v_2^*)$ satisfies $|v_i^* - v_i| \leq \varepsilon$ for each $i = 1, 2$.

**Proof of Lemma 2.** We use the technique of Lemma 2 in Ellison (1994). We define $\delta^* = \delta/\delta$, and choose any discount factor $\delta \in (\delta, 1)$. Then, we choose some integer $n^*$ that satisfies $\delta n^* \in [\delta, \delta]$. Then there exists a strategy $\sigma^*$ whose payoff vector is $(v_1^*, v_2^*)$ given $\delta n^*$. We divide the repeated game into $n^*$ distinct repeated games. The first repeated game is played in period $1, n^*+1, 2n^*+1, \ldots$, the second repeated game is played in period $2, n^*+1, 2n^*+2, \ldots$, and so on. Each repeated game can be regarded as a repeated game with discount factor $\delta n^*$. Let us consider the following strategy $\sigma^L$. In the 1st game, players follow strategy $\sigma^*$. In the 2nd game, players follow strategy $\sigma^*$. In the $n(n \leq n^*)$th game, players follow strategy $\sigma^*$. Then, strategy $\sigma^L$ is a sequential equilibrium because strategy $\sigma^*$ is a sequential equilibrium in each game. As the equilibrium payoff vector in each game satisfies $|v_i^* - v_i| \leq \varepsilon$ for each $i = 1, 2$, the equilibrium payoff of strategy $\sigma^L$ also satisfies $|v_i^* - v_i| \leq \varepsilon$ for each $i = 1, 2$. □

We obtain efficiency for a sufficiently high discount factor.

**Proposition 3.** Fix any $\zeta > 0$. Suppose that Assumptions 2 and 3 are satisfied. For any $\varepsilon > 0$, there exist $\delta^* \in (0, 1)$, $\lambda > 0$, and $\eta > 0$ such that for any discount factor $\delta \in (\delta^*, 1)$, any $\lambda \in (0, \bar{\lambda})$, and any $\eta_1, \eta_2 \in [0, \bar{\eta}]$, there exists a sequential equilibrium whose payoff vector $(v_1^*, v_2^*)$ satisfies $v_i^* \geq 1 - \varepsilon$ for each $i = 1, 2$.

**Proof of Proposition 3.** Apply Lemma 2 to Proposition 1. □

**Remark 3.** Proposition 3 shows monotonicity of efficiency on the discount factor. If efficiency holds given some $\varepsilon$, observation cost $\lambda$, $\eta_1, \eta_2$ and discount factor $\delta$, then efficiency holds given a sufficiently large discount factor $\delta' > \delta$.

### 4.2 Folk theorem

In what follows, we introduce an interim public randomization device at the end of the action phase. Public signal $x$ is uniformly distributed over $[0, 1]$ independently of the action profile chosen. Each player observes the interim public signal without cost. The purpose of interim public randomization is to prove a folk theorem (Theorem 1).

Let

- $\mathcal{F} \equiv \text{convex hull of } \{u(a) | a \in A\}$,
- $\mathcal{F}^* \equiv \{v \in \mathcal{F} | v_1 \geq 0 \text{ and } v_2 \geq 0\}$.

**Theorem 1** (Approximate folk theorem). Suppose that an interim public randomization is available, and Assumptions 2 and 3 are satisfied. Fix any positive $\zeta > 0$. Fix any interior point $v = (v_1, v_2)$ of $\mathcal{F}^*$. Fix any $\varepsilon > 0$. There exist a discount factor $\delta \in \left(\frac{\eta}{1+g}, 1\right)$, observation cost $\lambda > 0$, and $\eta > 0$ such that for any $\delta \in [\delta, 1)$, any $\lambda \in (0, \bar{\lambda})$, and any $\eta_1, \eta_2 \in [0, \bar{\eta}]$, there exists a sequential equilibrium whose payoff vector $v^F = (v^F_1, v^F_2)$ satisfies $|v_i^F - v_i| \leq \varepsilon$.
To prove Theorem 1, we prove the following proposition first.

**Proposition 4.** Suppose that a public randomization device is available, and \( \eta_1 = \eta_2 = 0 \). Assume Assumptions 2 and 3 are satisfied. For any \( \varepsilon > 0 \), there exist \( \delta \in \left( \frac{\varepsilon}{1+g}, 1 \right) \), \( \tilde{\delta} \in (\delta, 1) \), and \( \tilde{\lambda} > 0 \) such that for any discount factor \( \delta \in [\tilde{\delta}, \tilde{\delta}] \) and for any observation cost \( \lambda \in (0, \tilde{\lambda}) \), there exists a sequential equilibrium \( \sigma^{**} \) whose payoff vector \((v_1^{**}, v_2^{**})\) satisfies \( v_1^{**} = 0 \) and \( v_2^{**} \geq \frac{1+g+\ell}{1+\ell} - \varepsilon \).

**Strategy**

First, we define strategy \( \sigma^{**} \) independently of private signal \( z \), which will be used to present Proposition 4.

Fix any \( \varepsilon > 0 \). We define \( \varepsilon, \tilde{\delta}, \tilde{\lambda} \) and \( \overline{\lambda} \) as follows.

\[
\varepsilon \equiv \frac{\ell^2}{54(1+g)^3} \frac{\varepsilon}{1+\varepsilon},
\]

\[
\delta \equiv \frac{g}{1+g} + \varepsilon,
\]

\[
\tilde{\delta} \equiv \frac{g}{1+g} + 2\varepsilon < 1,
\]

\[
\overline{\lambda} \equiv \frac{1}{16} \frac{\ell}{(1+2g)^2}\varepsilon^2.
\]

We fix an arbitrary discount factor \( \delta \in [\tilde{\delta}, \overline{\lambda}] \) and an arbitrary observation cost \( \lambda \in (0, \overline{\lambda}) \). We show that there exists a sequential equilibrium whose payoff vector \((v_1^{**}, v_2^{**})\) satisfies \( v_1^{**} = 0 \) and \( v_2^{**} \geq \frac{1+g+\ell}{1+\ell} - \varepsilon \).

Applying the strategy in Section 4.1, let us consider another automaton strategy profile \( \sigma^{**} \). Player 1 has five types of states: Initial state \( \hat{\omega}_1 \), adjustment state \( \omega_1^A \), cooperation states \((\omega_1^{C,t})_{t=3}^{\infty}\), transition states \((\omega_1^{E,t})_{t=3}^{\infty}\), and defection state \( \omega_1^D \). Player 2 also has five types of states: Initial state \( \hat{\omega}_2 \), adjustment state \( \omega_2^A \), cooperation states \((\omega_2^{C,t})_{t=3}^{\infty}\), transition states \((\omega_2^{E,t})_{t=3}^{\infty}\), and defection state \( \omega_2^D \).

The stage behaviors and transition functions in the cooperation states \((\omega_1^{C,t})_{t=3}^{\infty}\) and \((\omega_2^{C,t})_{t=2}^{\infty}\), transition states \((\omega_1^{E,t})_{t=3}^{\infty}\) and \((\omega_2^{E,t})_{t=2}^{\infty}\), and the defections state \( \omega_i^D \) are the same as those given in strategy \( \sigma^* \). Note that \( \hat{\omega}_1, \omega_1^A, \omega_1^C, \omega_2^C, \omega_2^A, \omega_2^C \) are new states.

To define the stage behaviors and transition functions in the new states, we use the sequence \((\beta_{i,t})_{i=1,2,t=1}^{\infty}\), which is defined in the proof of Proposition 1. Let us define

\[
\hat{x} \equiv \frac{\ell}{\delta(1-\beta_{2,1})(1+g)}.
\]

Player 1 chooses stage behavior \( C_1 \) with probability \( \beta_{1,1,1} \), and \( D_1 \) with probability \( 1 - \beta_{1,1,1} \) in the initial state \( \hat{\omega}_1 \). Irrespective of player 1’s action, he chooses \( m_1 = 0 \). The state remains the same if realized \( x \) is greater than \( \hat{x} \). Player 1 moves to the adjustment state \( \omega_1^A \) if player 1 chose \( C_1 \) and realized \( x \) is smaller than \( \hat{x} \). Player 1 moves to the defection state \( \omega_1^D \) if player 1 chose \( C_1 \) and realized \( x \) is smaller than \( \hat{x} \). In the adjustment state \( \omega_1^A \), player 1 chooses \( C_1 \) with probability \( 1 - \beta_{1,2} \). If player 1 chooses \( C_1 \), he chooses \( m_1 = 1 \) with probability \( 1 - \beta_{1,3} \). When player 1 chooses \( D_1 \), he never observes the opponent. The transition function in the adjustment state \( \omega_1^A \) is the same as the one in the transition state \( \omega_1^{E,2} \).
The prescribed actions and observational decisions, and state transition function are summarized in the table and figure below.

<table>
<thead>
<tr>
<th>State</th>
<th>$\hat{\omega}_1^1$</th>
<th>$\omega_1^A$</th>
<th>$\omega_1^{C,t}$</th>
<th>$\omega_1^{E,t}$</th>
<th>$\omega_1^D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action</td>
<td>$C_1$ w.p. $1 - \beta_{1,1}$</td>
<td>$C_1$ w.p. $1 - \beta_{1,2}$, $D_1$ w.p. $\beta_{1,1}$</td>
<td>Same as in strategy $\sigma^*$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_1$ given $C_1$</td>
<td>$m_1 = 0$</td>
<td>$m_1 = 1$ w.p. $1 - \beta_{1,3}$, $m_1 = 0$ w.p. $\beta_{1,3}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_1$ given $D_1$</td>
<td>$m_1 = 0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Actions and observational decisions of player 1

Player 2 chooses $D_2$ in the initial state $\hat{\omega}_1^1$. Player 2 observes player 1 with probability $1 - \beta_{2,1}$ irrespective of her action when realized $x$ is not greater than $\hat{x}$. The state remains the same if realized $x$ is smaller than $\hat{x}$. Player 2 moves to the adjustment state $\omega_1^A$ if she observes $C_1$ and realized $x$ is smaller than $\hat{x}$. Player 1 moves to the defection state $\omega_1^D$ if she observes $C_1$ and realized $x$ is smaller than $\hat{x}$. Player 2 moves to the adjustment state $\omega_1^{C,3}$ if she chooses $m_2 = 0$ and realized $x$ is smaller than $\hat{x}$. In the adjustment state $\omega_1^A$, player 2 chooses $C_2$. When player 2 chooses $C_2$, she observes player 1 with probability $1 - \beta_{2,2}$ If player 2 chooses $D_2$, she does not observe the opponent. In the transition state $\omega_2^{E,1}$, player 2 chooses $D_2$ and $m_2 = 0$ irrespective of her action. The transition functions in the adjustment state $\omega_2^A$ and the transition state $\omega_2^{E,1}$ are the same as the one in the initial state $\omega_1^1$ given strategy $\sigma^*$. That is, if player 2 observes $(C_2, C_1)$, she moves to the cooperation state $\omega_2^{C,2}$. If player 2 chooses $C_2$ but does not observe, she moves to the transition state $\omega_2^{E,2}$. When player 2 chooses $D_2$ or observes $D_1$, she moves to the defection state $\omega_2^D$.

The prescribed actions and observational decisions, and state transition function are summarized in the table and figure below.
Let strategy $\sigma^{**}$ be the strategy defined by the above automaton. Next, we define a consistent system of beliefs with strategy profile $\sigma^{**}$. We consider a sequence of behavioral strategy profiles $(\hat{\sigma}^n)_{n=1}^\infty$ such that each strategy profile attaches a positive probability to every move, but puts far greater weights on the trembles on $C_i$ in the defection state $\omega_i^D$ compared with other stage behaviors in the other states. These trembles induce a consistent system of beliefs that player $i$ at any defection state $\omega_i^D$ is sure that the state of their opponent is the defection state $\omega_j^D$ or transition state $\omega_j^{E,t}$ for some $t \geq 2$.

**Proof of Proposition 4.** Here we prove Proposition 4 using strategy $\sigma^{**}$.

Let us consider the sequential rationality of player 1. We consider the defection state $\omega_1^D$. As in the proof of Proposition 1, player 1 in the defection state $\omega_1^D$ is certain that player 2 is in the defection state $\omega_2^D$ or transition state $\omega_2^{E,t}$ for some $t \geq 1$. Therefore, it is optimal for player 1 to choose action $D_1$ and choose $m_1 = 0$ irrespective of his action.

Next, let us consider a cooperation state $\omega_1^{C,t}(t \geq 3)$. Player 1 believes that player 2 is in the cooperation state $\omega_1^{C,t-1}$ with probability $1 - \beta_{2,t-1}$ and the transition state $\omega_1^{E,t-1}$ with the remaining probability $\beta_{2,t-1}$. This is the same belief over the opponent’s state as the one that player 1 has in the cooperation state $\omega_1^{C,t-1}$ given strategy $\sigma^*$. Hence, the optimal stage behavior is also the same as the one of the cooperation state $\omega_1^{C,t-1}$ given strategy $\sigma^*$. Therefore, it is optimal for player 1 to choose $C_1$. When player 1 chooses $C_1$, player 1 is
indifferent to his observational decision. Using the same argument, the sequential rationality in the transition state $\omega_t^E,3(t \geq 3)$ is also straightforward.

Let us consider the adjustment state $\omega_t^A$. Player 2 is in the adjustment state $\omega_t^A$ with probability $1 - \beta_{2,1}$. Player 2 observes player 1 with probability $1 - \beta_{2,3}$ given action $C_2$ in the adjustment state $\omega_t^A$. Furthermore, the state transition functions in the adjustment state $\omega_t^A$ and the state transition state $\omega_t^E,1$ are the same as the one in the initial state $\omega_t^1$ given strategy $\sigma^*$. This conjecture over the continuation play of player 1 is the same as the one in the initial state $\omega_t^1$ given strategy $\sigma^*$. Therefore, player 1 is indifferent among $(C_1, 1)$, $(C_1, 0)$, and $(D_1, 0)$. When player 1 chooses $D_1$, player 1 prefers $m_1 = 0$.

Finally, let us consider the initial state $\hat{\omega}_t^1$. If player 1 chooses $D_1$, he obtains zero payoff. If player 1 chooses $C_1$ and $x \leq \hat{x}$ is realized, player 1 will move to the adjustment state. Then, choosing $(D_1, 0)$ in the adjustment state, player 1 obtains $(1 - \delta)(1 - \beta_{2,1})(1 + g)$. Therefore, the indifference condition between action $C_1$ and action $D_1$ is given by

$$0 = -\ell + \hat{x}\delta(1 - \delta)(1 - \beta_{2,1})(1 + g).$$

This condition is ensured by the definition of $\hat{x}$. In addition, $m_1 = 0$ is optimal irrespective of his actions because player 2 chooses $D_2$ with certainty. Therefore, it is optimal for player 1 to follow the strategy $\sigma^{**}$.

Next, let us consider player 2. Applying similar arguments of player 1 to states $\omega_t^D$, $\omega_t^{C,t}(t \geq 2)$, and $\omega_t^{E,t}(t \geq 2)$, we can show the sequential rationality in those states. The sequential rationality in the defection state $\omega_t^D$ is straightforward because player 2 is sure that player 1 is in the transition state or the defection state. In the cooperation state $\omega_t^{C,t}$, player 1 is in the cooperation state $\omega_t^{C,t+1}$ with probability $1 - \beta_{1,t+1}$. This belief over the continuation play of player 1 is the same as the one that player 2 has in the cooperation state $\omega_t^{C,t+1}$ given strategy $\sigma^*$. Therefore, choosing $C_2$ is optimal and player 2 is indifferent to her observational decision given $C_2$. When player 2 chooses $D_2$, she prefers $m_2 = 0$. Similarly, it is obvious that $D_2$ and $m_2 = 0$ irrespective of his action are optimal in the transition state $\omega_t^{E,t}$.

Let us consider the adjustment state $\omega_t^A$. Then, player 2 is certain that player 1 is in the adjustment state $\omega_t^A$. Then, player 1 chooses $C_1$ with probability $1 - \beta_{1,2}$, and observes player 2 with probability $1 - \beta_{1,3}$ given $C_1$. Furthermore, the state transition function of player 1 is the same as the one in the cooperation state $\omega_t^{C,2}$. The conjecture is the same as the one in the cooperation state $\omega_t^{C,2}$ given strategy $\sigma^*$. Therefore, choosing $C_2$ is optimal, and player 2 is indifferent to her observation decisions given $C_2$. When player 2 chooses $D_2$, she prefers $m_2 = 0$. We apply the same argument to the transition state $\omega_t^{E,1}$ and obtain that it is optimal for player 2 to choose $D_2$ and $m_2 = 0$ irrespective of her action.

Using similar arguments again, we can consider the initial state $\hat{\omega}_t^1$ as well. Consider observation phase after $x \leq \hat{x}$ is realized. If player 2 observes $C_1$, player 1 moves to the adjustment state $\omega_t^A$ for sure. As we confirmed before, the belief in the adjustment state $\omega_t^A$ is the same as the one in the cooperation state $\omega_t^{C,2}$ given strategy $\sigma^*$. If player 2 observes $D_1$, player 1 moves to the adjustment state $\omega_t^D$ with certainty. This conjecture is the same as the one player 2 faces in the observation phase given $C_2$ in the initial state $\omega_t^1$ given strategy $\sigma^*$. Therefore, player 2 is indifferent between $m_2 = 1$ and $m_2 = 0$. Furthermore, it is obvious that player 2 has no incentive to choose $D_2$ in the action phase in the initial state $\hat{\omega}_t^1$ because player 1 does not observe player 2 in the initial state $\hat{\omega}_t^1$. It has been proved that this strategy $\sigma^{**}$ is a sequential equilibrium.
Finally, let us consider the equilibrium payoff. It is obvious that $v^*_2$ equals zero because player 1 (weakly) prefers action $D_1$ in the initial state $\hat{\omega}^1_1$. Player 2 prefers action $D_2$ in the initial state $\hat{\omega}^2_1$. In the adjustment state $\omega^A$, one of the best responses is choosing $C_2$ and $m_2 = 0$, and the payoff is bounded below by the one of choosing $D_2$ and $m_2 = 0$. Therefore, player 2’s payoff is bounded below by

$$v^*_2 > (1 - \delta) \{(1 - \beta_{1,1})(1 + g) - \hat{x}\lambda\} + \delta \hat{x}(1 - \delta)(1 - \beta_{1,2})(1 + g) + \delta(1 - \hat{x}) v^*_2$$

$$> (1 - \delta) \{(1 - \beta_{1,1})(1 + g) - \lambda\} + \hat{x}(1 - \delta)(1 - \beta_{1,2})g + \delta(1 - \hat{x}) v^*_2$$

$$> (1 - \delta)(1 - \beta_{1,1})(1 + g + \hat{x}g) - (1 - \delta)\lambda + \delta(1 - \hat{x}) v^*_2.$$  

The second inequality holds because $\delta > \frac{g}{1 + g}$ and $\hat{x} < 1$ hold. Lemma 1 ensures $\beta_{1,2} < \beta_{1,1}$ and the third inequality.

Subtracting $\delta(1 - \hat{x}) v^*_2$ from both sides, we obtain

$$v^*_2 > \frac{(1 - 1 + \frac{1 + g + \ell \varepsilon'}{g + \ell}) \{(1 + g + \hat{x} g) - \lambda\} + \frac{1 + g + \hat{x} g - 2(1 + g)\frac{1 + g + \ell \varepsilon'}{g + \ell}}{1 + \frac{\delta}{1 - \delta} \hat{x}}}{1 + \frac{\delta}{1 - \delta} \hat{x}}.$$  

In what follows, we often use the following lemma.

**Lemma 3.** For any $y \in (0, \frac{1}{2})$, it holds that

$$1 + y < \frac{1}{1 - y} < 1 + 2y,$$

$$1 - y < \frac{1}{1 + y} < 1.$$  

**Proof of Lemma 3.** This can be shown with simple calculations. \hfill \square

Let us consider the denominator.

$$1 + \frac{\delta}{1 - \delta} \hat{x} = 1 + \frac{g + (1 + g)\varepsilon'}{1 - (1 + g)\varepsilon'} \hat{x} = 1 + \left(\frac{1 + g}{1 - (1 + g)\varepsilon'} - 1\right) \hat{x}$$

$$< 1 + \{(1 + g)(1 + 2(1 + g)\varepsilon') - 1\} \hat{x}$$

$$= 1 + \{g + 2g(1 + g)^2 \varepsilon'\} \hat{x}.$$  

Lemma 3 ensures the inequality.

The value of $\hat{x}$ is bounded above by

$$\hat{x} = \frac{\ell}{\delta(1 - \beta_{2,2})(1 + g)} < \frac{1}{1 - \frac{1 + g + \ell \varepsilon'}{g + \ell}} \frac{\ell}{g} < \left(1 + 2\frac{1 + g + \ell \varepsilon'}{g + \ell}\right) \frac{\ell}{g}.$$  

Lemma 3 ensures the last inequality. Therefore, we have an upper bound of the denominator as follows.

$$1 + \frac{\delta}{1 - \delta} \hat{x} < 1 + \{g + 2(1 + g)^2 \varepsilon'\} \left(1 + \frac{1 + g + \ell \varepsilon'}{g + \ell}\right) \frac{\ell}{g}$$

$$< 1 + \{\ell + 2(1 + g)^2 \varepsilon'\} \left(1 + \frac{1 + 2g \varepsilon'}{g}\right)$$

$$< 1 + \ell + 2(1 + g)^2 \varepsilon' + 2(1 + 2g)\varepsilon' + (1 + g)^2 \varepsilon'$$

$$< 1 + \ell + 5(1 + 2g)^2 \varepsilon'.$
The third inequality follows from Assumption 3 and $\varepsilon' < 2\varepsilon$.

Next, let us consider a lower bound of the numerator.

$$1 + g + \hat{x}g - 2(1 + g)\frac{1 + g + \ell}{g + \ell} \varepsilon' > 1 + g + \hat{x}g - 2\frac{(1 + 2g)^2}{g} \varepsilon'.$$

The value of $\hat{x}$ has the following lower bound.

$$\hat{x} > \frac{\ell}{g + (1 + g)\varepsilon'} = \frac{1}{1 + \frac{1 + g \varepsilon'}{g}} > \left(1 - \frac{1 + g \varepsilon'}{g}\right) \frac{\ell}{g} = \frac{\ell}{g} - \frac{1 + g \varepsilon'}{g}.$$

Thus, the numerator is bounded below by

$$1 + g + \left(\frac{\ell}{g} - \frac{1 + g \varepsilon'}{g}\right) g - 2\frac{(1 + 2g)^2}{g} \varepsilon' > 1 + g + \ell - 3\frac{(1 + 2g)^2}{g} \varepsilon'.$$

The last inequality is ensured by Lemma 3.

Therefore, we obtain a lower bound of $v_2^{**}$ as follows.

$$v_2^{**} > 1 + g + \ell - 3\frac{(1 + 2g)^2}{g} \varepsilon' > \frac{1 + g + \ell}{1 + \ell} \frac{1}{g(1 + g + \ell)} \varepsilon' > \frac{1 + g + \ell}{1 + \ell} \frac{1}{g} \varepsilon' > \frac{1 + g + \ell}{1 + \ell} - 8\frac{(1 + 2g)^3}{g} \varepsilon' > 1 - \varepsilon.$$ 

Let us explain why we need an interim public randomization device and why we cannot use a public randomization device at the end of the observation phase instead of interim public randomization. In our strategy, the defection state $\omega_i^D$ is an absorbing state. It is also obvious that the payoff vector of $(D_1, D_2)$ is Pareto inefficient. Therefore, to achieve a nearly Pareto-efficient outcome, the probability that each player $i$ moves to the defection state $\omega_i^D$ must be small enough. It means that the observation probability of player 2 in the initial state $\omega_2^i$ and the probability of $C_1$ in the initial state $\omega_1^i$ must be high enough.

However, taking Assumption 3 into account, player 1 has a stronger incentive to choose $C_1$ given strategy $\sigma^{**}$ than given strategy $\sigma^*$, and does not randomize actions $C_1$ and $D_1$.

To mitigate this strong incentive, we need a public randomization device. It is well known that we can decrease the efficient discount factor by dividing the game into several games (e.g., Ellison (1994)). Moving back to the initial state irrespective of stage behavior with a certain probability, player $i$ considers the continuation payoff to be less important. Let $\hat{\delta}$
be an efficient discount factor in the initial state. If player 1 chooses $D_1$ in the initial state, he obtains 0. If player 1 chooses $C_1$ in the initial state, he obtains a nonaveraged payoff $-\ell + \hat{\delta}(1 + g)$. Therefore, to make player 1 indifferent between actions $C_1$ and $D_1$ in the initial state $\hat{\omega}_1$, the efficient discount factor must be close to $\frac{\ell}{1+g}$.

It will affect not only player 1 but also player 2’s incentive. As the continuation payoff is less important, player 2’s observation incentive decreases. To keep the right-hand side of (3) unchanged, the probability $\gamma_{1,1}$ of $D_1$ in the initial state $\hat{\omega}_1$ must satisfy the following equation.

$$\delta(1 - \beta_{1,2}) = \hat{\delta}(1 - \gamma_{1,1})$$

or,

$$\gamma_{1,1} = 1 - \frac{\delta}{\delta}(1 - \beta_{1,2}).$$

Taking $\delta \sim \frac{g}{1+g}$, $\hat{\delta} \sim \frac{\ell}{1+g}$, Assumption 3, and $\beta_{1,2} \sim 0$ into account, we find that $\gamma_{1,1} \sim 1 - \frac{\delta}{g}$ is negative. Therefore, we cannot make player 2 indifferent to her observational decisions when player 1 is indifferent between actions $C_1$ and $D_1$. We need an interim public randomization device to mitigate player 1’s incentive independently of player 2’s incentive.

**Corollary 4.1.** Suppose that an interim public randomization device is available, and Assumptions 2 and 3 are satisfied. Fix any positive $\zeta > 0$. For any $\varepsilon > 0$, there exist $\delta \in \left(\frac{g}{1+g}, 1\right)$, $\bar{\delta} \in (\delta, 1)$, $\bar{\lambda} > 0$, and $\bar{\eta} > 0$ such that for any discount factor $\delta \in [\delta, \bar{\delta}]$, any observation cost $\lambda \in (0, \bar{\lambda})$, and any $\eta_1, \eta_2 \leq \bar{\eta}$, there exists a sequential equilibrium $\sigma^{**}$ whose payoff vector $(v_1^{**}, v_2^{**})$ satisfies $v_1^{**} = 0$ and $v_2^{**} \geq \frac{1+g+\ell}{1+\ell} - \varepsilon$.

**Proof of Corollary 4.1.** Let us show that strategy $\sigma^{**}$ is a sequential equilibrium if $\eta_1$ and $\eta_2$ is sufficiently small. If player $i$ is in the cooperation state $\omega_i^{C,t}$, the private signal $z_i$ has no effect on the belief of player $i$ because player $i$ directly observed player $j$’s action, $C_j$, in the previous period. In the adjustment state $\omega_i^A$, player 1 is certain that player 2 chose $D_2$ in the initial state $\hat{\omega}_2$. Hence, the private signal $z_i$ does not change the belief and best response stage-behavior of player 1 when player 1 is in the cooperation or adjustment states. In the transition or defection states, player 1 strictly prefers action $D_1$ and $m_1 = 0$ when $\eta_2 = 0$. Therefore, because of continuity of expected utility function, it is optimal for player 1 to choose action $D_1$ and $m_1 = 0$ if $\eta_2$ is sufficiently small. Thus, it is optimal for player 1 to follow strategy $\sigma^{**}$ if $\eta_2$ is sufficiently small.

Let us consider player 2. In any transition state $\omega_2^{E,t}(t \geq 1)$, player 2 strictly prefers action $D_2$ and $m_2 = 0$ when $\eta_1 = 0$. Thus, it is optimal for player 2 choose action $D_2$ and $m_2 = 0$ when $\eta_1$ is sufficiently small. In adjustment and cooperation states, the private signal $z_2$ has no effect to player 2’s belief because player 2 observed $C_1$ in the previous period. Hence, it is optimal for player 2 to follow strategy $\sigma^{**}$ if $\eta_1$ is sufficiently small. Thus, the strategy $\sigma^{**}$ is a sequential equilibrium if $\eta_1$ and $\eta_2$ are sufficiently small.

**Corollary 4.2.** Suppose that an interim public randomization device is available, and Assumptions 2 and 3 are satisfied. Fix any $\zeta > 0$. For any $\varepsilon > 0$, there exist $\delta \in \left(\frac{g}{1+g}, 1\right)$, $\bar{\lambda} > 0$, and $\bar{\eta} > 0$ such that for any discount factor $\delta \in [\delta, 1)$, any observation cost $\lambda \in (0, \bar{\lambda})$, and any $\eta_1, \eta_2 \in (0, \bar{\eta})$, there exists a sequential equilibrium $\sigma^{**}$ whose payoff vector $(v_1^{**}, v_2^{**})$ satisfies $v_1^{**} = 0$ and $v_2^{**} \geq \frac{1+g+\ell}{1+\ell} - \varepsilon$. 

20 / 29
Proof of Corollary 4.2. Use Lemma 2.

We have shown that two payoff vectors can be approximated by sequential equilibria (Propositions 1 and 4) when the discount factor is sufficiently large and the observation cost is sufficiently small. It is straightforward that a payoff vector \((\frac{1+g+\ell}{1+g}, 0)\) can be approximated by a sequential equilibrium exchanging the roles of player 1 and player 2.

Using the technique in Ellison (1994) again and alternating four strategies \(\sigma^*, \sigma^{**}\), and the repetition of the stage game Nash equilibrium, we can approximate any payoff vector in \(\mathcal{F}^*\).

Proof of Theorem 1. See Appendix C.

Remark 4. As Miyagawa et al. (2008) mentioned, some previous literature requires a very complicate strategy and a very high discount factor for their results. On the other hand, our strategy is much simpler than theirs and a required discount factor is not high. For the payoff vector \((1, 1)\) or \((\frac{1+g+\ell}{1+g}, 0)\), a slightly larger discount factor than \(\frac{g}{1+g}\) is required (See Propositions 1 and 4). If we can use a public randomization device at the beginning of the repeated game, our folk theorem holds with the same level of discount factor.

Remark 5. Let us discuss what happens if the prisoner’s dilemma is asymmetric, as in Table 6.

<table>
<thead>
<tr>
<th>Player 1</th>
<th>C1</th>
<th>1 , 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1 + g_1)</td>
<td>(-\ell_2)</td>
</tr>
<tr>
<td>D1</td>
<td>(-\ell_1)</td>
<td>(1 + g_2)</td>
</tr>
</tbody>
</table>

Table 6: Asymmetric prisoner’s dilemma

In the proofs of the propositions and theorems, we require that the discount factor \(\delta\) is sufficiently close to \(\frac{g}{1+g}\). This condition is needed to approximate a Pareto-efficient payoff vector. If \(g_1 \neq g_2\), it is impossible to ensure that the discount factor \(\delta\) is sufficiently close to both \(\frac{g_1}{1+g_1}\) and \(\frac{g_2}{1+g_2}\). Therefore, we have to confine our attention to the case of \(g_1 = g_2 = g\).

Let us consider Propositions 1 and 3. In the construction of the strategy, the randomization probability of player \(i\) is defined based on the incentive constraint of the opponent only, or, it is determined based on \(\delta, g, \ell_j\) and is independent of \(\ell_i\). Hence, if \(g_1 = g_2\) and Assumptions 2 and 3 for each \(\ell_i\) \((i = 1, 2)\) hold, our approximate efficiency result and approximate folk theorem hold under a small observation cost. Symmetricity of \(\ell_1\) and \(\ell_2\) is not important for our strategy although symmetricity of \(g_1 = g_2\) is crucial.

5 Concluding Remarks

Prisoner’s dilemma is a minimal model to describe cooperation because it has only two actions: cooperation and uncooperation. Prisoner’s dilemma under costly observation has some difficulties in cooperation.

First, the number of actions is limited. This means that players cannot communicate using a variety of actions. If more than two actions are available, we can consider an equilibrium strategy where each player randomizes some two actions on the equilibrium path. If a player has an incentive to randomize actions \(C_i\) and \(D_i\) on the path in infinitely repeated prisoner’s
dilemma, it means that the repetition of $D_i$ is one of the optimal strategies. Player $i$ loses an incentive to observe because one of his optimal strategies is unchanged irrespective of his observation result.

Second, the number of players is limited. If there are three players $A, B,$ and $C$, it is easy to check the observation deviation of the opponents. Player $A$ can monitor the observational decisions of players $B$ and $C$ by comparing their actions. If players $B$ and $C$ choose inconsistent actions toward each other, player $A$ finds that players $B$ or $C$ do not observe some of the players. Third, there is no free-cost informative signal. To obtain information about the actions chosen by their opponents, players have to observe. Despite the above limitations, we have shown our efficiency without randomization device.

We considered an interim public randomization device and obtained a folk theorem. It is worth mentioning that our folk theorem holds in some asymmetric prisoner’s dilemma. Our results might be applied to more general games.

Appendix

A Proof of Proposition 1

Proof. We prove Proposition 1. Now, let us show that the strategy profile $\sigma^*$ is a sequential equilibrium. The equilibrium payoff and the sequential rationalities in the initial, cooperation, and defection states have already been shown in Section 4. We consider the sequential rationality in the transition state $\omega^{E,t}_i$ in detail.

We consider any history in period $t$ ($\geq 2$) associated with the transition state. Strategy $\sigma^*$ prescribes $D_i$ and $m_i = 0$ irrespective of his actions in the transition state. Let us consider a nonaveraged continuation payoff when player $i$ chooses action $C_i$. Let $p$ be the belief of player $i$ that his opponent is in the cooperation state $\omega^{C,t}_j$. Therefore, if player $i$ observes his opponent in period $t$, then $(a_i^t, o_j^t) = (C_i, C_j)$ is realized with probability $p$ and the state moves to the cooperation state $\omega^{C,t+1}_i$. Let

$$W_{i,t} \equiv \{(1 - \beta_{j,t}) \cdot 1 - \beta_{j,t} \cdot \ell\} + \delta(1 - \beta_{j,t})(1 - \beta_{j,t+1})(1 + g). \tag{4}$$

The value of $W_{i,t}$ is the nonaveraged continuation payoff from the cooperation state $\omega^{C,t}_i$ when player $i$ follows strategy $\sigma^*_i$. Therefore, the upper bound of the nonaveraged payoff when player $i$ chooses action $C_i$ in period $t$ is given by

$$p - (1 - p)\ell + \delta p W_{i,t+1}.$$ 

The nonaveraged payoff when player $i$ chooses $D_i$ is bounded above by $p(1 + g)$. Therefore, action $D_i$ is profitable if the following value is negative.

$$p - (1 - p)\ell + \delta p W_{i,t+1} - p(1 + g).$$
We can rewrite the above value as follows.

\[
p - (1 - p)\ell + \delta pW_{i,t+1} - p(1 + g) = (1 - \beta_{j,t}) - \beta_{j,t}\ell - \lambda + \delta (1 - \beta_{j,t})W_{i,t+1} - (1 - \beta_{j,t})(1 + g) + \lambda + \{p - (1 - \beta_{j,t})\}\{1 + \ell + \delta W_{i,t+1} - (1 + g)\}
\]

\[
= W_{i,t} - (1 - \beta_{j,t})(1 + g) + \lambda + \{p - (1 - \beta_{j,t})\}\{\delta W_{i,t+1} - (g - \ell)\}
\]

\[
= \frac{\lambda}{\delta(1 - \beta_{j,t-1})} + \lambda + \{p - (1 - \beta_{j,t})\}\{\delta W_{i,t+1} - (g - \ell)\}. \tag{5}
\]

The second inequality follows from equation (4) in period \(t\). The last equality is ensured by (3) in period \(t - 1\).

Using equation (3), we obtain the lower bound of \(\delta W_{i,t+1} - (g - \ell)\) as follows.

\[
\delta W_{i,t+1} - (g - \ell) \geq \delta(1 - \beta_{j,t+1})(1 + g) - (g - \ell)
\]

\[
\geq \{g + (1 + g)e\prime\}\left(1 - \frac{1 + g + \ell e\prime}{g + \ell}\right) - (g - \ell)
\]

\[
\geq \frac{\ell}{2}. \tag{6}
\]

The second inequality follows from \(\beta_{i,t} \leq \frac{1 + g + \ell e\prime}{g + \ell}\) by Lemma 1. The last inequality is ensured by \(e\prime \leq 2\varepsilon\). The maximum value of \(p\) is \((1 - \beta_{j,t-1})(1 - \beta_{j,t})\). Taking (6) into account, we show that (5) is negative as follows.

\[
\frac{\lambda}{\delta(1 - \beta_{j,t-1})} + \lambda - \{(1 - \beta_{j,t}) - p\}\{\delta W_{j,t+1} - (g - \ell)\}
\]

\[
\leq \delta(1 - \beta_{j,t-1}) + \lambda - (1 - \beta_{j,t})\beta_{j,t-1}\frac{\ell}{2}
\]

\[
\leq \frac{1 + g + \ell e\prime}{g + \ell} - \lambda + \left(1 - \frac{1 + g + \ell e\prime}{g + \ell}\right)\frac{1}{2}\frac{1 + g - \ell e\prime}{g + \ell} < 0.
\]

The second inequality is ensured by \(\delta \in [\delta, \overline{\delta}]\) by Lemma 1 and \(\beta_{j,t}, \beta_{j,t-1} \in \left[\frac{1 + g - \ell e\prime}{g + \ell}, \frac{1 + g + \ell e\prime}{g + \ell}\right]\). Therefore, player \(i\) prefers \(D_i\) to \(C_i\). Hence, it has been proven that it is optimal for player \(i\) to follow strategy \(\sigma^*\). The strategy \(\sigma^*\) is a sequential equilibrium. Proposition 1 has been proved. \(\square\)

### B Proof of Lemma 1

**Proof of Lemma 1.** To prove Lemma 1, we will use the following Lemma 4 holds.

**Lemma 4.** Suppose that Assumptions 2 and 3 are satisfied. Fix any discount factor \(\delta \in [\delta, \overline{\delta}]\) and observation cost \(\lambda \in (0, \overline{\lambda})\). Then, \(\beta_{i,1} - \beta_{i,2} \geq \frac{\ell}{2g}\varepsilon\prime\) holds and, for any \(t \in \mathbb{N}\), it holds that

\[
0 < \frac{\ell}{2g} < -\frac{\beta_{i,t+2} - \beta_{i,t+1}}{\beta_{i,t+1} - \beta_{i,t}} < 1.
\]
Assume that Lemma 4 holds. Using \( \beta_{i,t}, \beta_{i,t+1}, \) and \( -\frac{\beta_{i,t+2} - \beta_{i,t+1}}{\beta_{i,t+1} - \beta_{i,t}} \), we can express \( \beta_{i,t+2} \) as follows.

\[
\beta_{i,t+2} = \beta_{i,t} + (\beta_{i,t+1} - \beta_{i,t}) + (\beta_{i,t+2} - \beta_{i,t+1}) = \beta_{i,t} + (\beta_{i,t+1} - \beta_{i,t}) \left\{ 1 - \left( \frac{-\beta_{i,t+2} - \beta_{i,t+1}}{\beta_{i,t+1} - \beta_{i,t}} \right) \right\} = \left( \frac{-\beta_{i,t+2} - \beta_{i,t+1}}{\beta_{i,t+1} - \beta_{i,t}} \right) \beta_{i,t} + \left\{ 1 - \left( \frac{-\beta_{i,t+2} - \beta_{i,t+1}}{\beta_{i,t+1} - \beta_{i,t}} \right) \right\} \beta_{i,t+1}.
\]

Therefore, if \( \beta_{i,t}, \beta_{i,t+1} \in [0, 1], \) and \( \frac{\epsilon}{2g} < -\frac{\beta_{i,t+2} - \beta_{i,t+1}}{\beta_{i,t+1} - \beta_{i,t}} < 1 \) hold, we obtain \( \beta_{i,t+2} \in (\min\{\beta_{i,t}, \beta_{i,t+1}\}, \max\{\beta_{i,t}, \beta_{i,t+1}\}) \) because \( \beta_{i,t+2} \) is a convex combination of \( \beta_{i,t} \) and \( \beta_{i,t+1} \).

Let us compare \( \beta_{i,1}, \beta_{i,2}, \) and \( \beta_{i,3}. \) By Lemma 4, \( \beta_{i,1} - \beta_{i,2} \) is greater than \( \frac{\epsilon}{g + \ell} \). Furthermore, we have \( \beta_{i,2} < \beta_{i,3} < \beta_{i,1} \) because \( -\frac{\beta_{i,t+2} - \beta_{i,t+1}}{\beta_{i,t+1} - \beta_{i,t}} \in (0, 1) \) by Lemma 4 and, then, \( \beta_{i,3} \) is a convex combination of \( \beta_{i,1} \) and \( \beta_{i,2}. \) Next, let us compare \( \beta_{i,2}, \beta_{i,3}, \) and \( \beta_{i,4}. \) As we find, \( \beta_{i,2} \) is smaller than \( \beta_{i,3}. \) Therefore, we have \( \beta_{i,2} < \beta_{i,4} < \beta_{i,3} \) because \( \beta_{i,4} \) is a convex combination of \( \beta_{i,2} \) and \( \beta_{i,3}. \) Similarly, for any \( s \in \mathbb{N}, \) it holds that \( (\beta_{i,2s} < \beta_{i,2s+1} < \beta_{i,2s-1}, \) and \( \beta_{i,2s} < \beta_{i,2s+2} (< \beta_{i,2s+1}). \)

Next, we prove Lemma 4.

**Proof of Lemma 4.** First, let us derive \( -\frac{\beta_{i,3} - \beta_{i,2}}{\beta_{i,2} - \beta_{i,1}}. \) By (1), we have

\[
0 = -(1 - \beta_{i,1})g - \beta_{i,1}\ell + \delta(1 + g)(1 - \beta_{i,1})(1 - \beta_{i,2}).
\]

Furthermore, by (3), we have

\[
\frac{\lambda}{\delta(1 - \beta_{i,1})} = -(1 - \beta_{i,2})g - \beta_{i,2}\ell + \delta(1 + g)(1 - \beta_{i,2})(1 - \beta_{i,3}).
\]

By (7) and (8), we obtain

\[
(\beta_{i,2} - \beta_{i,1})(g - \ell) - \delta(1 + g)(1 - \beta_{i,2}) \left\{ (\beta_{i,3} - \beta_{i,2}) + (\beta_{i,2} - \beta_{i,1}) \right\} = \frac{\lambda}{\delta(1 - \beta_{i,1})}.
\]

Let us consider the lower bound of \( \beta_{i,2}. \) As \( \epsilon' \in [\epsilon, 2\epsilon] \) and \( 0 < \frac{\ell}{g} < 1 \) hold, we have

\[
\beta_{i,2} = \frac{1 + g - \frac{\ell}{g} \ell - (1 + g + \ell)\frac{1 + g}{g}\epsilon'}{1 + \frac{\ell}{g}(1 + g + \ell)(\epsilon')^2 g + \ell}\epsilon' > \frac{\frac{3}{2}(1 + g - \ell)}{g + \ell}\epsilon' > \frac{1 + g - \ell}{g + \ell}\epsilon'.
\]

Next, let us consider the upper bound of \( \beta_{i,2}. \)

\[
\beta_{i,2} = \frac{1 + g - \frac{\ell}{g} \ell - (1 + g + \ell)\frac{1 + g}{g}\epsilon'}{1 + \frac{\ell}{g}(1 + g + \ell)(\epsilon')^2 g + \ell}\epsilon' < \frac{1 + g + \ell}{g + \ell}\epsilon'.
\]
The last inequality is ensured by \( \varepsilon' < 2\varepsilon \). Thus, we obtain
\[
\frac{11 + g - \ell}{2} \varepsilon' < \beta_{i,2} < \frac{1 + g}{g + \ell} \varepsilon'.
\]

As \( \beta_{i,2} < \frac{1 + g + \ell}{g + \ell} \varepsilon' < \beta_1 = \frac{1 + g + \ell}{g + \ell} \varepsilon' \), we can divide both sides of (9) by \( \beta_{i,2} - \beta_{i,1} \) and obtain

\[
\frac{\beta_{i,3} - \beta_{i,2}}{\beta_{i,2} - \beta_{i,1}} = \frac{\ell + \delta(1 + g)(1 - \beta_{i,2}) - g + \frac{\lambda}{\delta(1 - \beta_{i,1})(1 - \beta_{i,2})}}{\delta(1 + g)(1 - \beta_{i,2})}.
\]

As Assumption 3, \( \beta_{i,1} \), \( \beta_{i,2} < 1 \), and \( \beta_{i,2} - \beta_{i,1} > 0 \) hold, we find an upper bound of \( -\frac{\beta_{i,3} - \beta_{i,2}}{\beta_{i,2} - \beta_{i,1}} \).

\[
-\frac{\beta_{i,3} - \beta_{i,2}}{\beta_{i,2} - \beta_{i,1}} \leq \frac{\delta(1 + g)(1 - \beta_{i,2}) + \frac{\lambda}{\delta(1 - \beta_{i,1})(1 - \beta_{i,2})}}{\delta(1 + g)(1 - \beta_{i,2})} < 1.
\]

Taking \( \beta_{i,1} = \frac{1 + g + \ell}{g + \ell} \varepsilon' \), \( \beta_{i,2} < \frac{1 + g + \ell}{g + \ell} \varepsilon' \), and \( -(\beta_{i,2} - \beta_{i,1}) > \frac{\ell}{g + \ell} \varepsilon' > \frac{\ell}{2g} \varepsilon' \) into account, we have a lower bound of \( -\frac{\beta_{i,3} - \beta_{i,2}}{\beta_{i,2} - \beta_{i,1}} \) as follows.

\[
-\frac{\beta_{i,3} - \beta_{i,2}}{\beta_{i,2} - \beta_{i,1}} > \frac{\ell + g \left( 1 - \frac{1 + g + \ell}{g + \ell} \varepsilon' \right) - g - \frac{4(1 + g)}{1 + g} \left( 1 - \frac{1 + g}{g + \ell} \varepsilon' \right)}{\frac{\ell}{g} \left( 1 + g \varepsilon' \right) (1 + g)}\]
\[
> \frac{\ell - \frac{1 + g + \ell}{g + \ell} g \varepsilon' - \frac{4(1 + g)}{1 + g} \left( 1 - \frac{1 + g}{g + \ell} \varepsilon' \right)}{\frac{\ell}{g} \left( 1 + g \varepsilon' \right) (1 + g)} > \frac{3}{2} \frac{\ell}{g} > \frac{\ell}{2g}.
\]

The first inequality follows from \( \delta = \frac{g}{1 + g} + \varepsilon' > \frac{g}{1 + g} \). The third inequality is ensured by \( \varepsilon' < 2\varepsilon \) and \( \lambda < \lambda' \). Therefore, we have obtained \( \frac{\ell}{2g} < \frac{\beta_{i,3} - \beta_{i,2}}{\beta_{i,2} - \beta_{i,1}} < 1 \) and \( \beta_{i,3} \in (\beta_{i,2}, \beta_{i,2}) \). That is, \( \beta_{i,3} - \beta_{i,2} > 0 \).

Next, let us derive \( -\frac{\beta_{i,s+3} - \beta_{i,s+2}}{\beta_{i,s+2} - \beta_{i,s+1}} \) inductively. Suppose that \( \frac{\ell}{2g} < -\frac{\beta_{i,s+2} - \beta_{i,s+1}}{\beta_{i,s+1} - \beta_{i,s}} < 1 \) and \( \beta_{i,s+2} \in (\min \{ \beta_{i,s}, \beta_{i,s+1} \}, \max \{ \beta_{i,s}, \beta_{i,s+1} \}) \) hold for period \( s = 1, 2, \ldots, t \). We have shown that this supposition holds for \( t = 1 \). We show that \( \frac{\ell}{2g} < \frac{\beta_{i,t+3} - \beta_{i,t+2}}{\beta_{i,t+2} - \beta_{i,t+1}} < 1 \) and \( \beta_{i,t+3} \in (\min \{ \beta_{i,t+1}, \beta_{i,t+2} \}, \max \{ \beta_{i,t+1}, \beta_{i,t+2} \}) \) hold.

By (3) for \( t + 1 \) and \( t + 2 \), we have
\[
\begin{cases}
\frac{\lambda}{\delta(1 - \beta_{i,t})} = -(1 - \beta_{i,t+1})g - \beta_{i,t+1} \ell + \delta(1 - \beta_{i,t+1})(1 - \beta_{i,t+2})(1 + g), \\
\frac{\lambda}{\delta(1 - \beta_{i,t+1})} = -(1 - \beta_{i,t+2})g - \beta_{i,t+2} \ell + \delta(1 - \beta_{i,t+2})(1 - \beta_{i,t+3})(1 + g),
\end{cases}
\]
or,

\[
-\frac{\beta_{i,t+1} - \beta_{i,t}}{\delta(1 - \beta_{i,t})(1 - \beta_{i,t+1})} \lambda
\]
\[
= -(\beta_{i,t+2} - \beta_{i,t+1})(g - \ell) + \delta(1 - \beta_{i,t+2}) \{(\beta_{i,t+3} - \beta_{i,t+2}) + (\beta_{i,t+2} - \beta_{i,t+1})\} (1 + g).
\]

The suppositions ensure \( \beta_{i,t+2} - \beta_{i,t+1} \neq 0 \). Divide both sides of the above equation by \( \beta_{i,t+2} - \beta_{i,t+1} \). Then, we obtain
\[
-\frac{\beta_{i,t+3} - \beta_{i,t+2}}{\beta_{i,t+2} - \beta_{i,t+1}} = \frac{\ell + \delta(1 + g)(1 - \beta_{i,t+2}) - g + \frac{1}{\delta(1 - \beta_{i,t})(1 - \beta_{i,t+1})(1 - \beta_{i,t+2})} \frac{\beta_{i,t+2} - \beta_{i,t+1}}{\beta_{i,t+1} - \beta_{i,t}}}{\delta(1 + g)(1 - \beta_{i,t+2})}.
\]
As Assumption 3 and \( \frac{\beta_{i,t+3} - \beta_{i,t+2}}{\beta_{i,t+2} - \beta_{i,t+1}} < 0 \) hold, \( \frac{\beta_{i,t+3} - \beta_{i,t+2}}{\beta_{i,t+2} - \beta_{i,t+1}} \) is bounded above by

\[
\frac{\beta_{i,t+3} - \beta_{i,t+2}}{\beta_{i,t+2} - \beta_{i,t+1}} = \beta_{i,t+3} - \beta_{i,t+2} \leq \frac{\delta(1 + g)(1 - \beta_{i,t+2})}{\delta(1 + g)(1 - \beta_{i,t+2})} < 1.
\]

Taking 0 < \( \beta_{i,t+1}, \beta_{i,t+2} < \frac{1 + g + \ell}{g + \ell} \) = \( \beta_{i,1} \), and \( \frac{\ell}{2g} < -\frac{\beta_{i,t+3} - \beta_{i,t+2}}{\beta_{i,t+2} - \beta_{i,t+1}} < 1 \) into account, we find the following lower bound of \( -\frac{\beta_{i,t+3} - \beta_{i,t+2}}{\beta_{i,t+2} - \beta_{i,t+1}} \).

\[
-\frac{\beta_{i,t+3} - \beta_{i,t+2}}{\beta_{i,t+2} - \beta_{i,t+1}} = \frac{\ell + \delta(1 - \beta_{i,t+2})(1 + g) - g + \frac{1}{\delta(1 - \beta_{i,t+2})(1 - \beta_{i,t+1})\beta_{i,t+2} - \beta_{i,t+1}}}{\delta(1 + g)(1 - \beta_{i,t+2})} > \frac{\ell + g(1 + g + \ell \varepsilon') - g - \frac{1}{g + (1 + g)(1 + g + \ell \varepsilon')^2} g}{g + (1 + g)(1 + g + \ell \varepsilon')} > \frac{3 \ell}{2g}.
\]

Therefore, we obtain \( \frac{\ell}{2g} < -\frac{\beta_{i,t+3} - \beta_{i,t+2}}{\beta_{i,t+2} - \beta_{i,t+1}} < 1 \) and \( \beta_{i,t+3} \in \{ \min \{ \beta_{i,t+1}, \beta_{i,t+2} \}, \max \{ \beta_{i,t+1}, \beta_{i,t+2} \} \} \).

C Proof of Theorem 1

Proof. Let us fix \( \pi \) such that:

\[
\pi \geq \frac{4 + 2g}{\varepsilon}.
\]

We use the same technique as in Lemma 2. We divide the repeated game into \( \pi \) distinct repeated games. The first repeated game is played in period 1, \( \pi + 1, 2\pi + 1 \ldots \), the second repeated game is played in period 2, \( \pi + 1, 2\pi + 2 \ldots \), and so on. Each repeated game can be regarded as a repeated game with discount factor \( \delta^\pi \).

We can find a sequential equilibrium strategy \( \hat{\sigma}^\pi \) whose payoff vector \( \hat{v}^\pi = (v_1^\pi, v_2^\pi) \) satisfies \( |\hat{v}_i^\pi - 1| \leq \frac{1}{\pi} \) when discount factor \( \delta^\pi \) is sufficiently large by Proposition 3. By Corollary 4.2, there exists a sequential equilibrium strategy \( \hat{\sigma}^{**} \) whose payoff vector \( \hat{v}^{**} = (v_1^{**}, v_2^{**}) \) satisfies \( \hat{v}_1^{**} = 0 \) and \( |\hat{v}_2^{**} - \frac{1 + g + \ell}{1 + g + \ell} | < \frac{1}{\pi} \) when discount factor \( \delta^{**} \) is sufficiently large.

Let us assume that \( v_1^\pi \leq v_2^\pi \). We choose sufficiently large discount factor \( \delta \) so that we can use Proposition 4 and Corollary 4.2, and the discount factor \( \delta \) satisfies the following condition:

\[
\frac{1 - \delta}{1 - \delta^\pi} < \frac{2}{\pi}.
\]

The desired payoff vector \( v \) can be expressed uniquely as a convex combination of \( \hat{v}^\pi, \hat{v}^{**} \) and \( (0, 0) \) with some \( \alpha_1, \alpha_2 \in (0, 1) \) as below.

\[
v = \alpha_1 \hat{v}^\pi + \alpha_2 \hat{v}^{**} + (1 - \alpha_1 - \alpha_2) \cdot 0.
\]
Let us define $n_1$ and $n_2$ as follows.

$$n_1 \equiv \arg \min_{n \in \mathbb{N} \cup \{0\}} \left| \frac{1 - \delta^n}{1 - \delta \pi} - \alpha_1 \right|, \quad n_2 \equiv \arg \min_{n \in \mathbb{N} \cup \{0\}} \left| \frac{\delta^{n_1} - \delta^{n_1+n_2}}{1 - \delta \pi} - \alpha_2 \right|.$$  

Then, $n_1$ and $n_2$ satisfy

$$\left| \frac{1 - \delta^{n_1}}{1 - \delta \pi} - \alpha_1 \right| \leq \left( \frac{1 - \delta}{1 - \delta \pi} \right) \frac{2}{n}, \quad \left| \frac{\delta^{n_1} - \delta^{n_1+n_2}}{1 - \delta \pi} - \alpha_2 \right| \leq \frac{2}{n}.$$  

Let us consider the following strategy $\sigma^{F}$. In the first $n_1$-th games, players play strategy $\hat{\sigma}^*$. From the $n_1 + 1$-th game to the $n_1 + n_2$-th game, players play strategy $\hat{\sigma}^{**}$. From the $n_1 + n_2 + 1$-th to $n$-th game, players play the stage game Nash equilibrium repetitively. It is straightforward that the strategy $\sigma^{F}$ is a sequential equilibrium.

The payoff $v_i^{F}$ for player $i$ is given by

$$v_i^{F} = \frac{(1 - \delta^{n_1})\hat{v}_i^* + (\delta^{n_1} - \delta^{n_1+n_2})\hat{v}_i^{**} + (\delta^{n_1+n_2} - \delta \pi) \cdot 0}{1 - \delta \pi}.$$  

Therefore, we have

$$|v_i^{F} - v| < \left| \frac{1 - \delta^{n_1}}{1 - \delta \pi} \hat{v}_i^* - \alpha_1 \hat{v}_i^* \right| + \left| \frac{\delta^{n_1} - \delta^{n_1+n_2}}{1 - \delta \pi} \hat{v}_i^{**} - \alpha_2 \cdot v_i^{**} \right| + 0$$

$$< \frac{2}{n} \cdot 1 + \frac{2}{n} \cdot (1 + g) = \frac{4 + 2g}{n} < \varepsilon.$$  

We obtain that the payoff vector $v$ can be approximated by a sequential equilibrium payoff vector when $v_1 \leq v_2$ holds.

By symmetricity of the payoff matrix, it is straightforward that the payoff vector $v$ can be approximated by a sequential equilibrium payoff vector when $v_1 \geq v_2$ also holds.  

\begin{enumerate}
\item Let us define $n_1$ and $n_2$ as follows.
\item Then, $n_1$ and $n_2$ satisfy
\item Let us consider the following strategy $\sigma^{F}$. In the first $n_1$-th games, players play strategy $\hat{\sigma}^*$. From the $n_1 + 1$-th game to the $n_1 + n_2$-th game, players play strategy $\hat{\sigma}^{**}$. From the $n_1 + n_2 + 1$-th to $n$-th game, players play the stage game Nash equilibrium repetitively. It is straightforward that the strategy $\sigma^{F}$ is a sequential equilibrium.
\item The payoff $v_i^{F}$ for player $i$ is given by
\item Therefore, we have
\item We obtain that the payoff vector $v$ can be approximated by a sequential equilibrium payoff vector when $v_1 \leq v_2$ holds.
\item By symmetricity of the payoff matrix, it is straightforward that the payoff vector $v$ can be approximated by a sequential equilibrium payoff vector when $v_1 \geq v_2$ also holds.  
\end{enumerate}

\textbf{References}


