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

Hino, Yoshifumi

Vietnam-Japan University

10 December 2018

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

Yoshifumi Hino†

This version: September 13, 2019

Abstract

We consider an infinitely repeated prisoner's dilemma under costly monitoring. If a player observes his opponent, then he pays an observation cost and knows the action chosen by his opponent. If a player does not observe his opponent, he cannot obtain any information about his opponent's action. Furthermore, no player can statistically identify the observational decision of his opponent. We prove efficiency without any signals. Then, we extend the idea with a public randomization device and 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 now standard insight in the theory of repeated games is that repetition enables players to obtain collusive and efficient outcomes in a repeated game. 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 or monitor his opponent's previous action at a (small) cost and a player's monitoring decision is unobservable to his opponent. We establish an approximate efficient result together with a corresponding 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, 129060, Vietnam
Tel: +84 24 7306 6001
Email: y.hino@vju.ac.vn

In our model, we consider costly monitoring as a monitoring structure. Each player chooses his action and 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 cannot obtain any information about his opponent in that period if he chooses not to observe that player.

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 without any signals under some assumptions regarding the payoff matrix when players are patient and the observation cost is small (efficiency). The second result is a type of folk theorem. We introduce a public randomization device. The public randomization device is realized at the end of each stage game, and players see the public randomization device without any cost. We present a folk theorem with a public randomization device under some assumptions regarding the payoff matrix when players are patient and the observation cost is small. The first result shows that cooperation is possible without any signals or communication in the venture company example. The second result implies that companies need a coordination device to achieve an asymmetric cooperation in the venture company example.

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.

As with Chen (2010), our analysis is tractable. By construction, the only concern of each player at each period is whether his opponent is in the cooperation state. It is sufficient to keep track of this belief, which is the probability that the opponent is in the cooperation state.

Our main contribution is the efficiency result and folk theorem in an infinitely repeated prisoner's dilemma. Some previous studies show that the efficiency result holds if tools to share information are available. For example, certain studies assume that some information is available even if players do not observe their opponent. We describe these tools and discuss previous studies in Section 2 before we define our model in Section 3. Our efficiency result shows that players can construct a cooperative relationship without any randomization device.

Another contribution of the paper is a new approach to the construction of a sequential equilibrium. We consider randomization of monitoring, whereas previous studies confine their attention to randomization of actions. In most cases, the observational decision is supposed to be unobservable in costly monitoring 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 monitoring. Section 3 introduces a repeated prisoner's dilemma model with costly monitoring. We present our main idea and results in Section 4, including an efficiency result with a small observation cost. We extend our main idea with a public randomization device and present a folk theorem in Section 5. Section 6 provides concluding remarks.

2 Literature Review

We only review previous studies on repeated games with costly monitoring.

One of the biggest difficulties in costly monitoring is monitoring the monitoring activity of opponents, because observational behavior under costly monitoring is often assumed to be unobservable. Each player has to check this unobservable monitoring 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); Lehrer and Solan (2018) assume that players can observe other players' observational decisions.

Next approach is communication. 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 a communication through actions, and they call it implicit communication. They assume 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 implicit communication and present a folk theorem for any level of observation cost.

Another approach is introduction of nonpublic randomization device. The nonpublic randomization device enables players randomize actions even though they are certain that both players are in the cooperation state. Hino (2019) shows that if 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 no costless information is available, then cooperation is difficult. Two other papers present folk theorems without costless information. Flesch and Perea (2009) consider monitoring 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 monitoring 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 implicit communication through mixed actions. To use implicit communication 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 whether players can achieve an efficiency result and a folk theorem in 2-action games, even though the monitoring cost is sufficiently small. We show an efficiency result without any randomization device using a grim trigger strategy and mixed monitoring rather than mixed actions when observation cost is small. Since we use a grim trigger strategy, players have no incentive to observe the opponent once after the punishment has started. In addition, in our strategy, players choose different actions when they are in the cooperation state and when they are in another state. Therefore, the observation in the current period gives player i enough information to check whether the punishment starts or not. It means that our efficiency result holds under the same structure in Flesch and Perea (2009) or Miyagawa et al. (2003). 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))

Consider the bilateral trade game with moral hazard in Bhaskar and van Damme (2002) simplified by Miyagawa et al. (2003).

		Player 2 C_2	D_2	E_2
	C_1	1 , 1	-1 , 2	-1 , -1
Player 1	D_1	2 , -1	0 , 0	-1 , -1
	E_1	-1 , -1	-1 , -1	0 , 0

Table 1: Extended prisoner's dilemma

Players choose whether he observes the opponent or not at the same time with his action choice. Miyagawa et al. (2003) consider the following keep-them-guessing grim trigger strategies to approximate payoff vector $(1, 1)$. There are three states: cooperation, punishment, and defection. In the defection state, both players choose E_i, and the state remains the same. In the punishment state, both players choose E_i for some periods, and then the state moves back to a cooperation state. In both the punishment state and the defection state, the players do not observe their opponent. In the cooperation state, each player chooses C_i with sufficiently high probability and chooses D_i with the remaining probability. Players observe their opponent in the cooperation state. If players observe (C_1, C_2) or (D_1, D_2), the state remains the same. The state moves to the defection state if player i chooses E_i or observes E_j. When (C_1, D_2) or (D_1, C_2) is realized, the state moves to the punishment state.

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 the state of the opponent in the next period. If the state of the opponent is the cooperation state, then action E_i is a suboptimal action because the opponent never chooses action E_j. That is, choosing action E_i has some opportunity cost because the state of the opponent is the cooperation state with a high probability. However, if the state of the opponent is the defection state, then action E_i is a unique optimal action. Choosing actions C_i or D_i also has opportunity costs because the state of the opponent is 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. Let us consider the prisoner's dilemma as an example. If players randomize C_i and D_i in the cooperation state, then their best response action includes action D_i at any history. As a result, choosing D_i and not observe player j every period is one of the best response strategy. The strategy fails to give players monitoring incentive.

I consider the following strategy. 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 out of the cooperation state and chooses D_i.

Using this strategy, we show an efficiency result without any randomization device, and we extend it with public randomization device and present a folk theorem.

The reason why our strategy works in a two-action game is that the strategy prescribes different actions based on the observation result. The strategy prescribes action C_i (resp., D_i) when player i observes action $o_i = C_i$ (resp., $o_i = D_j$). Hence, player i does not randomize actions C_i and D_i in each period except for the initial period, and the above-mentioned problem does not occur. The above-mentioned problem does not happen in the initial because there is no previous period of the initial period and player i does not observe the opponent in the previous period.

However, it causes another problem related to the monitoring incentive. As player j does not randomize his action, player i can easily guess player j's action through past observation. For example, if player i chooses C_i and observed C_j in the previous period, player i can guess that player j's action will be C_j. Then, player i loses the monitoring incentive again.

Our strategy can overcome this difficulty as well. Since player j randomize his observational decision in the cooperation state and it is unobservable, player i in the cooperation state cannot know whether player j observed player i or not. Suppose that player i chooses C_i in the previous period. Then, if player j chooses C_j and observed player i in the previous period, player j is in the cooperation state in the current period and chooses C_j. Otherwise, player j chooses D_j in the current period. Therefore, from the viewpoint of player i, it looks as if player j randomizes actions C_j and D_j, which gives player i an incentive to observe. This is why player i has an incentive to observe player j given our strategy.

3 Model

The base game is a symmetric prisoner's dilemma. 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. Given an action profile (a_1, a_2), the base game payoff for player i, $u_i(a_1, a_2)$, is displayed in Table 2.

		Player 2: C_2	Player 2: D_2
Player 1	C_1	1 , 1	$-\ell$, $1+g$
	D_1	$1+g$, $-\ell$	0 , 0

Table 2: Prisoner's dilemma

We make the usual assumptions about the above payoff matrix.

Assumption 1. (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 2. $g - \ell > 0$.

Assumption 2 is the same as Assumption 1 in Chen (2010).

The stage game is of simultaneous form. Each player i chooses an action a_i and the observational decision simultaneously. 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 incurs 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 the pair of base game actions 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 the pair b_1 and 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

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

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 average discounted stage game payoff is given by

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

During the repeated game, players does not receive any free signals regarding player actions (no free signal). It implies that a player receives no information about the action chosen by his opponent when he does not observe the opponent. This implies that no player receives the base game payoffs in the course of play. As in Miyagawa et al. (2003), we interpret the discount factor as the probability with which the repeated game continues, and it is assumed that each player receives the sum of the payoffs when the repeated game ends. Then, the assumption of no free signals regarding actions is less problematic.

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

Let h_i^t be the (private) history of player i at the beginning of period $t \geq 2$: $h_i^t = (a_i^k, o_i^k)_{k=1}^{t-1}$. This history is a sequence of his own actions and observation results up to period $t - 1$. We omit the observational decisions from h_i^t because observation result o_i^k implies the observational decision m_i^k for any k. Let H_i^t

denote the set of all the histories for player i at the beginning of period $t \geq 1$, where H_i^1 is an arbitrary singleton set.

A (behavior) strategy for player i in the repeated game is a function of the history of player i to his (mixed) stage behavior.

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

A strategy profile σ is a pair of strategies σ_1 and σ_2. Given a strategy profile σ, a sequence of completely mixed behavior strategy profiles $(\sigma^n)_{n=1}^{\infty}$ that converges to σ is called a *tremble*. Each completely mixed behavior strategy profile σ^n induces a unique system of beliefs ψ^n.

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

4 No public randomization

In this section, we show our efficiency result without any randomization device. The following proposition shows that the symmetric efficient outcome is approximated by a sequential equilibrium if the observation cost λ is small and the discount factor δ is moderately low.

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

Proof. See Appendix A. □

An illustration

While the proof in Appendix A provides the detailed construction of an equilibrium that approximates the Pareto-efficient payoff vector, we here present its main idea.

Let us consider the following four-state automaton: initial state ω_i^1, cooperation state $(\omega_i^t)_{t=2}^{\infty}$, transition state ω_i^E, and defection state ω_i^D. In the initial state ω_i^1, player i randomizes three stage-behaviors: $(C_i, 1)$, $(C_i, 0)$, and $(D_i, 0)$. Player i chooses $(C_i, 1)$ with sufficiently high probability. In the cooperation state $\omega_i^t (t \geq 2)$, player i chooses C_i and randomizes the observational decision. Player i chooses $(C_i, 1)$ with sufficiently high probability. In the transition state and defection state ω_i^D, player i chooses $(D_i, 0)$.

The state transition is described in Figure 1.

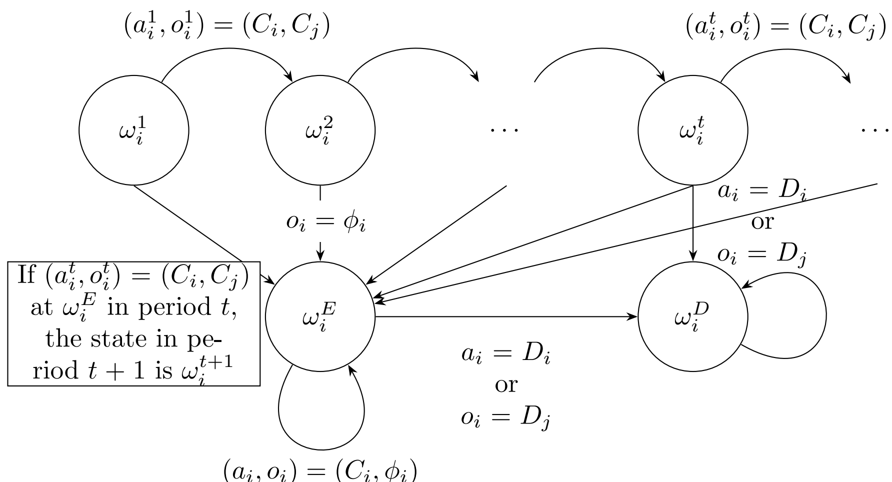

Figure 1: The state-transition rule

That is, a player remains in the cooperation state only when he chooses C_i and observes C_j. Player i moves to the defection state if he chooses D_i or observes D_j. If player i does not observe his opponent in the cooperation state, he moves to the transition state. Although, the stage-behavior in the transition state is the same as that in the defection state, the transition function is not. Player i moves back to the cooperation state from the transition state if he observes (C_i, C_j), which is the event off the equilibrium path. Let strategy σ^* be the strategy above.

Another property of this strategy is that players never randomize their actions in the cooperation state, whereas players randomize their observational decisions in the cooperation state. As a result, the player looks as if he randomizes actions from the viewpoint of his opponent although he chooses pure actions in each state. Furthermore, we show in the appendix that player i strictly prefers action C_i in the cooperation state although he strictly prefers action D_i in the defection state. This creates a monitoring incentive in order to know which action is optimal.

Let us consider sequential rationality in each state. First, we consider the defection state. A sufficient condition for sequential rationality in the defection state is that player i is certain that the state of the opponent is the transition or defection state. It implies that player i is sure that player j chooses D_j and never observes player i. Hence, player i does not have an incentive to choose C_i or $m_i = 1$ in the defection state on the equilibrium path.

The defection state occurs only when player i chooses D_i or observed D_j in the past. Therefore, this sufficient condition is obvious on the equilibrium path.

The defection state might be realized off the equilibrium path. Then, it is not obvious whether this sufficient condition holds off the equilibrium path or not. Let us consider the following history in period 3. Player i chooses $a_i = D_i$ and $m_i = 0$ in period 1, and he chooses $a_i = C_i$ and observes C_j in period 2. We can consider the following history. Player j chooses $a_j = C_j$ and $m_j = 0$ in period 1, and he chooses $a_j = C_j$ (by mistakes) and observes C_j (by mistake) in period 2. Then, player j is in the cooperation state in period 3.

To obtain the desired belief, we consider the same belief as Miyagawa et al. (2008). That is, 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 smaller weights on the trembles with respect to the observational decisions compared with those with respect to actions[1]. These trembles induce a consistent system of beliefs that player i at any defection state is sure that the state of their opponent is the defection state. It ensures the sufficient condition.

Let us discuss sequential rationality in the cooperation state. First of all, we consider how to find the sequence of randomization probabilities of stage-behaviors. We define the sequence as follows. We fix small probability of choosing D_j in the initial probability. Then, we can derive the payoff for player i when player i choose D_i in the initial period based on probability of D_j. Player i must be indifferent between C_i and D_i in the initial period. The continuation payoff from the cooperation state in period 2 is the continuation payoff when player i chooses C_i and does not observe the opponent in period 2. It is the function of the probabilities of choosing D_j and of monitoring in the initial period. We choose monitoring probability in the initial period to make player i indifferent between C_i and D_i. The similar argument holds in period 2 onwards. The continuation payoff from the cooperation state in period $t \geq 3$ is the continuation payoff when player i chooses C_i and does not observe the opponent in period t. It is the function of the probabilities of monitoring in the previous and current period. In period 2, we know the monitoring probability in the initial period. Therefore, we choose the monitoring probability in the period 2 so that player i is indifferent between observing and not observing in period 2. In period $t (\geq 3)$, we know the monitoring probability in the previous period, and we choose the monitoring probability in period t so that player i is indifferent between observing and not observing in period t.

Then, the sequential rationality in the initial and cooperation states is satisfied as follows. In the initial state, player i is indifferent between C_i and D_i, and is indifferent between observing and not observing because of the randomization probability of D_i and the definition of the monitoring probability in the initial state. It is suboptimal that choosing D_i but observing the opponent because player i is certain that the opponent is no longer in the cooperation state irrespective his observational result. In period $t \geq 2$, player i is indifferent between observing and not observing by the selection of the monitoring probability of player j in the current period. It is also suboptimal to choose D_i and observe the opponent as the same as in the initial period. Furthermore, this definition

[1] See Miyagawa et al. (2008) for further discussion.

ensures that player i strictly prefers action C_i to D_i in the cooperation state. Suppose that player i weakly prefers action D_i in the next cooperation state. Then, choosing action D_i is the best response action irrespective of his observation result. This means that player i strictly prefers $m_i = 0$ because he can save the observation cost by choosing $(C_i, 0)$ in the current period and $(D_i, 1)$ in the next period. Hence, player i strictly prefers action C_i in the next cooperation state after he chooses C_i and observes C_j in the cooperation state. Therefore, sequential rationality is satisfied in the cooperation state.

Next, let us consider the transition state. We show why player i prefers action D_i in the transition state. In the transition state, there are two types of situations. Situation A is where player i is in the cooperation state if he observed his opponent in the previous period. Situation B is where player i is in a defection state if he observed his opponent in the previous period. Of course, player i cannot distinguish between these two situations because he did not observe his opponent. To understand sequential rationality in the transition state, let us assume that the monitoring cost is almost zero. This means that the deviation payoff to $(D_i, 0)$ in the cooperation state is sufficiently close to the continuation payoff from the cooperation state. Otherwise, player i strictly prefers to observe in the cooperation state to avoid choosing action D_i in situation A. Therefore, player i is almost indifferent between choosing C_i and D_i in situation A, whereas player i strictly prefers action D_i in situation B. Hence, player i strictly prefers action D_i when the observation cost is sufficiently small because both situations are realized with a positive probability.

Third, let us consider the initial state. The indifference condition between C_i and D_i is ensured by the randomization probability between $(C_i, 1)$ and $(C_i, 0)$ in the initial state. If the monitoring probability in the cooperation state is high enough, then player i is willing to choose action C_i. The indifference condition between $(C_i, 1)$ and $(C_i, 0)$ in the initial state is ensured by the randomization probability between $(C_i, 1)$ and $(C_i, 0)$ in the cooperation state in period 2. There is no incentive to choose $(D_i, 1)$ because the state of the opponent in the next period is not the cooperation state for sure, irrespective of the observation.

Lastly, let us consider the payoff. It is obvious that the equilibrium payoff vector is close to 1 if the probabilities of $(C_i, 1)$ in the initial state and cooperation state are close to 1 and the observation cost is close to 0. In Appendix A, we show that the equilibrium payoff vector is close to 1 when the discount factor is close to $\frac{g}{1+g}$. Another remaining issue is whether our strategy is well-defined, which is shown by solving a difference equation of monitoring constraints in Appendix A when Assumption 2 is satisfied.

We extend Proposition 1 using Lemma 1.

Lemma 1. *Fix any payoff vector v and any $\varepsilon > 0$. Suppose that there exist $\underline{\delta} \in \left(\frac{g}{1+g}, 1\right)$, $\overline{\delta} \in (\underline{\delta}, 1)$ such that for any discount factor $\delta \in [\underline{\delta}, \overline{\delta}]$, 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 exist $\underline{\delta}^* \in (g/1+g, 1)$ such that for any discount factor $\delta \in [\underline{\delta}^*, 1)$, 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 1 . We use the technique of Lemma 2 in Ellison (1994). We define $\underline{\delta}^* \equiv \underline{\delta}/\overline{\delta}$, and choose any discount factor $\delta \in (\underline{\delta}^*, 1)$. Then, we choose some integer n^* that satisfies $\delta^{n^*} \in [\underline{\delta}, \overline{\delta}]$. Then there exists a strategy σ' whose payoff vector is (v_1^*, v_2^*) given δ^{n^*}. We divide the repeated game into n^* distinct repeated games. The first repeated game is played in period 1, $n^* + 1$, $2n^* + 1 \dots$, the second repeated game is played in period 2, $n^* + 1$, $2n^* + 2 \dots$, and so on. Each repeated game can be regarded as a repeated game with discount factor δ^{n^*}. Let us consider the following strategy σ^{L1}. In the 1st game, players follow strategy σ'. In the 2nd game, players follow strategy σ'. In the $n(n \leq n^*)$th game, players follow strategy σ'. Then, strategy σ^{L1} will be a sequential equilibrium. This is because strategy σ' 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 σ^{L1} also satisfies $|v_i^* - v_i| \leq \varepsilon$ for each $i = 1, 2$. □

We obtain efficiency for a sufficiently high discount factor.

Proposition 2. *Suppose that the base game satisfies Assumptions 1 and 2. For any $\varepsilon > 0$, there exist $\underline{\delta}^* \in (0,1)$ and $\overline{\lambda} > 0$ such that for any discount factor $\delta \in (\underline{\delta}^*, 1)$ and any $\lambda \in (0, \overline{\lambda})$, 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 2 . Apply Lemma 1 to Proposition 1. □

Remark 1. Proposition 2 shows monotonicity of efficiency on the discount factor. If efficiency holds given ε, observation cost λ and discount factor δ, then efficiency holds given a sufficiently large discount factor $\delta' > \delta$.

Lastly, let us consider what happens if Assumption 2 is not satisfied. Let β_1 be the probability that player i chooses $a_i = D_i$ in the first period. Let β_{t+1} be the probability that player i chooses $m_i = 0$ in the cooperation state ω_i^t. Then, the following proposition holds.

Proposition 3. *Suppose that Assumption 1 is satisfied, but 2 is not satisfied. Then, $(\beta_t)_{t=1}^{\infty}$ is not well defined for a small observation cost λ.*

Proof of Proposition 3 . See Appendix C. □

Proposition 3 implies that our strategy is not well-defined when 2 is not satisfied. Let us explain Proposition 3. A relationship between β_t and β_{t+1} is determined by the monitoring incentive and $\frac{\ell}{g}$. We show that the absolute value of β_t goes to infinity as t goes to infinity when $\frac{\ell}{g} > 1$.

Let us derive the relationship. If player i chooses C_i and observes C_j in state ω_i^{t-1}, his continuation payoff from period t is given by

$$(1 - \delta)\left\{(1 - \beta_t) \cdot 1 - \beta_t \ell\right\} + \delta(1 - \delta)(1 - \beta_t)(1 - \beta_{t+1})(1 + g).$$

On the other hand, if player i chooses C_i but does not observe player j, his continuation payoff from period t is

$$(1-\delta)(1-\beta_t)(1+g).$$

Therefore, only when the following equation holds, player i is indifferent between $m_i = 0$ and $m_i = 1$ in state ω_i^{t-1}.

$$\frac{\lambda}{\delta(1-\beta_{t-1})} = -g + \beta_t(g-\ell) + \delta(1-\beta_t)(1-\beta_{t+1})(1+g). \tag{1}$$

Rewards for choosing $m_i = 1$ comes from two parts: $-g + \beta_t(g-\ell)$ and $\delta(1-\beta_t)(1-\beta_{t+1})(1+g)$ in (1). Reward $\delta(1-\beta_t)(1-\beta_{t+1})(1+g)$ implies that when β_t is too small (resp., large), β_{t+1} must be large (resp., small) in order to keep the right-hand side in (1) unchanged. Reward $-g+\beta_t(g-\ell)$ implies that when $g-\ell > 0$ (resp., $g-\ell < 0$), an increase in β_t strengthens (resp., weakens) the monitoring incentive. This is because increase in β_t implies decrease of $(1-\beta_t)g$ and makes choosing D_i in the next period less attractive. Increase in β_t also implies decrease of $-\beta_t\ell$ and makes choosing C_i in the next period less attractive. Therefore, whether larger β_t strengthens the monitoring incentives or not depends on whether Assumption 2 holds or not.

Assume that Assumption 2 holds, then these two rewards has opposite effect. Suppose that β_t is high. Then, the reward $\delta(1-\beta_t)(1-\beta_{t+1})(1+g)$ prescribes smaller β_{t+1} to maintain the monitoring incentive. On the other hand, reward $-g+\beta_t(g-\ell)$ strengthens the monitoring incentive and prescribes higher β_{t+1} in order to reduce $\delta(1-\beta_t)(1-\beta_{t+1})$ and to weaken the monitoring incentive.

On the other hands, if Assumption 2 does not hold, then these two rewards have the same effect. Given high β_t, the reward $\delta(1-\beta_t)(1-\beta_{t+1})(1+g)$ prescribes smaller β_{t+1} to maintain the monitoring incentive. Reward $-g+\beta_t(g-\ell)$ weakens the monitoring incentive and prescribes smaller β_{t+1} in order to strengthen the monitoring incentive. As a result, the sequence $(\beta_t)_{t=1}^{\infty}$ will be a divergent sequence.

Let us show this fact by an approximation. Let define $\varepsilon' \equiv \delta - \frac{g}{1+g}$ and assume that ε' is sufficiently small. Then, assuming ε', β_t, and β_{t+1} are small, and λ is sufficiently small compared to ε', we can ignore λ, $\beta_t \cdot \beta_{t+1}$, $\varepsilon'\beta_t$ and $\varepsilon'\beta_{t+1}$. Then, we obtain

$$0 = -\beta_t\ell + (1+g)\varepsilon' - \beta_{t+1}g$$

or,

$$\beta_{t+1} - \frac{1+g}{g+\ell}\varepsilon' = -\frac{\ell}{g}\left(\beta_t - \frac{1+g}{g+\ell}\varepsilon'\right) \tag{2}$$

The above equation (2) shows the relationship between β_t and β_{t+1}.

The ratio $\frac{\ell}{g}$ determines how large β_{t+1} must be in order to satisfy the monitoring incentive. The equation (2) says that when Assumption 2 does not hold (i.e., $\frac{\ell}{g} > 1$), the absolute value of $\beta_{t+2} - \frac{1+g}{g+\ell}\varepsilon'$ must be larger than the absolute value of $\beta_{t+1} - \frac{1+g}{g+\ell}\varepsilon'$. As a result, the absolute value of $\beta_t - \frac{1+g}{g+\ell}\varepsilon'$ goes to infinity as t goes to infinity. More precise calculation without approximation can be found in Appendix C.

Remark 2. Let us refer to σ^* in the proof of Proposition 1 and σ^{L1} in Lemma 1 as keep-them-monitoring grim trigger strategy. Proposition 1 and Proposition 3 shows a sufficient and necessary condition for keep-them-monitoring grim trigger strategy. A keep-them-monitoring grim trigger strategy is a sequential equilibrium for a sufficiently large discount factor and a sufficiently small observation cost if and only if Assumption 2 is satisfied.

5 Public randomization

In this section, we assume the public randomization device is available at the end of each stage game. The distribution of the public signal is independent of the action profile chosen. Public signal x is uniformly distributed over $[0, 1)$ and each player observes the public signal without any cost.

The purpose of this section is to prove a folk theorem. To prove Theorem 1 (folk theorem), we present the following proposition first.

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

Proof of Proposition 4 . See Appendix D. □

An illustration

Here we explain our strategy and why we need a public randomization device for our result.

In our strategy, the players play (C_1, D_2) in the first period and only player 2 observe, and then play the strategy in the proof of Proposition 1 from period 2 onward. Applying the strategy in Section 4, let us consider the following strategy. At the initial state, player 1 randomizes actions C_1 and D_1, whereas player 2 chooses D_2. Player 2 randomly observes player 1. The state transition depends on public randomization. If realized x is greater than $\hat{x}$ in the initial state, the remains the same. If realized x is not greater than $\hat{x}$ in the initial state, the state transition depends on the stage-behaviors. If player 1 plays D_1 (resp., player 2 observes D_1) in the initial state, then he moves to the defection state in period 2 onward. If player 1 play C_1 (resp., player 2 observes C_1) in the initial state, then he moves to the cooperation state and play a sequential equilibrium whose payoff vector is sufficiently close to $(1, 1)$, which is similar to the one in Section 4. We show in Appendix D that our sequential equilibrium strategy is similar to that above. If player 2 observes nothing in the initial state, she moves to the transition state.

Let us consider sequential rationalities of players. The sequential rationality in defection state both on and off the equilibrium path holds in the same manner in the Section 4. The sequential rationality in the cooperation state from period 3 on holds as well.

Let us consider the sequential rationality of player 1 in the cooperation state in period 2. Player 1 cannot distinguish whether the state of the opponent is cooperation state or not because the observational decision is unobservable. If player 2 observes in the previous period, he chooses C_2. Otherwise, player 2 chooses D_2. Therefore, from the viewpoint of player 1, player 2 randomizes three stage-behavior: $(C_2, 1)$, $(C_2, 0)$, and $(D_2, 0)$ like the initial state in the proof of Proposition 1. Hence, if player 2 chooses appropriate randomization probability of $(C_2, 1)$, $(C_2, 0)$, and $(D_2, 0)$, then player 1 is indifferent between $(C_1, 1)$, $(C_1, 0)$, and $(D_1, 0)$. Next, let us consider the sequential rationality of player 2 in the cooperation state in period 2. As player 1 randomizes $(C_1, 1)$, $(C_1, 0)$, and $(D_1, 0)$, it is easily satisfied when player 1 chooses appropriate randomization probability.

Let us consider the sequential rationality in period 1. As Assumption 2 is satisfied, the deviation to action D_i is more profitable in terms of the stage game payoff when the opponent chooses C_j than when the opponent chooses D_j. The incentive for player 1 to choose C_1 is higher than the one in the proof of Proposition 1. Therefore, we use an public randomization device to decrease the incentive to choose action C_1.

Let us explain why we need a public randomization device in more detail. Let $\beta_{2,2}$ be the probability that player 2 chooses $m_2 = 0$ in the initial state. Then, without public randomization device, player 1 is indifferent between C_1 and D_1 when the following equation holds.

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

The left-hand (resp., right-hand) side is the payoff when player 1 chooses C_1 (resp., D_1). Our strategy σ^* in the proof of Proposition 1 will be a sequential equilibrium only when discount factor is slightly greater than $\frac{g}{1+g}$ and Assumption 2 holds. However, when discount factor is slightly greater than $\frac{g}{1+g}$ and Assumption 2 holds and $\beta_{2,2}$ is close to zero, the payoff when player 1 chooses C_1 is higher than zero. As a result, player 1 strictly prefers C_1 and does not randomize C_1 and D_1.

We can consider two approaches to this problem. The first one is choosing high $\beta_{2,2}$. However, it means that player 2 chooses D_2 in period 2 and (D_1, D_2) will be played with a high probability from period 3. We cannot approximate the Pareto efficient payoff vector $\left(0, \frac{1+g+\ell}{1+\ell}\right)$.

Let us consider what happens if we decrease an efficient discount factor only in period 1 using public randomization. To satisfy the indifference condition, the discount factor must be close to $\frac{\ell}{1+g}$ $\left(< \frac{g}{1+g}\right)$. It is well known that we can decrease the efficient discount factor dividing the game into several games (e.g., Ellison (1994)). However, there is no technique to increase the discount factor without public randomization device as game proceeds. Therefore, we need public randomization device to use a smaller efficient discount factor in the initial state. Public randomization device is indispensable because the discount factor must increase when players moves out of the initial state. The sequential rationality of player 2 holds as well because player 1 randomizes C_1 and D_1 with moderate probability and monitoring cost is sufficiently small. Therefore, the strategy will be a sequential equilibrium.

The last issue is the equilibrium payoff. Given this strategy, we have to consider the effect of public randomization device to the equilibrium payoff. Let V_i be the payoff for player i for each $i = 1, 2$. In the proof of Proposition 1, we have shown that Pareto efficient payoff vector $(1, 1)$ can be approximated by a sequential equilibrium when the discount factor is close to $\frac{g}{1+g}$. Therefore, the continuation payoff when player 1 moves to cooperation state in period 2 is close to 1. The value of $\hat{x}$ can be approximated as the solution of the following equation.

$$-(1-\delta)\ell + \delta\hat{x}\cdot 1 + \delta(1-\hat{x})V_1 = (1-\delta)\cdot 0 + \delta\hat{x}\cdot 0 + \delta(1-\hat{x})V_1$$

The left-hand side is the payoff when player 1 chooses C_1, and the right-hand side is the one when he chooses D_1. Therefore, $\hat{x}$ is close to $\frac{1-\delta}{\delta}\ell$. Then, the payoff V_2 of player 2 can be approximated by the following equation.

$$\begin{aligned} V_2 =& (1-\delta)(1+g) + \delta\hat{x}\cdot 1 + \delta(1-\hat{x})V_2 \\ =& \frac{(1-\delta)(1+g) + \delta\hat{x}\cdot 1}{1-\delta+\delta\hat{x}} \\ =& \frac{1+g+\ell}{1+\ell} \end{aligned}$$

We have obtained the desired result.

Corollary 4.1. *Suppose that a public randomization device is available, and the base game satisfies Assumptions 1 and 2. For any* $\varepsilon > 0$*, there exist* $\underline{\delta} \in \left(\frac{g}{1+g}, 1\right)$ *and* $\overline{\lambda} > 0$ *such that for any discount factor* $\delta \in [\underline{\delta}, 1)$ *and for any observation cost* $\lambda \in (0, \overline{\lambda})$*, there exists a sequential equilibrium whose payoff vector* (v_1^*, v_2^*) *satisfies* $v_1^* = 0$ *and* $v_2^* \geq \frac{1+g+\ell}{1+\ell} - \varepsilon$*.*

Proof of Corollary 4.1 . Use Lemma 1. □

We have shown that two types of 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 to show that payoff vector $\left(\frac{1+g+\ell}{1+\ell}, 0\right)$ can be approximated by a sequential equilibrium exchanging the roles of player 1 and player 2. Let us denote this strategy by $\hat{\sigma}$, and let us denote the strategy used in the proof of Proposition 4 by $\tilde{\sigma}$.

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

Theorem 1 (Approximate folk theorem)**.** *Suppose that a public randomization is available, and Assumptions 1 and 2 are satisfied. Fix any interior point* $v = (v_1, v_2)$ *of* $\mathcal{F}^*$*. Fix any* $\varepsilon > 0$*. There exist a discount factor* $\underline{\delta} \in \left(\frac{g}{1+g}, 1\right)$ *and observation cost* $\overline{\lambda} > 0$ *such that for any* $\delta \in [\underline{\delta}, 1)$ *and* $\lambda \in (0, \overline{\lambda})$*, there exists a sequential equilibrium whose payoff vector* $v^F = (v_1^F, v_2^F)$ *satisfies* $|v_i^F - v_i| \leq \varepsilon$*.*

Proof of Theorem 1. See Appendix E. □

Remark 3. Our approach can be applied under the monitoring structure of Flesch and Perea (2009) if a public randomization device is available. Our strategy is a variant of the grim trigger strategy. It means that players have no incentive to observe the opponent once after the punishment starts. In addition, in our strategy, player i chooses C_i and choose D_i if he is not in the cooperation state. Therefore, the observation in the current period gives player i enough information to check whether the punishment starts or not. Therefore, each player does not have an incentive to acquire information about past actions.

Remark 4. We have proven efficiency and the folk theorem in a repeated symmetric prisoner's dilemma. In this section, we discuss what happens if the prisoner's dilemma is asymmetric, as in Table 3.

		Player 2	
		C_2	D_2
Player 1	C_1	1 , 1	$-\ell_1$, $1+g_2$
	D_1	$1+g_1$, $-\ell_2$	0 , 0

Table 3: Asymmetric prisoner's dilemma

In the proofs of the propositions and theorems below, we require that the discount factor δ is sufficiently close to $\frac{g}{1+g}$. This condition is required to approximate a Pareto-efficient payoff vector. If $g_1 \neq g_2$, it is impossible to ensure that the discount factor δ 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 2. In the construction of the strategy, the randomization probability of player i is defined based on the incentive constraint of the opponent only. In other words, the randomization probability is determined independently of the payoffs of player i. This means that the randomization probability of player i is determined based on δ, g, ℓ_j and is independent of ℓ_i. Therefore, we can discuss the randomization probabilities of player 1 and 2 independently. Hence, if $g_1 = g_2$ and Assumptions 1 and 2 for each ℓ_i $(i = 1, 2)$ hold, our efficiency result and folk theorem under a small observation cost hold.

6 Concluding Remarks

Although only a few possible types of cooperation exist in a two-player, two-action prisoner's dilemma, prisoner's dilemma under costly monitoring is still a useful model to understand cooperation. Prisoner's dilemma under costly monitoring has some properties. First, the number of actions is limited. This means that players cannot communicate using a variety of actions. Second, the number of players is limited. If there are three players A, B, 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.

Originally, the prisoner's dilemma under costly monitoring has these constraints. Despite the above limitations, we have shown efficiency without any randomization device. Our paper is the first result to show that efficiency holds without any randomization device under an infinitely repeated prisoner's dilemma with costly monitoring, although it is the simplest model among those with costly monitoring considered in the literature (e.g., Miyagawa et al. (2003) and Flesch and Perea (2009)).

We considered a 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 can be applied to more general games.

A Proof of Proposition 1

Proof. We prove Proposition 1.

Strategy

We define a grim trigger strategy σ^*, and then we define a consistent system of beliefs ψ^*. Strategy σ^* is represented by an automaton that has four types of states: initial state ω_i^1, cooperation states $(\omega_i^t)_{t=2}^{\infty}$, transition state ω_i^E, and defection state ω_i^D. For any period $t \geq 2$, there is a unique cooperation state. Let ω_i^t be the cooperation state in period $t \geq 2$.

At the initial state ω_i^1, each player i chooses $(C_i, 1)$ with probability $(1-\beta_1)(1-\beta_2)$, chooses $(C_i, 0)$ with probability $(1-\beta_1)\beta_2$, and chooses $(D_i, 0)$ with probability β_1. We call (a_i, o_i) an action–observation pair. The state moves from the initial state to the cooperation state ω_i^2 if the action–observation pair in period 1 is (C_i, C_j). The state moves to the transition state ω_i^E in period 2 when (C_i, ϕ_i) is realized in period 1. Otherwise, the state moves to a defection state in period 2.

At the cooperation state ω_i^t, each player i chooses $(C_i, 1)$ with probability $1-\beta_{t+1}$ and $(C_i, 0)$ with probability β_{t+1}. That is, the randomization probability β_{t+1} depends on calendar time t. The state moves to the next cooperation state ω_i^{t+1} if the action–observation pair in period t is (C_i, C_j). The state moves to the transition state ω_i^E in period $t+1$ when (a_i^t, o_i^t) or (C_i, ϕ_i) is realized in period t. Otherwise, the state moves to the defection state in period $t+1$.

At the transition state ω_i^E in period t, each player i chooses $(D_i, 0)$ with certainty. The state moves to the defection state ω_i^D in period $t+1$ when $a_i^t = D_i$ or $o_i^t = D_j$ is realized. If player i chooses $(C_i, 0)$, the state remains the same. When player i chooses C_i and observes C_j in period t, the state in period $t+1$ moves to the cooperation state ω_i^{t+1}.

Players choose $(D_i, 0)$ and the state remains the same ω_i^D at the defection state ω_i^D, irrespective of the action–observation pair.

The state-transition rule is summarized in Figure 1. Let strategy σ^* be the strategy represented by the above automaton.

We define a system of beliefs consistent with strategy σ^* by the same approach as that in Section 4. Each behavioral strategy profile $\hat{\sigma}^n$ induces the system of beliefs ψ^n, and the consistent system of beliefs ψ^* is defined as the limit of $\lim_{n\to\infty} \psi^n$.

Selection of discount factor and observation cost

Fix any $\varepsilon > 0$. We define $\overline{\varepsilon}$, $\underline{\delta}$, $\overline{\delta}$, and $\overline{\lambda}$ as follows

$$\begin{aligned}
\overline{\varepsilon} \equiv& \frac{\ell^2}{54(1+g+\ell)^3} \frac{\varepsilon}{1+\varepsilon}, \\
\underline{\delta} \equiv& \frac{g}{1+g} + \overline{\varepsilon}, \\
\overline{\delta} \equiv& \frac{g}{1+g} + 2\overline{\varepsilon} < 1, \\
\overline{\lambda} \equiv& \frac{1}{16} \frac{g}{1+g} \frac{1}{1+g+\ell} \overline{\varepsilon}^2.
\end{aligned}$$

We fix an arbitrary discount factor $\delta \in [\underline{\delta}, \overline{\delta}]$ 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 $v_i^* \geq 1 - \varepsilon$ for each $i = 1, 2$.

Specification of strategy

Let us define $\varepsilon' \equiv \delta - \frac{g}{1+g}$. We set $\beta_1 = \frac{1+g+\ell}{g+\ell} \varepsilon'$. Given β_1, we define β_2 as the solution of the following indifference condition between $(C_i, 0)$ and $(D_i, 0)$ in period 1.

$$(1-\beta_1) \cdot 1 - \beta_1 \cdot \ell + \delta(1-\beta_1)(1-\beta_2)(1+g) = (1-\beta_1)(1+g). \tag{3}$$

Next, we define $(\beta_t)_{t=3}^{\infty}$. We choose β_{t+2} so that player j at state ω_i^t is indifferent between choosing $(C_i, 1)$ and $(C_i, 0)$.

To define $\beta_t (t \geq 3)$, let W_t $(t \geq 1)$ be the sum of the stage game payoffs from the cooperation state ω_i^t. That is, payoff W_t is given by

$$W_t = \left[\sum_{s=t}^{\infty} \delta^{s-1} u_i(a^s) \middle| \sigma^*, \psi^*, \omega_i^t \right].$$

Please note that W_t $(t \geq 1)$ is determined uniquely. There are several histories associated with the cooperation state ω_i^t (e.g., the ones where player i cooperated and observed cooperation in the transition state in the previous period). At any of those histories, player i believes that player j is at the cooperation state ω_j^t with probability $1 - \beta_t$ and at the transition state with probability β_t. Therefore, the continuation payoff W_t is uniquely determined.

At the cooperation state $\omega_i^t (t \geq 2)$, player i weakly prefers to play $(C_i, 0)$. Therefore, payoff W_t is given by

$$W_t = (1 - \beta_t) \cdot 1 - \beta_t \ell + \delta(1 - \beta_t)(1 - \beta_{t+1})(1 + g), \quad \forall\, t \geq 2. \tag{4}$$

Therefore, payoff W_t is a function of (β_t, β_{t+1}). We denote payoff W_t by $W_t(\beta_t, \beta_{t+1})$ when we should consider W_t as a function of (β_t, β_{t+1}).

Then, β_3 is given by

$$W_1 = (1 - \beta_1) \cdot 1 - \beta_1 \ell - \lambda + \delta(1 - \beta_1) W_2(\beta_2, \beta_3). \tag{5}$$

Next, let us consider the indifference condition between $(C_i, 1)$ and $(C_i, 0)$ at the cooperation state $\omega_i^t (t \geq 2)$. Let us consider the belief for each player i at the cooperation state ω_i^t in period t. Assume that $\beta_t \in (0, 1)$ for any $t \in \mathbb{N}$, which is proved later. Then, we show by mathematical induction that, for any period $t \geq 2$, player i at the cooperation state ω_i^t in period t believes that the state of his opponent is a cooperation state with positive probability $1 - \beta_t$. Let us consider period $t = 2$ first. The state moves to the cooperation state ω_i^2 in period 2 only when player i has observed the action–observation pair $(a_i^1, o_i^1) = (C_i, C_j)$ in period 1. Therefore, player i believes that the state of his opponent is the cooperation state with positive probability $1 - \beta_2$ by Bayes' rule. Thus, the statement is true in period 2. Next, suppose that the statement is true until period t and consider player i at the cooperation state ω_i^{t+1}. This means that player i has observed action–observation pair $(a_i^t, o_i^t) = (C_i, C_j)$ in period t. Thus, player i believes with certainty that player j was in the cooperation state in period t. Therefore, he believes that player j is the in the cooperation state with positive probability $1 - \beta_{t+1}$ by Bayes' rule. Hence, the statement is true.

Taking the belief at the cooperation state $\omega_i^t (t \geq 2)$ into account, β_{t+2} is defined as the solution of the equation below.

$$W_t(\beta_t, \beta_{t+1}) = (1 - \beta_t) \cdot 1 - \beta_t \ell - \lambda + \delta(1 - \beta_t) W_{t+1}(\beta_{t+1}, \beta_{t+2}). \tag{6}$$

Specifically, β_2 is defined by (3), and β_{t+2} ($t \in \mathbb{N}$) is defined by (6) as follows.

$$\begin{aligned}
\beta_2 =& \frac{(1-\beta_1)\{\delta(1+g)-g\} - \beta_1 \ell}{\delta(1-\beta_1)(1+g)} \\
=& \frac{g+g^2-\ell^2-(1+g+\ell)(1+g)\varepsilon'}{(g+\ell)\{g+(1+g)\varepsilon'\}\left(1-\frac{1+g+\ell}{g+\ell}\varepsilon'\right)}\varepsilon' \\
=& \frac{1+g-\frac{\ell}{g}\ell-(1+g+\ell)\frac{1+g}{g}\varepsilon'}{1+\frac{\ell}{g}\frac{1}{g+\ell}\varepsilon'-\frac{(1+g)(1+g+\ell)}{g(g+\ell)}(\varepsilon')^2}\frac{1}{g+\ell}\varepsilon' \\
\beta_{t+2} =& \frac{(1-\beta_{t+1})\{\delta(1+g)-g\}-\beta_{t+1}\ell-\frac{\lambda}{\delta(1-\beta_t)}}{\delta(1-\beta_{t+1})(1+g)}, \quad \forall\ t \in \mathbb{N}.
\end{aligned}$$

Now, to focus on a game theoretic argument, we assume the following Lemma 2, which is proved in Appendix B.

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

$$\frac{1}{2}\frac{1+g-\ell}{g+\ell}\varepsilon' < \beta_2 < \beta_4 < \beta_6 \cdots < \beta_5 < \beta_3 < \beta_1 = \frac{1+g+\ell}{g+\ell}\varepsilon'.$$

Thanks to Lemma 2, we obtain a lower bound and an upper bound of β_t for any $t \in \mathbb{N}$.

Now, let us show that the grim trigger strategy σ^* is a sequential equilibrium.

Sequential rationality at the initial state

At the initial state, the indifference condition between $(C_i, 0)$ and $(D_i, 0)$ is ensured by the construction of β_2. The indifference condition between $(C_i, 1)$ and $(C_i, 0)$ is ensured by the construction of β_3. Furthermore, if player i chooses action D_i, then his opponent chooses action D_j with certainty from the next period on, irrespective of his observation result. Thus, player i has no incentive to choose $(D_i, 1)$. Therefore, it is optimal for player i to follow strategy σ^* at the initial state.

Sequential rationality in the cooperation state

Next, consider a history associated with the cooperation state ω_i^t in period t (≥ 2). Then, strategy σ^* prescribes to randomize $(C_i, 1)$ and $(C_i, 0)$. As we explained earlier, at any history associated with a cooperation state in period t (≥ 2), player i believes that player j is at the cooperation state ω_j^t with probability $1-\beta_t$ and at the transition state with probability β_t. The definition of β_{t+2} ensures that $(C_i, 1)$ and $(C_i, 0)$ are indifferent for player i in period t. When player i chooses $(D_i, 0)$ or $(D_i, 1)$, then the continuation payoff is bounded above by $(1-\beta_t)(1+g)$. Equation (6) implies that, for any $t \in \mathbb{N}$, it holds that

$$W_{t+1} = (1-\beta_{t+1})(1+g) + \frac{\lambda}{\delta(1-\beta_t)}. \tag{7}$$

The above equality ensures that, for any period $t \geq 1$, $(1-\beta_{t+1})(1+g)$ is strictly smaller than W_{t+1}, which is the payoff when player i chooses $(C_i, 1)$ in period $t+1$. Thus, both $(D_i, 0)$ and $(D_i, 1)$ are suboptimal in any period $t \geq 2$. Therefore, it is optimal for player i to follow strategy σ^* in the cooperation state.

Sequential rationality at the defection state

Consider any history associated with the defection state. Then, σ^* prescribes $(D_i, 0)$. Since we consider the belief construction similar to the one in Miyagawa et al. (2008), player i believes that player j never deviates from prescribed observational decision as the same as Miyagawa et al. (2008). Therefore, player i is certain that the state of his opponent is either the transition or defection state, and player i's action in that period does not affect the continuation play of his opponent. Furthermore, player i believes that player j chooses action D_j with certainty and has no incentive to observe his opponent. Therefore, it is optimal for player i to follow strategy σ^* in the defection state.

Sequential rationality in the transition state

We consider any history in period t (≥ 2) associated with the transition state. Strategy σ^* prescribes $(D_i, 0)$ in the transition state.

Let us consider a continuation payoff when player i chooses action C_i in period t. Let p be the belief of player i in the transition state in period t that his opponent is in the cooperation state. If player i observes his opponent, then $(a_i^t, o_i^t) = (C_i, C_j)$ is realized with probability p and the state moves to the cooperation state (ω_i^{t+1}). The continuation payoff in the cooperation state in period $t+1$ is bounded above by W_{t+1}. This is because W_{t+1} is a continuation payoff when player i chooses action C_i from ω_i^{t+1}, and W_{t+1} is strictly greater than payoff $(1-\beta_{t+1})(1+g)$, which is the upper bound of the payoff when player i chooses action D_i at ω_i^{t+1}. Therefore, the upper bound of the non-averaged payoff when player i chooses action C_i in period t is given by

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

The non-averaged 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.

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

We can rewrite the above value as follows.

$$
\begin{aligned}
&p-(1-p)\ell+\delta p W_{t+1}-p(1+g)\\
=&(1-\beta_t)-\beta_t\ell-\lambda+\delta(1-\beta_t)W_{t+1}-(1-\beta_t)(1+g)\\
&\quad+\lambda+\{p-(1-\beta_t)\}\{1+\ell+\delta W_{t+1}-(1+g)\}\\
=&W_t-(1-\beta_t)(1+g)+\lambda+\{p-(1-\beta_t)\}\{\delta W_{t+1}-(g-\ell)\}\\
=&\frac{\lambda}{\delta(1-\beta_{t-1})}+\lambda+\{p-(1-\beta_t)\}\{\delta W_{t+1}-(g-\ell)\}. \qquad (8)
\end{aligned}
$$

The second equality follows from equation (6) in period t. The last equality is ensured by (7) in period $t-1$.

Using equation (7), we obtain the lower bound of $\delta W_{t+1}-(g-\ell)$ as follows.

$$
\begin{aligned}
\delta W_{t+1}-(g-\ell)\geq&\delta(1-\beta_{t+1})(1+g)-(g-\ell)\\
\geq&\{g+(1+g)\varepsilon'\}\left(1-\frac{1+g+\ell}{g+\ell}\varepsilon'\right)-(g-\ell)\\
\geq&\frac{\ell}{2}. \qquad (9)
\end{aligned}
$$

The second inequality follows from $\beta_t\leq\frac{1+g+\ell}{g+\ell}\varepsilon'$ by Lemma 2. The last inequality is ensured by $\varepsilon'\leq 2\overline{\varepsilon}$. The maximum value of p is $(1-\beta_{t-1})(1-\beta_t)$. Taking (9) into account, we show that (8) is negative as follows.

$$
\begin{aligned}
&\frac{\lambda}{\delta(1-\beta_{t-1})}+\lambda-\{(1-\beta_t)-p\}\{\delta W_{t+1}-(g-\ell)\}\\
\leq&\frac{\lambda}{\delta(1-\beta_{t-1})}+\lambda-(1-\beta_t)\beta_{t-1}\frac{\ell}{2}\\
\leq&\frac{1+g}{g}\frac{1}{1-\frac{1+g+\ell}{g+\ell}\varepsilon'}\lambda+\lambda-\left(1-\frac{1+g+\ell}{g+\ell}\varepsilon'\right)\frac{1}{2}\frac{1+g-\ell}{g+\ell}\varepsilon'\frac{\ell}{2}\\
<&0.
\end{aligned}
$$

The second inequality is ensured by $\delta\in[\underline{\delta},\overline{\delta}]$ by Lemma 2 and $\beta_t,\beta_{t-1}\in\left[\frac{1}{2}\frac{1+g-\ell}{g+\ell}\varepsilon',\frac{1+g+\ell}{g+\ell}\varepsilon'\right]$. Therefore, player i prefers D_i to C_i. Hence, it has been proven that it is optimal for player i to follow strategy σ^*. The strategy σ^* is a sequential equilibrium.

The payoff

Finally, we show that the sequential equilibrium payoff v_i^* is strictly greater than $1-\varepsilon$. Player i chooses $(D_i,0)$ in period 1 at the initial state. Therefore, the equilibrium payoff v_i^* is given by

$$
v_i^*=(1-\delta)(1-\beta_1)(1+g)=\{1-(1+g)\varepsilon'\}\left(1-\frac{1+g+\ell}{g+\ell}\varepsilon'\right)>1-\varepsilon.
$$

Therefore, Proposition 1 has been proven. □

B Proof of Lemma 2

Proof of Lemma 2. To prove Lemma 2, we will use the following Lemma 3 holds.

Lemma 3. *Suppose that Assumptions 1 and 2 are satisfied. Fix any discount factor $\delta \in [\underline{\delta}, \overline{\delta}]$ and observation cost $\lambda \in (0, \overline{\lambda})$. Then, $\beta_1 - \beta_2 \geq \frac{\ell}{g+\ell}\varepsilon'$ holds and, for any $t \in \mathbb{N}$, it holds that*

$$0 < \frac{\ell}{2g} < -\frac{\beta_{t+2} - \beta_{t+1}}{\beta_{t+1} - \beta_t} < 1.$$

Assume that Lemma 3 holds. Using β_t, β_{t+1}, and $-\frac{\beta_{t+2}-\beta_{t+1}}{\beta_{t+1}-\beta_t}$, we can express β_{t+2} as follows.

$$\begin{aligned}
\beta_{t+2} =& \beta_t + (\beta_{t+1} - \beta_t) + (\beta_{t+2} - \beta_{t+1}) \\
=& \beta_t + (\beta_{t+1} - \beta_t)\left\{1 - \left(-\frac{\beta_{t+2} - \beta_{t+1}}{\beta_{t+1} - \beta_t}\right)\right\} \\
=& \left(-\frac{\beta_{t+2} - \beta_{t+1}}{\beta_{t+1} - \beta_t}\right)\beta_t + \left\{1 - \left(-\frac{\beta_{t+2} - \beta_{t+1}}{\beta_{t+1} - \beta_t}\right)\right\}\beta_{t+1}.
\end{aligned}$$

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

Let us compare β_1, β_2, and β_3. By Lemma 3, $\beta_1 - \beta_2$ is greater than $\frac{\ell}{g+\ell}\varepsilon'$. Furthermore, we have $\beta_2 < \beta_3 < \beta_1$ because $-\left(-\frac{\beta_{t+2}-\beta_{t+1}}{\beta_{t+1}-\beta_t}\right) \in (0,1)$ by Lemma 3 and, then, β_3 is a convex combination of β_1 and β_2. Next, let us compare β_2, β_3, and β_4. As we find, β_2 is smaller than β_3. Therefore, we have $\beta_2 < \beta_4 < \beta_3$ because β_4 is a convex combination of β_2 and β_3. Similarly, for any $s \in \mathbb{N}$, it holds that $(\beta_{2s} <)\beta_{2s+1} < \beta_{2s-1}$, and $\beta_{2s} < \beta_{2s+2} (< \beta_{2s+1})$.

□

Next, we prove Lemma 3.

Proof of Lemma 3. First, let us derive $-\frac{\beta_3-\beta_2}{\beta_2-\beta_1}$. By (3), we have

$$0 = -(1-\beta_1)g - \beta_1\ell + \delta(1+g)(1-\beta_1)(1-\beta_2). \tag{10}$$

Furthermore, by (4) and (5), we have

$$\frac{\lambda}{\delta(1-\beta_1)} = -(1-\beta_2)g - \beta_2\ell + \delta(1+g)(1-\beta_2)(1-\beta_3) \tag{11}$$

By (10) and (11), we obtain

$$(\beta_2 - \beta_1)(g - \ell) - \delta(1+g)(1-\beta_2)\left\{(\beta_3 - \beta_2) + (\beta_2 - \beta_1)\right\} = \frac{\lambda}{\delta(1-\beta_1)}. \tag{12}$$

Let us consider the lower bound of β_2. As $\varepsilon' \in [\bar{\varepsilon}, 2\bar{\varepsilon}]$ and $0 < \frac{\ell}{g} < 1$ hold, we have

$$\begin{aligned}\beta_2 =& \frac{1+g-\frac{\ell}{g}\ell-(1+g+\ell)\frac{1+g}{g}\varepsilon'}{1+\frac{\ell}{g}\frac{1}{g+\ell}\varepsilon'-\frac{(1+g)(1+g+\ell)}{g(g+\ell)}(\varepsilon')^2}\frac{1}{g+\ell}\varepsilon' \\ >& \frac{\frac{3}{4}(1+g-\ell)}{\frac{3}{2}}\frac{1}{g+\ell}\varepsilon > \frac{1}{2}\frac{1+g+\ell}{g+\ell}\varepsilon'.\end{aligned}$$

Next, let us consider the upper bound of β_2.

$$\begin{aligned}\beta_2 =& \frac{1+g-\frac{\ell}{g}\ell-(1+g+\ell)\frac{1+g}{g}\varepsilon'}{1+\frac{\ell}{g}\frac{1}{g+\ell}\varepsilon'-\frac{(1+g)(1+g+\ell)}{g(g+\ell)}(\varepsilon')^2}\frac{1}{g+\ell}\varepsilon' \\ <& \frac{1+g-\frac{\ell}{g}}{1-\frac{(1+g)(1+g+\ell)}{g(g+\ell)}(\varepsilon')^2}\frac{1}{g+\ell}\varepsilon' \\ <& \frac{1+g-\frac{\ell}{g}}{1-\frac{(1+g)(1+g+\ell)}{g(g+\ell)}\varepsilon'}\frac{1}{g+\ell}\varepsilon' < \frac{1+g}{g+\ell}\varepsilon'.\end{aligned}$$

The last inequality is ensured by $\varepsilon' < 2\bar{\varepsilon}$. Thus, we obtain

$$\frac{1}{2}\frac{1+g-\ell}{g+\ell}\varepsilon' < \beta_2 < \frac{1+g}{g+\ell}\varepsilon'.$$

As $\beta_2 < \frac{1+g}{g+\ell}\varepsilon' < \beta_1 = \frac{1+g+\ell}{g+\ell}\varepsilon'$, we can divide both sides of (12) by $\beta_2 - \beta_1$ and obtain $-\frac{\beta_3-\beta_2}{\beta_2-\beta_1}$.

$$-\frac{\beta_3-\beta_2}{\beta_2-\beta_1} = \frac{\ell+\delta(1+g)(1-\beta_2)-g+\frac{\lambda}{\delta(1-\beta_1)(\beta_2-\beta_1)}}{\delta(1+g)(1-\beta_2)}.$$

As Assumption 2, $\beta_1, \beta_2 < 1$, and $\beta_2 - \beta_1 < 0$ hold, we find an upper bound of $-\frac{\beta_3-\beta_2}{\beta_2-\beta_1}$.

$$-\frac{\beta_3-\beta_2}{\beta_2-\beta_1} \leq \frac{\delta(1+g)(1-\beta_2)+\frac{\lambda}{\delta(1-\beta_1)(\beta_2-\beta_1)}}{\delta(1+g)(1-\beta_2)} < 1.$$

Taking $\beta_1 = \frac{1+g+\ell}{g+\ell}\varepsilon'$, $\beta_2 < \frac{1+g}{g+\ell}\varepsilon'$, and $-(\beta_2-\beta_1) > \frac{\ell}{g+\ell}\varepsilon'$ into account, we have a lower bound of $-\frac{\beta_3-\beta_2}{\beta_2-\beta_1}$ as follows.

$$\begin{aligned}-\frac{\beta_3-\beta_2}{\beta_2-\beta_1} >& \frac{\ell+g\left(1-\frac{1+g}{g+\ell}\varepsilon'\right)-g-\frac{\ell}{\left(\frac{g}{1+g}+\varepsilon'\right)\left(1-\frac{1+g+\ell}{g+\ell}\varepsilon'\right)}\frac{\lambda}{\varepsilon'}}{\left(\frac{g}{1+g}+\varepsilon'\right)(1+g)} \\ >& \frac{\ell-\frac{1+g}{g+\ell}g\varepsilon'-\frac{\ell}{\left(\frac{g}{1+g}+\varepsilon'\right)\left(1-\frac{1+g+\ell}{g+\ell}\varepsilon'\right)}\frac{\lambda}{\varepsilon'}}{g+(1+g)\varepsilon'} \\ >& \frac{\frac{3}{4}\ell}{\frac{3}{2}g} > \frac{\ell}{2g}.\end{aligned}$$

The first inequality follows from $\delta = \frac{g}{1+g} + \varepsilon' > \frac{g}{1+g}$. The third inequality is ensured by $\varepsilon' < 2\bar{\varepsilon}$ and $\lambda < \overline{\lambda}$. Therefore, we have obtained $\frac{\ell}{2g} < -\frac{\beta_3-\beta_2}{\beta_2-\beta_1} < 1$ and $\beta_3 \in (\beta_2, \beta_2)$. That is, $\beta_3 - \beta_2 > 0$.

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

By (4), (5), and (6), for any $t \in \mathbb{N}$, we have

$$\begin{cases} \frac{\lambda}{\delta(1-\beta_t)} = -(1-\beta_{t+1})g - \beta_{t+1}\ell + \delta(1-\beta_{t+1})(1-\beta_{t+2})(1+g) \\ \frac{\lambda}{\delta(1-\beta_{t+1})} = -(1-\beta_{t+2})g - \beta_{t+2}\ell + \delta(1-\beta_{t+2})(1-\beta_{t+3})(1+g), \end{cases}$$

or,

$$\begin{aligned} &-\frac{\beta_{t+1}-\beta_t}{\delta(1-\beta_t)(1-\beta_{t+1})}\lambda \\ =& -(\beta_{t+2}-\beta_{t+1})(g-\ell) + \delta(1-\beta_{t+2})\left\{(\beta_{t+3}-\beta_{t+2}) + (\beta_{t+2}-\beta_{t+1})\right\}(1+g). \end{aligned}$$

The suppositions ensure $\beta_{t+2}-\beta_{t+1} \neq 0$. Divide both sides of the above equation by $\beta_{t+2}-\beta_{t+1}$. Then, we obtain

$$-\frac{\beta_{t+3}-\beta_{t+2}}{\beta_{t+2}-\beta_{t+1}} = \frac{\ell + \delta(1+g)(1-\beta_{t+2}) - g + \frac{1}{\delta(1-\beta_t)(1-\beta_{t+1})\frac{\beta_{t+2}-\beta_{t+1}}{\beta_{t+1}-\beta_t}}\lambda}{\delta(1+g)(1-\beta_{t+2})}. \tag{13}$$

As Assumption 2 and $\frac{\beta_{t+2}-\beta_{t+1}}{\beta_{t+1}-\beta_t} < 0$ hold, $-\frac{\beta_{t+3}-\beta_{t+2}}{\beta_{t+2}-\beta_{t+1}}$ is bounded above by

$$-\frac{\beta_{t+3}-\beta_{t+2}}{\beta_{t+2}-\beta_{t+1}} \leq \frac{\delta(1+g)(1-\beta_{t+2}) + \frac{1}{\delta(1-\beta_t)(1-\beta_{t+1})\frac{\beta_{t+2}-\beta_{t+1}}{\beta_{t+1}-\beta_t}}\lambda}{\delta(1+g)(1-\beta_{t+2})} < 1.$$

Taking $0 < \beta_{t+1}, \beta_{t+2} < \frac{1+g+\ell}{g+\ell}\varepsilon' = \beta_1$, and $\frac{\ell}{2g} < -\frac{\beta_{t+2}-\beta_{t+1}}{\beta_{t+1}-\beta_t} < 1$ into account, we find the following lower bound of $-\frac{\beta_{t+3}-\beta_{t+2}}{\beta_{t+2}-\beta_{t+1}}$.

$$\begin{aligned} -\frac{\beta_{t+3}-\beta_{t+2}}{\beta_{t+2}-\beta_{t+1}} =& \frac{\ell + \delta(1-\beta_{t+2})(1+g) - g + \frac{1}{\delta(1-\beta_t)(1-\beta_{t+1})\frac{\beta_{t+2}-\beta_{t+1}}{\beta_{t+1}-\beta_t}}\lambda}{\delta(1+g)(1-\beta_{t+2})} \\ >& \frac{\ell + g\left(1 - \frac{1+g+\ell}{g+\ell}\varepsilon'\right) - g - \frac{1}{\left(\frac{g}{1+g}+\varepsilon'\right)\left(1-\frac{1+g+\ell}{g+\ell}\varepsilon'\right)^2\frac{2g}{\ell}}\lambda}{\left(\frac{g}{1+g}+\varepsilon'\right)(1+g)} \\ >& \frac{\ell - \frac{1+g+\ell}{g+\ell}g\varepsilon' - \frac{1}{\frac{g}{1+g}\cdot\frac{1}{4}\cdot 2}\varepsilon'}{g + (1+g)\varepsilon'} > \frac{\frac{3}{4}\ell}{\frac{3}{2}g} > \frac{\ell}{2g}. \end{aligned}$$

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

□

C Proof of Proposition 3

Proof of Proposition 3 . By (13), we have

$$-\frac{\beta_{t+3}-\beta_{t+2}}{\beta_{t+2}-\beta_{t+1}} = 1-\frac{g-\ell}{\delta(1-\beta_{t+2})(1+g)}-\frac{1}{\delta^2(1-\beta_t)(1-\beta_{t+1})(1-\beta_{t+1})\frac{\beta_{t+2}-\beta_{t+1}}{\beta_{t+1}-\beta_t}}\lambda.$$

Therefore, if $g-\ell<0$ and λ is small, then $-\frac{\beta_{t+3}-\beta_{t+2}}{\beta_{t+2}-\beta_{t+1}}>1$, and $|\beta_t|$ goes to infinity as t goes to infinity. That is, we have obtained a necessary condition for the efficiency result. □

D Proof of Proposition 4

Proof. Fix any $\varepsilon>0$. We define $\overline{\varepsilon}$, $\underline{\delta}$, $\overline{\delta}$, and $\overline{\lambda}$ as follows:

$$\begin{aligned}
\overline{\varepsilon} &\equiv \frac{\ell^2}{54(1+g+\ell)^2}\frac{\varepsilon}{1+\varepsilon},\\
\underline{\delta} &\equiv \frac{g}{1+g}+\overline{\varepsilon},\\
\overline{\delta} &\equiv \frac{g}{1+g}+2\overline{\varepsilon},\\
\overline{\lambda} &\equiv \frac{1}{16}\frac{g}{1+g}\frac{1}{1+g+\ell}\overline{\varepsilon}^2.
\end{aligned}$$

Fix any $\delta\in\left[\underline{\delta},\overline{\delta}\right]$ and $\lambda\in(0,\overline{\lambda})$. We show a sequential equilibrium whose payoff vector (v_1^*,v_2^*) satisfies $v_1^*=0$ and $v_2^*\geq\frac{1+g+\ell}{1+\ell}-\varepsilon$.

Strategy

We define a grim trigger strategy $\tilde{\sigma}$. Strategy $\tilde{\sigma}$ is represented by an automaton that has four types of states: initial $\tilde{\omega}_i^1$, cooperation $(\tilde{\omega}_i^t)_{t=2}^{\infty}$, transition ω_i^E, and defection ω_i^D. Players use the public randomization only at the initial state.

At the initial state $\tilde{\omega}_1^1$, player 1 chooses C_1 with probability $1-\beta_{1,1}$ and D_1 with probability $\beta_{1,1}$. Player 1 does not observe player 2 irrespective of his action. The transition state depends on a realized sunspot. If the realized sunspot is greater than $\hat{x}$, the state remains the same. If the realized sunspot is not greater than $\hat{x}$ and player 1 chooses C_1, then the state in the next period moves to the cooperation state $\tilde{\omega}_1^2$. If the realized sunspot is not greater than $\hat{x}$ and player 1 chooses D_1, then the state in the next period moves to the defection state ω_1^D.

At the initial state $\tilde{\omega}_2^1$, player 2 chooses D_2. Irrespective of his action, player 2 observes player 1 with probability $1-\beta_{2,2}$ and does not observe him with probability $\beta_{2,2}$. The transition state depends on the realized sunspot x. If the realized sunspot is greater than $\hat{x}$, the state remains the same. Suppose that the realized sunspot is not greater than $\hat{x}$. If player 2 observes C_1, then the state in the next period moves to the cooperation state $\tilde{\omega}_2^2$. If player 2 observes D_1, then the state in the next period is the defection state ω_2^D. If player 2 does not observe his opponent in period 1, then the state in the next period is the transition state ω_2^E.

At the cooperation state $\tilde{\omega}_1^2$, player 1 chooses action C_1 with probability $1-\beta_{1,2}$. When player 1 chooses action C_1, he observes his opponent with probability $1-\beta_{1,3}$. When player 1 chooses action D_1, he does not observe. If player 1 chooses C_1 and observes C_2, then the state in the next period is the cooperation state $\tilde{\omega}_1^3$. If player 1 chooses D_1 or observes D_2, then the state in the next period is the defection state ω_1^D. If player 1 chooses C_1 but does not observe, then the state in the next period is the transition state ω_1^E.

At the cooperation state $\tilde{\omega}_1^t (t \geq 3)$, player 1 chooses action C_1. Player 1 observes his opponent with probability $1-\beta_{1,t+1}$. If player 1 chooses C_1 and observes C_2, then the state in the next period is the cooperation state $\tilde{\omega}_1^{t+1}$. If player 1 chooses D_1 or observes D_2, then the state in the next period is the defection state ω_1^D. If player 1 chooses C_1 but does not observe, then the state in the next period is the transition state ω_1^E.

At the cooperation state $(\tilde{\omega}_2^t)_{t=2}^{\infty}$, player 2 chooses action C_2. He observes player 1 with probability $1-\beta_{2,t+1}$. If player 2 chooses C_2 and observes C_1, then the state in the next period is the cooperation state $\tilde{\omega}_2^{t+1}$. If player 2 chooses D_2 or observes D_1, then the state in the next period is the defection state ω_2^D. If player 2 chooses C_2 but does not observe, then the state in the next period is the transition state ω_2^E.

The output and transition functions at the transition state and the defection state are defined in the same manner as in the proof of Proposition 1. At the transition state ω_i^E in period t, each player i chooses D_i and does not observe irrespective of his action. The state moves to the defection state ω_i^D in period $t+1$ when $a_i^t = D_i$ or $o_i^t = D_j$ is realized. If player i chooses $a_i = C_i$ and $m_i = 0$, the state remains the same. When player i chooses C_i and observes C_j in period t, the state in period $t+1$ moves to the cooperation state $\tilde{\omega}_i^{t+1}$. At the defection state ω_i^D, the state remains the same, the defection state ω_i^D, irrespective of the action–observation pair.

The belief ψ_i^* for player i is determined in the same manner in the proof of Proposition 1. We consider a tremble that attaches far less weight to the deviations with respect to observations at any history h_t^i compared with those with respect to action for any i and any $t \in \mathbb{N}$. The above tremble induces the unique belief ψ_j^* for player j for each j. We denote by ψ^* the system of beliefs (ψ_1^*, ψ_2^*). The belief ψ^* has a similar property to the one in the proof of Proposition 1. That is, given ψ^*, player i is certain that the state of his opponent is the defection state when player i chose D_i or observed D_j in the past.

We define $(\beta_{1,t})_{t=1}^{\infty}$ and $(\beta_{2,t})_{t=2}^{\infty}$. First, let us define $\beta_{1,1}$ and $\beta_{1,2}$. We define $\varepsilon' \equiv \delta - \frac{g}{1+g}$. It is obvious that $\varepsilon' \in [\bar{\varepsilon}, 2\bar{\varepsilon}]$ We set $\beta_{1,1} = \frac{1+g+\ell}{g+\ell}\varepsilon'$. We define

$\beta_{1,2}$ as follows.

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

Let $W_{i,t}(t \geq 2)$ be the sum of the stage game payoffs for player i from the cooperation state ω_i^t. At any cooperation state $\omega_2^{t+1}(t \in \mathbb{N})$, player 2 believes that the state of his opponent is the cooperation state ω_1^{t+1} with probability $1-\beta_{1,t+1}$, and with the remaining probability $\beta_{1,t+1}$, the state is either ω_1^E or ω_1^D. Therefore, $W_{2,t+1}$ is given by

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

At the initial state, player 2 is indifferent between $m_2 = 1$ and $m_2 = 0$. Therefore, $\beta_{1,3}$ is given by

$$\frac{\lambda}{\hat{x}\delta(1-\beta_{1,1})} = W_{2,2}(\beta_{1,2},\beta_{1,3}) - (1-\beta_{1,2})(1+g).$$

At any cooperation state, player 2 is indifferent between $m_2 = 1$ and $m_2 = 0$. Therefore, for any $t \in \mathbb{N}$, $\beta_{1,t+2}$ is given by

$$\frac{\lambda}{\delta(1-\beta_{1,t})} = W_{2,t+1}(\beta_{1,t+1},\beta_{1,t+2}) - (1-\beta_{1,t+1})(1+g). \tag{14}$$

Next, we define $(\beta_{2,t})_{t=2}^{\infty}$. We define $\beta_{2,2}$ so that player 1 is indifferent between choosing $(C_1, 0)$ and $(D_1, 0)$ at the initial state. That is, $\beta_{2,2}$ is given by the equation below.

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

Player 1 randomizes $(C_1, 0)$ and $(D_1, 0)$ at the cooperation state $\tilde{\omega}_1^2$. Hence, $\beta_{2,3}$ is given by the following equation.

$$(1-\beta_{2,2}) - \beta_{2,2}\ell + \delta(1-\beta_{2,2})(1-\beta_{2,3})(1+g) = (1-\beta_{2,2})(1+g).$$

In the cooperation state $\tilde{\omega}_1^t$ $(t \geq 2)$, player 1 believes that the state of his opponent is the cooperation state with probability $1-\beta_{2,t}$. Therefore, $W_{1,t}(t \geq 2)$ is given by

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

Furthermore, player 1 randomizes $(C_1, 1)$ and $(C_1, 0)$ at the cooperation state $\tilde{\omega}_1^2$. At the cooperation state $\tilde{\omega}_1^2$, player 1 believes that the state of player 2 is $\tilde{\omega}_2^2$ with probability $1-\beta_{2,2}$. Therefore, $\beta_{2,4}$ is determined as the solution of the following equation.

$$\frac{\lambda}{\delta(1-\beta_{2,2})} = W_{1,3}(\beta_{2,3},\beta_{2,4}) - (1-\beta_{2,3})(1+g).$$

In addition, player 1 randomizes $(C_1, 1)$ and $(C_1, 0)$ at the cooperation state $\tilde{\omega}_1^t$ $(t \geq 3)$. At the cooperation state $\tilde{\omega}_1^t$ $(t \geq 3)$, player 1 believes that the state of player 2 is $\tilde{\omega}_2^t$ with probability $1 - \beta_{2,t}$. We choose $\beta_{2,t+1}$ as the solution of the equation below so that player 1 is indifferent between choosing $(C_1, 1)$ and $(C_1, 0)$.

$$\frac{\lambda}{\delta(1-\beta_{2,t})} = W_{1,t+1}(\beta_{2,t+1}, \beta_{2,t+2}) - (1-\beta_{2,t+1})(1+g). \tag{15}$$

Taking into account the definition of δ, $(\beta_{1,t})_{t=2}^{\infty}$ and $(\beta_{2,t})_{t=2}^{\infty}$ are chosen as follows.

$$\begin{aligned}
\beta_{1,1} =& \frac{1+g+\ell}{g+\ell}\varepsilon' \\
\beta_{1,2} =& \frac{(1-\beta_{1,1})\{\delta(1+g)-g\} - \beta_{1,1}\ell}{\delta(1-\beta_{1,1})(1+g)} \\
\beta_{1,3} =& \frac{(1-\beta_{1,2})\{\delta(1+g)-g\} - \beta_{1,2}\ell - \frac{\lambda}{\hat{x}\delta(1-\beta_{1,1})}}{\delta(1+g)(1-\beta_{1,2})}, \\
\beta_{1,t+2} =& \frac{(1-\beta_{1,t+1})\{\delta(1+g)-g\} - \beta_{1,t+1}\ell - \frac{\lambda}{\delta(1-\beta_{1,t})}}{\delta(1+g)(1-\beta_{1,t+1})}, \quad \forall\, t \geq 2. \\
\beta_{2,2} =& \frac{\hat{x}\delta(1+g)-\ell}{\hat{x}\delta(1+g)} \\
\beta_{2,3} =& \frac{(1-\beta_{2,2})\{\delta(1+g)-g\} - \beta_{2,2}\ell}{\delta(1+g)(1-\beta_{2,2})} \\
\beta_{2,t+2} =& \frac{(1-\beta_{2,t+1})\{\delta(1+g)-g\} - \beta_{2,t+1}\ell - \frac{\lambda}{\delta(1-\beta_{2,t})}}{\delta(1+g)(1-\beta_{2,t+1})}, \quad \forall\, t \geq 2.
\end{aligned}$$

Finally, we choose $\hat{x}$. We define $\hat{x}$ as the solution below.

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

When $\hat{x} = \frac{\ell}{g}$, the left-hand side is greater than the right-hand side.

$$\frac{\frac{\ell}{g}(1+g)\varepsilon'}{\frac{\ell}{g}\delta(1+g)} = \frac{1}{\delta}\varepsilon' > \frac{1+g+\ell}{g+\ell}\varepsilon'.$$

Furthermore, if $\hat{x} = \frac{\ell}{\delta(1+g)}$, then the left-hand side is smaller than the right-hand side. Therefore, $\hat{x} \in \left(\frac{\ell}{\delta(1+g)}, \frac{\ell}{g}\right)$ is well defined.

By similar discussion in Lemma 2, we have

$$\begin{aligned}
&\frac{1}{2}\frac{1+g-\ell}{g+\ell}\varepsilon' < \beta_{1,t} < \frac{1+g+\ell}{g+\ell}\varepsilon', && \forall t \in \mathbb{N}, \text{ and} \\
&\frac{1}{2}\frac{1+g-\ell}{g+\ell}\varepsilon' < \beta_{2,t+1} < \frac{1+g+\ell}{g+\ell}\varepsilon', && \forall t \in \mathbb{N}.
\end{aligned}$$

Following the proof of Proposition 1, we show sequential rationality and that the equilibrium payoff vector (v_1^*, v_2^*) satisfies $v_1^* = 0$ and $v_2^* \geq \frac{1+g+\ell}{1+\ell} - \varepsilon$.

Sequential rationality at the defection state

Let us confine our attention to show sequential rationality. At the defection state, player i is certain that the state of his opponent is the defection state, and the opponent chooses $(D_j, 0)$ with certainty from the current period onwards. Player i has no incentive to choose C_i or $m_i = 1$. Therefore, it is optimal for player i to choose $(D_i, 0)$.

Sequential rationality at the initial state and the cooperation state

Let us consider a cooperation state $\tilde{\omega}_i^t (t \geq 2)$. Once player i chooses D_i, the strategy σ^* prescribes D_i every period, irrespective of his observation result. Therefore, at any cooperation state, each player i has no incentive to choose $(D_i, 1)$.

First, let us consider player 1's sequential rationality at the initial state $\tilde{\omega}_1^1$. The definition of $\beta_{2,2}$ ensures that player 1 is indifferent between $(C_1, 0)$ and $(D_1, 0)$. It is obvious that player 1 has no incentive to observe player 2 because player 2 chooses action D_2 with certainty.

Next, let us consider the decision of player 1 at cooperation states. At the cooperation state $\tilde{\omega}_1^2$, the stage-behaviors $(C_1, 1)$, $(C_1, 0)$ and $(D_1, 0)$ are indifferent by the definitions of $\beta_{2,3}$ and $\beta_{2,4}$. At the cooperation state $\tilde{\omega}_1^{t+2}$ $(t \geq 1)$, the definition of $\beta_{2,t+4}$ ensures that $(C_1, 1)$ and $(C_1, 0)$ are indifferent. In addition, the equation (15) in period $t + 1$ implies that the payoff $W_{1,t+2}$ for choosing action C_1 is greater than the payoff $(1-\beta_{2,t+2})(1+g)$ when he chooses action D_2. It is optimal for player 1 to follow the strategy $\tilde{\sigma}$ at cooperation states $(\tilde{\omega}_{1,t})_{t=2}^{\infty}$.

Lastly, let us consider player 2's choice at the initial state $\tilde{\omega}_2^1$. By the definition of $\beta_{1,3}$, player 2 is indifferent between choosing $(C_2, 1)$ and $(C_2, 0)$. Player 2 does not prefer action D_2 because player 1 never observes him. Next, let us confine our attention to player 2's choice at the cooperation state $\tilde{\omega}_2^t$ $(t \geq 2)$. By the definition of $\beta_{1,t+2}$, player 2 is indifferent between choosing $(C_2, 1)$ and g $(C_2, 0)$. If player 2 chooses $(D_2, 0)$, his payoff is $(1 - \beta_{1,t})(1 + g)$. The inequality (14) in period $t - 1$ ensures that the payoff $W_{2,t}$ for choosing C_1 is greater than $(1 - \beta_{1,t})(1 + g)$. That is, action D_2 is suboptimal.

Thus, it is optimal for both players to follow strategy $\tilde{\sigma}$ in the cooperation state.

Sequential rationality in the transition state

We consider sequential rationality at any period t (≥ 2) associated with the transition state.

First, let us consider the transition state for player 1 in period t ($t \geq 3$). Let p be the probability with which player 1 believes that the state of his opponent is the cooperation state. Therefore, the upper bound of the payoff when player 1 chooses action C_1 in period t is given by

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

Furthermore, the payoff for $(D_1, 0)$ is bounded above by $p(1+g)$. Therefore, $(D_1, 0)$ is profitable if the following value is negative.

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

Using (15), we can rewrite the above value as follows.

$$\begin{aligned}
&p - (1-p)\ell + \delta p W_{1,t+1} - p(1+g) \\
=&(1-\beta_{2,t}) - \beta_{2,t}\ell - \lambda + \delta(1-\beta_{2,t})W_{1,t+1} - (1-\beta_{2,t})(1+g) \\
&\quad + \lambda + \{p - (1-\beta_{2,t})\}\{1 + \ell + \delta W_{1,t+1} - (1+g)\} \\
=&\frac{\lambda}{\delta(1-\beta_{2,t-1})} + \lambda - \{(1-\beta_{2,t}) - p\}\{\delta W_{1,t+1} - (g-\ell)\}. \qquad (16)
\end{aligned}$$

The second equality follows from equation (15) for $t-1$.

Furthermore, the payoff $\delta W_{1,t+1}$ is greater than that for choosing $(D_1, 0)$. Therefore, the payoff $\delta W_{1,t+1}$ is bounded below by

$$\begin{aligned}
\delta W_{1,t+1} - (g-\ell) \geq& \delta(1-\beta_{2,t+1})(1+g) - (g-\ell) \\
\geq& \{g + (1+g)\varepsilon'\}\left(1 - \frac{1+g+\ell}{g+\ell}\varepsilon'\right) - (g-\ell) \\
\geq& \frac{\ell}{2}. \qquad (17)
\end{aligned}$$

The second inequality follows from $\delta = \frac{g}{1+g} + \varepsilon'$ and $\beta_{2,t+1} \leq \frac{1+g+\ell}{g+\ell}\varepsilon'$.

The maximum value of p in period t ($t \geq 3$) is $(1-\beta_{2,t-1})(1-\beta_{2,t})$. Taking (17) into account, the value of (16) has the following upper bound.

$$\begin{aligned}
&\frac{\lambda}{\delta(1-\beta_{2,t-1})} + \lambda - \{(1-\beta_{2,t}) - p\}\delta W_{1,t+1} \\
<&\frac{1+g}{g}\frac{\lambda}{1-\beta_{2,t-1}} + \lambda - (1-\beta_{2,t-1})\beta_{2,t}\frac{\ell}{2} \\
<&\frac{1+g}{g}\frac{\lambda}{1-\frac{1+g+\ell}{g+\ell}\varepsilon'} + \lambda - \left(1 - \frac{1+g+\ell}{g+\ell}\varepsilon'\right)\frac{1}{2}\frac{1+g-\ell}{1+g+\ell}\varepsilon'\frac{\ell}{2} \\
<&0.
\end{aligned}$$

The second inequality follows from $\frac{1}{2}\frac{1+g-\ell}{1+g+\ell}\varepsilon' < \beta_{2,t-1}, \beta_{2,t} < \frac{1+g+\ell}{g+\ell}\varepsilon'$. Therefore, choosing $(D_1, 0)$ is optimal at the transition state ω_1^E.

Next, let us consider the transition state for player 2 in period 2. Then, player 2 believes that the state of his opponent is the cooperation state $\tilde{\omega}_1^2$ with probability $1-\beta_{1,1}$. If player 2 chooses C_2, the continuation payoff is bounded above by

$$(1-\beta_{1,1})W_{2,2}-\beta_{1,1}\ell.$$

However, the payoff of choosing $(D_2,0)$ is given by $(1-\beta_{1,1})(1-\beta_{1,2})(1+g)$. Therefore, it is optimal for player 2 to choose $(D_2,0)$ if the following value is negative.

$$(1-\beta_{1,1})W_{2,2}-\beta_{1,1}\ell-(1-\beta_{1,1})(1-\beta_{1,2})(1+g).$$

Or, equivalently

$$\begin{aligned}
&(1-\beta_{1,1})\left\{W_{2,2}-(1-\beta_{1,2})(1+g)\right\}-\beta_{1,1}\ell\\
=&(1-\beta_{1,1})\frac{\lambda}{\delta(1-\beta_{1,1})}-\beta_{1,1}\ell\\
=&\frac{\lambda}{\delta}-\beta_{1,1}\ell<0.
\end{aligned}$$

Therefore, it is optimal for player 2 to choose $(D_2,0)$.

Finally, let us consider the transition state for player 2 in period t ($t\geq 3$). Let us denote by p the probability with which player 2 believes that the state of his opponent is a cooperation state. Then, the upper bound of the payoff when player 2 chooses action C_2 in period t is given by

$$p-(1-p)\ell+\delta pW_{2,t+1}.$$

The payoff for $(D_2,0)$ is given by $p(1+g)$. Therefore, $(D_2,0)$ is profitable if the following value is negative.

$$p-(1-p)\ell+\delta pW_{2,t+1}-p(1+g).$$

We can rewrite the above value as follows.

$$\begin{aligned}
&p-(1-p)\ell+\delta pW_{2,t+1}-p(1+g)\\
=&(1-\beta_{1,t})-\beta_{1,t}\ell-\lambda+\delta(1-\beta_{1,t})W_{2,t+1}-(1-\beta_{1,t})(1+g)\\
&\quad+\lambda+\{p-(1-\beta_{1,t})\}\{1+\ell+\delta W_{2,t+1}-(1+g)\}\\
=&W_{2,t}-(1-\beta_{1,t})(1+g)+\lambda+\{p-(1-\beta_{1,t})\}\{\delta W_{2,t+1}-(g-\ell)\}\\
=&\frac{\lambda}{\delta(1-\beta_{1,t-1})}+\lambda-\{(1-\beta_{1,t})-p\}\{\delta W_{2,t+1}-(g-\ell)\}.
\end{aligned}\tag{18}$$

The third equality follows from equation (14) for $t-1$.

Furthermore, $\delta W_{2,t+1}$ is bounded below by

$$\begin{aligned}
\delta W_{2,t+1}-(g-\ell)\geq&\delta(1-\beta_{1,t+1})(1+g)-(g-\ell)\\
\geq&\{g+(1+g)\varepsilon'\}\left(1-\frac{1+g+\ell}{g+\ell}\varepsilon'\right)-(g-\ell)\\
\geq&\frac{\ell}{2}.
\end{aligned}$$

The second inequality follows from $\delta = \frac{g}{1+g} + \varepsilon'$ and $\beta_{1,t+1} \leq \frac{1+g+\ell}{g+\ell}\varepsilon'$.

The maximum value of p in period t is $(1-\beta_{1,t-1})(1-\beta_{1,t})$. Taking (17) into account, we can show that (18) is negative as follows.

$$\begin{aligned}
&\frac{\lambda}{\delta(1-\beta_{1,t-1})} + \lambda - \{(1-\beta_{1,t}) - p\}\,\delta W_{2,t+1} \\
\leq &\frac{\lambda}{\delta(1-\beta_{1,t-1})} + \lambda - (1-\beta_{1,t})\beta_{1,t-1}\frac{\ell}{2} \\
\leq &\frac{1+g}{g}\frac{1}{1-\frac{1+2g}{2g}\varepsilon'}\lambda + \lambda - \left(1 - \frac{1+g+\ell}{g+\ell}\varepsilon'\right)\frac{1}{2}\frac{1+g-\ell}{1+g+\ell}\varepsilon'\frac{\ell}{2} \\
<&0.
\end{aligned}$$

The second inequality is ensured by $\beta_{1,t}, \beta_{1,t-1} \in \left(\frac{1}{2}\frac{1+g-\ell}{1+g+\ell}\varepsilon', \frac{1+g+\ell}{g+\ell}\varepsilon'\right)$. Therefore, player 2 prefers D_2 to C_2 at the transition state.

Hence, it has been proven that it is optimal for both players to follow strategy $\tilde{\sigma}$. The strategy $\tilde{\sigma}$ is a sequential equilibrium.

The payoff

Finally, let us consider the equilibrium payoff. The equilibrium payoff for player 1 is 0 because player 1 weakly prefers $(D_1, 0)$ in period 1.

Similarly, player 2 weakly prefers $(D_2, 0)$ in period 2. Thus, his equilibrium payoff v_2^* is given by

$$\begin{aligned}
v_2^* \geq &(1-\delta)(1-\beta_{1,1})\{(1+g) + \hat{x}\delta(1-\beta_{1,2})(1+g)\} + (1-\hat{x})\delta v_2^* \\
= &\frac{(1-\delta)(1-\beta_{1,1})\{(1+g) + \hat{x}\delta(1-\beta_{1,2})(1+g)\}}{1-(1-\hat{x})\delta} \\
= &\frac{(1-\beta_{1,1})\{1 + \hat{x}\delta(1-\beta_{1,2})\}}{1+\hat{x}\frac{\delta}{1-\delta}}(1+g).
\end{aligned}$$

Taking $\hat{x} \in \left(\frac{\ell}{\delta(1+g)}, \frac{\ell}{g}\right)$ into consideration, we obtain a lower bound of v_2^* below.

$$\begin{aligned}
v_2^* > &\frac{(1-\beta_{1,1})\left\{1 + \frac{\ell}{1+g}(1-\beta_{1,2})\right\}}{1+\frac{\ell}{g}\frac{\delta}{1-\delta}}(1+g) \\
> &\frac{\left(1 - \frac{1+g+\ell}{g+\ell}\varepsilon'\right)\left(1+g+\ell - \frac{1+g+\ell}{g+\ell}\varepsilon'\ell\right)}{1+\frac{\ell}{g}\frac{g+\varepsilon'}{1-(1+g)\varepsilon'}} > \frac{1+g+\ell}{1+\ell} - \varepsilon.
\end{aligned}$$

The second inequality follows from the upper bound of $\beta_{1,1}$ and $\beta_{1,2}$. Therefore, Proposition 4 has been proven. □

E Proof of Theorem 1

Proof. Let us fix $\overline{n}$ such that:

$$\overline{n} \geq \frac{8+2g}{\varepsilon}.$$

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

By Corollary 4.1, there exists a sequential equilibrium strategy $\tilde{\sigma}$ whose payoff vector $v^* = (v_1^*, v_2^*)$ is sufficiently close to payoff vector $\hat{v}^* = \left(0, \frac{1+g+\ell}{1+\ell}\right)$ and satisfies $|v_i^* - \hat{v}_i| < \frac{1}{\overline{n}}$ when discount factor $\delta^{\overline{n}}$ is sufficiently large. We can also find a sequential equilibrium strategy σ^* whose payoff vector $v^{**} = (v_1^{**}, v_2^{**})$ satisfies $|v_i^{**} - 1| < \frac{1}{\overline{n}}$ when discount factor $\delta^{\overline{n}}$ is sufficiently large by Proposition 2.

Let us assume that $v_1 \leq v_2$. We choose sufficiently large discount factor δ so that we can use Corollary 4.1 and Proposition 4, and the discount factor δ satisfies the following condition:

$$\frac{1-\delta}{1-\delta^{\overline{n}}} \leq \frac{2}{\overline{n}}.$$

The desired payoff vector v can be expressed uniquely as a convex combination of v^*, v^{**} and $(0,0)$ as below.

$$v = \alpha_1 \delta v^{**} + \alpha_2 \delta v^* + (1-\alpha_1-\alpha_2)\cdot 0.$$

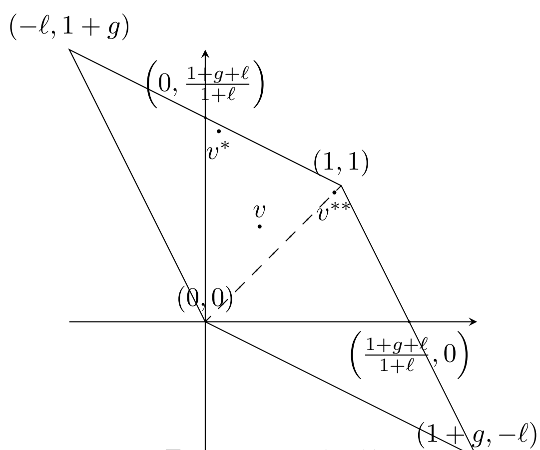

Figure 2: v, v^*, v^{**}

Let us define n_1 and n_2 as follows.

$$n_1 \equiv \underset{n\in\mathbb{N}\cup\{0\}}{\arg\min} \left| \frac{1-\delta^n}{1-\delta^{\overline{n}}} - \alpha_1 \right|, \qquad n_2 \equiv \underset{n\in\mathbb{N}\cup\{0\}}{\arg\min} \left| \frac{\delta^{n_1}-\delta^{n_1+n}}{1-\delta^{\overline{n}}} - \alpha_2 \right|.$$

Then, n_1 and n_2 satisfy

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

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

The payoff v_i^F for player i is given by

$$v_i^F = \frac{(1-\delta^{n_1})v_i^* + (\delta^{n_1}-\delta^{n_1+n_2})v_i^{**} + (\delta^{n_1+n_2}-\delta^{\overline{n}})\cdot 0}{1-\delta^{\overline{n}}}$$

Then, we have

$$\begin{aligned}\left| \frac{1-\delta^{n_1}}{1-\delta^{\overline{n}}} v_i^* - \alpha_1 \hat{v}_i^* \right| &= \left| \alpha_1(v_i^*-\hat{v}_i^*) + \left(\frac{1-\delta^{n_1}}{1-\delta^{\overline{n}}} - \alpha_1 \right) \hat{v}_i^* + \left(\frac{1-\delta^{n_1}}{1-\delta^{\overline{n}}} - \alpha_1 \right)(v_i^*-\hat{v}_i^*) \right| \\ &\leq \varepsilon' + \frac{2}{\overline{n}}(1+g) + \frac{2}{\overline{n}}\varepsilon' = \frac{2(2+g)}{\overline{n}} \\ &\leq \frac{1}{\overline{n}} + \frac{2}{\overline{n}}(1+g) + \frac{1}{\overline{n}} = \frac{2(2+g)}{\overline{n}}.\end{aligned}$$

Similarly, we have

$$\begin{aligned}&\left| \frac{\delta^{n_1}-\delta^{n_1+n_2}}{1-\delta^{\overline{n}}} v_i^{**} - \alpha_2 \cdot 1 \right| \\ =&\left| \alpha_2(v_i^{**}-1) + \left(\frac{\delta^{n_1}-\delta^{n_1+n_2}}{1-\delta^{\overline{n}}} - \alpha_2 \right) + \left(\frac{\delta^{n_1}-\delta^{n_1+n_2}}{1-\delta^{\overline{n}}} - \alpha_2 \right)(v_i^{**}-1) \right| \leq \frac{4}{\overline{n}}.\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}&\left| \frac{(1-\delta^{n_1})v_i^* + (\delta^{n_1}-\delta^{n_1+n_2})v_i^{**} + (\delta^{n_1+n_2}-\delta^{\overline{n}})\cdot 0}{1-\delta^{\overline{n}}} - v_i \right| \\ \leq& \left| \frac{1-\delta^{n_1}}{1-\delta^{\overline{n}}} v_i^* - \alpha_1 \hat{v}_i^* \right| + \left| \frac{\delta^{n_1}-\delta^{n_1+n_2}}{1-\delta^{\overline{n}}} v_i^{**} - \alpha_2 \cdot 1 \right| \\ \leq& \frac{8+2g}{\overline{n}} \leq \varepsilon.\end{aligned}$$

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 game, it is straightforward that the payoff vector v can be approximated by a sequential equilibrium payoff vector when $v_1 \geq v_2$ holds as well. □

References

[1] Ben-Porath, Elchanan and Michael Kahneman (2003) "Communication in repeated games with costly monitoring," *Games and Economic Behavior*, Vol. 44, No. 2, pp. 227–250, DOI: https://doi.org/10.1016/S0899-8256(03)00022-8.

[2] Bhaskar, V. and Eric van Damme (2002) "Moral Hazard and Private Monitoring," *Journal of Economic Theory*, Vol. 102, No. 1, pp. 16–39, DOI: https://doi.org/10.1006/JETH.2001.2861.

[3] Bhaskar, V. and Ichiro Obara (2002) "Belief-Based Equilibria in the Repeated Prisoners' Dilemma with Private Monitoring," *Journal of Economic Theory*, Vol. 102, No. 1, pp. 40–69, DOI: https://doi.org/10.1006/jeth.2001.2878.

[4] Chen, Bo (2010) "A belief-based approach to the repeated prisoners' dilemma with asymmetric private monitoring," *Journal of Economic Theory*, Vol. 145, No. 1, pp. 402–420, DOI: https://doi.org/10.1016/J.JET.2009.05.006.

[5] Ellison, G. (1994) "Cooperation in the Prisoner's Dilemma with Anonymous Random Matching," *The Review of Economic Studies*, Vol. 61, No. 3, pp. 567–588, DOI: https://doi.org/10.2307/2297904.

[6] Ely, Jeffrey C. and Juuso Välimäki (2002) "A Robust Folk Theorem for the Prisoner's Dilemma," *Journal of Economic Theory*, Vol. 102, No. 1, pp. 84–105, DOI: https://doi.org/10.1006/JETH.2000.2774.

[7] Ely, Jeffrey C., Johannes Horner, and Wojciech Olszewski (2005) "Belief-Free Equilibria in Repeated Games," *Econometrica*, Vol. 73, No. 2, pp. 377–415, DOI: https://doi.org/10.1111/j.1468-0262.2005.00583.x.

[8] Flesch, János and Andrés Perea (2009) "Repeated Games with Voluntary Information Purchase," *Games and Economic Behavior*, Vol. 66, No. 1, pp. 126–145, DOI: https://doi.org/10.1016/j.geb.2008.04.015.

[9] Hino, Yoshifumi (2019) "An efficiency result in a repeated prisoner's dilemma game under costly observation with nonpublic randomization," *Mathematical Social Sciences*, Vol. 101, pp. 47–53, DOI: https://doi.org/10.1016/J.MATHSOCSCI.2019.06.005.

[10] Hörner, Johannes and Wojciech Olszewski (2006) "The Folk Theorem for Games with Private Almost-Perfect Monitoring," *Econometrica*, Vol. 74, No. 6, pp. 1499–1544, DOI: https://doi.org/10.1111/j.1468-0262.2006.00717.x.

[11] ——— (2009) "How Robust Is the Folk Theorem?," *Quarterly Journal of Economics*, Vol. 124, No. 4, pp. 1773–1814, DOI: `https://doi.org/10.1162/qjec.2009.124.4.1773`.

[12] Kandori, Michihiro and Ichiro Obara (2004) "Endogenous Monitoring," *UCLA Economics Online Papers*, Vol. 398, URL: `http://www.econ.ucla.edu/people/papers/Obara/Obara398.pdf`.

[13] Lehrer, Ehud and Eilon Solan (2018) "High Frequency Repeated Games with Costly," *Theoretical Economics*, Vol. 13, No. 1, pp. 87–113, DOI: `https://doi.org/10.3982/TE2627`.

[14] Mailath, George J. and Stephen Morris (2002) "Repeated Games with Almost-Public Monitoring," *Journal of Economic Theory*, Vol. 102, No. 1, pp. 189–228, DOI: `https://doi.org/10.1006/jeth.2001.2869`.

[15] Mailath, George J and Stephen Morris (2006) "Coordination failure in repeated games with almost-public monitoring," *Theoretical Economics*, Vol. 1, pp. 311–340, URL: `https://econtheory.org/ojs/index.php/te/article/viewFile/20060311/791/30`.

[16] Mailath, George J. and Wojciech Olszewski (2011) "Folk theorems with bounded recall under (almost) perfect monitoring," *Games and Economic Behavior*, Vol. 71, No. 1, pp. 174–192, DOI: `https://doi.org/10.1016/j.geb.2010.11.002`.

[17] Miyagawa, Eiichi, Yasuyuki Miyahara, and Tadashi Sekiguchi (2003) "Repeated Games with Observation Costs," *Columbia University Academic Commons*, No. 203-14, DOI: `https://doi.org/https://doi.org/10.7916/D8VX0TRW`.

[18] ——— (2008) "The folk theorem for repeated games with observation costs," *Journal of Economic Theory*, Vol. 139, No. 1, pp. 192–221, DOI: `https://doi.org/10.1016/j.jet.2007.04.001`.

[19] Piccione, Michele (2002) "The Repeated Prisoner's Dilemma with Imperfect Private Monitoring," *Journal of Economic Theory*, Vol. 102, No. 1, pp. 70–83, DOI: `https://doi.org/10.1006/JETH.2001.2819`.

[20] Sekiguchi, Tadashi (1997) "Efficiency in Repeated Prisoner's Dilemma with Private Monitoring," *Journal of Economic Theory*, Vol. 76, No. 2, pp. 345–361, DOI: `https://doi.org/10.1006/jeth.1997.2313`.

[21] Yamamoto, Yuichi (2007) "Efficiency results in N player games with imperfect private monitoring," *Journal of Economic Theory*, Vol. 135, No. 1, pp. 382–413, DOI: `https://doi.org/10.1016/J.JET.2006.05.003`.

[22] ——— (2009) "A limit characterization of belief-free equilibrium payoffs in repeated games," *Journal of Economic Theory*, Vol. 144, No. 2, pp. 802–824, DOI: `https://doi.org/10.1016/J.JET.2008.07.005`.