The Nash-Threat Folk Theorem in Repeated Games with Private Monitoring and Public Communication

Takuo Sugaya[*][†]
Stanford Graduate School of Business

November 7, 2012

Abstract

Assuming that cheap talk is available, we show that the Nash-threat folk theorem holds for repeated games with private monitoring if the individual full rank condition is satisfied.

Journal of Economic Literature Classification Numbers: C72, C73, D82

Keywords: repeated game, folk theorem, private monitoring

[*]tsugaya@gsb.stanford.edu
[†]This paper stems from Sugaya (2012b), Sugaya (2012c) and Sugaya (2012d).

1 Introduction

One of the key results in the literature on infinitely repeated games is the folk theorem: any feasible and individually rational payoff can be sustained in equilibrium when players are sufficiently patient. Even if a stage game does not have an efficient Nash equilibrium, the repeated game does. Hence, the repeated game gives a formal framework to analyze a cooperative behavior. Fudenberg and Maskin (1986) establish the folk theorem under perfect monitoring, where players can directly observe the action profile. Fudenberg, Levine, and Maskin (1994) extend the folk theorem to imperfect public monitoring, where players can observe only public noisy signals about the action profile.

The driving force of the folk theorem is reciprocity: if a player deviates today, she will be punished in future. For this mechanism to work, each player needs to coordinate her action with the other players' histories.

This coordination is straightforward if players' strategies only depend on the public component of histories, such as action profiles in perfect monitoring or public signals in public monitoring. Since this public information is common knowledge, players can coordinate a punishment contingent on the public information (reciprocity), and thereby provide dynamic incentives to choose actions that are not static best responses. On the other hand, with private monitoring, where players can observe only private noisy signals about action profiles, since they do not share common information about histories, this coordination could become complicated as periods proceed.

Hörner and Olszewski (2006) and Hörner and Olszewski (2009) show the robustness of this coordination to private monitoring if monitoring is almost perfect and almost public, respectively. If monitoring is almost perfect, then players can believe that every player observes the same signal corresponding to the true action profile with a high probability. If monitoring is almost public, then players can believe that every player observes the same signal with a high probability. Hence, almost common knowledge about relevant histories still exists.

However, with general private monitoring, almost common knowledge may not exist and coordination is difficult. Nevertheless, a series of papers, Sugaya (2012b), Sugaya (2012c) and Sugaya (2012d), show that the folk theorem with the lower bound calculated by individually-mixed minimax values generically holds in repeated games with private monitoring, *without* public randomization or cheap talk.

The proof of these papers goes as follows: first, assuming that cheap talk is available, construct a sequential equilibrium to support an arbitrarily fixed payoff profile. Second, dispense with cheap talk by showing that players can communicate by actions. In this paper, we set aside the second issue and focus on the first component.

Since the messages by cheap talk are public, introducing cheap talk and letting a strategy depend on the messages by the cheap talk helps to overcome the difficulty of coordination through private signals. In fact, folk theorems have been proven by Compte (1998), Kandori and Matsushima (1998), Aoyagi (2002), Fudenberg and Levine (2007) and Obara (2009).

There are two key differences between this paper and the other papers: first, in this paper, the communication is carefully constructed so that we can dispense with cheap talk later.[1] Note that, when the players communicate with actions, the common knowledge about the messages will disappear.

Second, even with cheap talk, it is hard to incentivize the players to tell the truth when the monitoring of actions is private since there is no precise evidence to show that a player tells a lie. This is why the existing papers in the literature need to assume more than individual identifiability of actions. In this paper, we show that individual identifiability is sufficient if we carefully construct an equilibrium.

To this end, it simplifies the equilibrium construction to concentrate on the Nash-threat folk theorem rather than the minimax-threat folk theorem. The reason is related to one well known in mechanism design: if there are only two players and they send messages that are statistically rare, then the players cannot tell which one of them is more suspicious. Even in such a case, they can mutually punish each other by going to a long repetition of a static Nash equilibrium to discourage lies. See Sugaya (2012c) and Sugaya (2012d) for how to deal

[1] See Sugaya (2012b), Sugaya (2012c) and Sugaya (2012d) for this part.

with mixed-strategy minimax values.

To show the folk theorem with general monitoring, we unify and improve on belief-free equilibria that has been used extensively to show the partial results so far in the literature on private monitoring. A strategy profile is belief-free if, after any history profile, the continuation strategy of each player is optimal conditional on the histories of the opponents. Hence, coordination never becomes an issue. The belief-free approach has been successful in showing the folk theorem in prisoners' dilemma with almost perfect monitoring. See, among others,[2] Piccione (2002), Ely and Välimäki (2002), Ely, Hörner, and Olszewski (2005), Yamamoto (2007) and Yamamoto (2009).

There are two extensions necessary for the folk theorem in general games with general monitoring. First, without any assumption on the precision of monitoring, Matsushima (2004) and Yamamoto (2012) show the folk theorem in prisoners' dilemma if the monitoring is conditionally independent. The main idea is to recover the precision of monitoring by "review strategies." The intuition is as follows: suppose two players play prisoners' dilemma. Each player i has two signals, g_i (good signal) and b_i (bad signal). If player j (the opponent) takes C_j (cooperation), then player i observes the good signal more likely. For example, the probability of g_i given C_j is 0.6 while that given D_j is 0.3. (If these numbers were almost equal to 1 and 0, respectively, then the monitoring would be almost perfect.) For a simple exposition, let us see one period as a day. Even if a signal per day is not so precise, if player j has an incentive to take a constant action over a year, then player i can get an almost precise idea of player j's action by aggregating information over the year. With conditionally independent monitoring, since player j cannot obtain any information about how player i's review on player j is going over the year, it is optimal for player j to adhere to one constant action.

However, without conditionally independent monitoring, player j has an incentive to

[2]Kandori and Obara (2006) use a similar concept to analyze a private strategy in public monitoring. Kandori (2011) considers "weakly belief-free equilibria," which is a generalization of belief-free equilibria. Apart from a typical repeated-game setting, Takahashi (2010) and Deb (2011) consider the community enforcement and Miyagawa, Miyahara, and Sekiguchi (2008) consider the situation where a player can improve the precision of monitoring by paying cost.

defect after some history. Since player i observes g_i with probability 0.6 under player j's cooperation, what player i can expect for one year is to observe approximately $365 \times 0.6 \approx 200$ days of good signals. Hence, player i cannot punish player j after excessively many days of good signals (say, 250 days) since otherwise, punishment would be triggered too easily and efficiency would be destroyed.

Hence, if the monitoring is not conditionally independent, then players i and j need to coordinate on when player i should switch to defection. Previously, attempts to generalize Matsushima (2004) to conditionally dependent monitoring have shown only limited results because coordination is difficult in private monitoring.[3] In this paper, the players use cheap talk to overcome this difficulty. Note that, since player i's switch to defection hurts player j, constructing an incentive compatible equilibrium is not straightforward.

Second, Hörner and Olszewski (2006) show the folk theorem in a general game but with almost perfect monitoring. Hörner and Olszewski (2006) consider the following phase-belief-free equilibrium: they see the repeated game as a repetition of L-period review phase and the belief free property holds at the beginning of each review phase. However, the players coordinate their play within a phase.

Given our first generalization, it is natural to replace each period of Hörner and Olszewski (2006) with a T-period review round ($T = 365$ in our example above), and so consider a LT-period review phase. One difficulty to make this idea work is that, in the equilibrium of Hörner and Olszewski (2006), player i's optimal action in period $l \leq L$ depends on player j's history until period $l-1$. Hence, player i calculates the belief of player j's history from player i's history and takes an action. That is, player i's action in period l depends on player i's history until period $l-1$. Symmetrically, player j's action in period l depends on player j's history until period $l-1$.

If we replace one period of Hörner and Olszewski (2006) with a T-period review round, then player i's optimal action in round $l \leq L$ depends on player j's history until round $l-1$. At the same time, player j's action in round l depends on player j's history until round $l-1$.

[3] See Fong, Gossner, Hörner, and Sannikov (2010) and Sugaya (2012a).

Therefore, player i, which playing the stage game T times in round l, gradually learns player j's action, which affects player i's belief about player j's history, which, in turn, affects player i's belief about her optimal action. This belief update can be getting very complicated if T becomes large.

In this paper, we use the communication to simplify this belief calculation. At the end of round $l-1$, player i announces what action player i will take in round l. If this announcement is different from what would be optimal given player j's history (that is, player i announces a wrong action), then player j changes her strategy so that it is actually optimal for player i to follow player i's own announcement. To incentivize player i to tell the truth, we make sure that when player i announces a wrong action, player j punishes player i. This implies that, to know what announcement is correct, player i at the end of round $l-1$ needs to calculate the belief about player j's history. Hence, our paper is related to belief-based approach.[4]

The rest of the paper is organized as follows: Section 2 introduces the model and Section 3 states the assumptions and main result. Section 4 relates the infinitely repeated game to a finitely repeated game with an auxiliary scenario (reward function) and derives sufficient conditions on the finitely repeated game to show the folk theorem in the infinitely repeated game. The remaining parts of the paper are devoted to the proof of the sufficient conditions. Section 5 offers the overview of the structure of the proof. Section 6 defines the equilibrium. While defining the equilibrium, we define variables with various conditions. In Section 7, we verify that we take all the variables satisfying all the conditions. Section 8 finishes proving the sufficient conditions. Section 9 concludes. Some proofs are relegated to Appendix (Section 10).

[4] Among papers using the belief-based approach, Sekiguchi (1997) shows that the payoff of the mutual cooperation is approximately attainable and Bhaskar and Obara (2002) show the folk theorem in the prisoners' dilemma with almost perfect monitoring. Phelan and Skrzypacz (2012) characterize the set of possible beliefs about opponents' states in a finite-state automaton strategy and Kandori and Obara (2010) offer a way to verify if a finite-state automaton strategy is an equilibrium. However, without almost perfect or public monitoring, the belief calculation is too complicated to find an equilibrium to support the folk theorem.

2 Model

2.1 Stage Game

We consider the general multi-player repeated game, where the stage game is given by $\{I, \{A_i, Y_i, U_i\}_{i \in I}, q\}$. $I = \{1, ..., N\}$ is the set of players, A_i is the set of player i's pure actions, Y_i is the finite set of player i's private signals, and U_i is the finite set of player i's ex-post utilities. Let $A \equiv \prod_{i \in I} A_i$, $Y \equiv \prod_{i \in I} Y_i$ and $U \equiv \prod_{i \in I} U_i$ be the set of action profiles, signal profiles and ex post utility profiles, respectively.

In every stage game, player i chooses an action $a_i \in A_i$, which induces an action profile $a \equiv (a_1, ..., a_N) \in A$. Then, a signal profile $y \equiv (y_1, ..., y_N) \in Y$ and an ex post utility profile $\tilde{u} \equiv (\tilde{u}_1, ..., \tilde{u}_N) \in U$ are realized according to a joint conditional probability function $q(y, \tilde{u} \mid a)$.

Following the convention in the literature, we assume that $\tilde{u}_i$ is a deterministic function of a_i and y_i so that observing the ex post utility does not give any further information than (a_i, y_i). If this were not the case, then we could see a pair of a signal and an ex post utility $(y_i, \tilde{u}_i)$ as a new signal.

Given this, we see $q(y \mid a)$ as the conditional joint distribution of signal profiles. $q_i(y_i \mid a)$ denotes the marginal distribution of player i's signals derived from q. In addition, let

$$q_i(a) \equiv (q_i(y_i \mid a))_{y_i \in Y_i} \tag{1}$$

denote a $|Y_i| \times 1$ vector of player i's signal distribution given a.

Player i's expected payoff from $a \in A$ is the ex ante value of $\tilde{u}_i$ given a and is denoted by $u_i(a)$. Without loss, we assume

$$u_i(a) \geq 0 \tag{2}$$

for all $i \in I$ and $a \in A$. For each $a \in A$, let $u(a)$ represent the payoff vector $(u_i(a))_{i \in I}$.

2.2 Repeated Game

Consider the infinitely repeated game with the (common) discount factor $\delta \in (0,1)$. Let $a_{i,\tau}$, $y_{i,\tau}$ and m_τ respectively, denote the action played by player i in period τ, the private signal observed by player i in period τ and the message sent in period τ. Since the result of communication is public, m_τ does not have a player-specific index.

Player i's private history up to period $t \geq 1$ is given by $h_i^t \equiv \{a_{i,\tau}, y_{i,\tau}, m_\tau\}_{\tau=1}^{t-1}$. With $h_i^1 \equiv \{\emptyset\}$, for each $t \geq 1$, let H_i^t be the set of all h_i^t. As we will see, in each period t, player i first sends a message simultaneously, second takes an action simultaneously, and finally observes an signal. Hence, a strategy for player i is defined to be $\sigma_i \equiv (\sigma_i^a, \sigma_i^m)$ such that

- $\sigma_i^a : \bigcup_{t=1}^{\infty} H_i^t \to \triangle(A_i)$ maps player i's histories in period t at the instant when player i takes an action to player i's actions;

- $\sigma_i^m \equiv (\sigma_{i,t}^m)_{t=1}^{\infty}$, where $\sigma_{i,t}^m$ is player i's strategy $\sigma_{i,t}^m : H_i^t \times A_i \times Y_i \to \triangle(M_{i,t})$ which maps player i's histories in period t at the instant when player i sends a message (after taking $a_{i,t}$ and observing $y_{i,t}$) to player i's messages. The message space for each period, $M_{i,t}$, will be defined later.

Let Σ_i be the set of all strategies for player i.

Finally, let $E(\delta)$ be the set of sequential equilibrium payoffs with a common discount factor δ.

3 Assumptions and Result

In this section, we state the three assumptions and main result. First, we assume the full dimensionality condition. Let α^* be a static Nash equilibrium in the stage game with $v^* \equiv u(\alpha^*)$. If there are multiple, then the argument below holds for any arbitrarily fixed static Nash equilibrium α^*. Then, the Nash-threat feasible payoff set is given by

$$F^{\text{Nash}} = \{v \in \mathbb{R}^N : v \in \text{co}(\{u(a)\}_{a\in A}) \text{ and } v_i \geq v_i^* \text{ for all } i\}.$$

We assume F^{Nash} has full dimension:

Assumption 1 F^{Nash} *has full dimension:* $\dim(F^{\text{Nash}}) = N$.

Second, we assume that the marginal distribution of the signals has full support:

Assumption 2 *For any* $i \in I$, $a \in A$ *and* $y_i \in Y_i$, $q_i(y_i \mid a) > 0$.

Third, we assume that player i's signal statistically identifies player j's action (see (1) for the definition of $q_i(a)$):

Assumption 3 *For any* $j \in I$, *there exists* $i \in -j$ *such that, for all* $a \in A$, *the collection of* $|Y_i|$*-dimensional vectors* $(q_i(a_j, a_{-j}))_{a_j \in A_j}$ *is linearly independent with respect to* a_j.

If cheap talk communication devices are available and these three assumptions are satisfied, then we can show that any payoff profile in F^{Nash} is sustainable in a sequential equilibrium.

Theorem 1 *If cheap talk communication devices are available and Assumptions 1, 2 and 3 are satisfied, then for any* $v \in \text{int}(F^{\text{Nash}})$, *there exists* $\bar{\delta} < 1$ *such that, for all* $\delta > \bar{\delta}$, $v \in E(\delta)$.

Two remarks: first, with Assumption 2, the set of sequential equilibrium payoffs is equal to that of Nash equilibrium payoffs. See Sekiguchi (1997) for the proof. Hence, we will consider Nash equilibrium below.

Second, all the assumptions are generic if $|Y_i| \geq |A_j|$ for all i and j. Especially, we can allow public monitoring with $|Y| = \max_{i \in I} |A_i|$, where Y is the set of public signals. Hence, it is the restricted attention on the perfect public equilibrium that causes efficiency loss in Radner, Myerson, and Maskin (1986).[5]

[5] Kandori and Obara (2006) create an equilibrium with private strategies that Pareto-dominates the most efficient public perfect equilibrium. However, their equilibrium cannot support the mutual cooperation payoff unless there exists an action after taking which a player can identify the other player's defection almost perfectly.

4 Finitely Repeated Game

As we see in the Introduction, we see repeated game as a repetition of T_P-period review phase with $T_P \equiv LT$, where L is the number of review rounds and T is the length of review round. Instead of considering the infinitely repeated game directly, we consider T_P-period finitely repeated game with a "reward function." Intuitively, a finitely repeated game corresponds to a review phase in the infinitely repeated game and a reward function corresponds to changes in the continuation payoff.

We derive sufficient conditions on strategies and reward functions in the finitely repeated game such that we can construct a strategy in the infinitely repeated game to support the targeted payoff v. The sufficient conditions are summarized in Lemma 1, which are the same sufficient conditions for the block equilibrium in Hörner and Olszewski (2006) to work. In other words, the main contribution of this paper is to offer the proof of these sufficient conditions in a general private monitoring from the next section while Hörner and Olszewski (2006) consider almost perfect monitoring.

Let $\sigma_i^{T_P} \equiv (\sigma_i^{a,T_P}, \sigma_i^{m,T_P})$ with $\sigma_i^{a,T_P} : \bigcup_{t=1}^{T_P} H_i^t \to \triangle(A_i)$, $\sigma_i^{m,T_P} \equiv (\sigma_{i,t}^m)_{t=1}^{T_P}$, $\sigma_{i,t}^m : H_i^t \times A_i \times Y_i \to \triangle(M_{i,t})$ and be player i's strategy in the finitely repeated game and $\Sigma_i^{T_P}$ be the set of all strategies in the finitely repeated game. Each player i has a state $x_i \in \{G, B\}$. In state x_i, player i plays $\sigma_i(x_i) \in \Sigma_i^{T_P}$.

In addition, each player i with x_i gives a "reward function" $\pi_{i+1}(x_i, \cdot : \delta) : H_i^{T_P+1} \to \mathbb{R}$ to player $i+1$, that is, the reward function is a mapping from player i's histories in the finitely repeated game to the real numbers. Throughout the paper, we identify player $n \notin \{1, ..., N\}$ with player $n \pmod N$.

Our task is to find $\{\sigma_i(x_i)\}_{x_i,i}$ and $\{\pi_{i+1}(x_i, \cdot : \delta)\}_{x_i,i}$ such that, for each $i \in I$, there are two numbers $\underline{v}_i$ and $\bar{v}_i$ to contain v between them:

$$\underline{v}_i < v_i < \bar{v}_i, \tag{3}$$

such that there exists T_P with $\lim_{\delta \to 1} \delta^{T_P} = 1$, and such that the following conditions are satisfied: for sufficiently large δ, for any $i \in I$,

1. for any combination of the other players' states $x_{-i} \equiv (x_n)_{n \neq i} \in \{G, B\}^{N-1}$, it is optimal to take $\sigma_i(G)$ and $\sigma_i(B)$: for any $x_{-i} \in \{G, B\}^{N-1}$,

$$\sigma_i(G), \sigma_i(B) \in \arg \max_{\sigma_i^{T_P} \in \Sigma_i^{T_P}} \mathbb{E}\left[\sum_{t=1}^{T_P} \delta^{t-1} u_i(a_t) + \pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta) \mid \sigma_i^{T_P}, \sigma_{-i}(x_{-i})\right]; \tag{4}$$

2. regardless of $x_{-(i-1)}$, the discounted average of the expected sum player i's instantaneous utilities and player $i-1$'s reward function on player i is equal to $\bar{v}_i$ if player $i-1$'s state is good and equal to $\underline{v}_i$ if player $i-1$'s state is bad: for all $x_{-(i-1)} \in \{G, B\}^{N-1}$,

$$\frac{1-\delta}{1-\delta^{T_P}} \mathbb{E}\left[\sum_{t=1}^{T_P} \delta^{t-1} u_i(a_t) + \pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta) \mid \sigma(x)\right] = \begin{cases} \bar{v}_i \text{ if } x_{i-1} = G, \\ \underline{v}_i \text{ if } x_{i-1} = B. \end{cases} \tag{5}$$

   Intuitively, since $\lim_{\delta \to 1} \frac{1-\delta}{1-\delta^{T_P}} = \frac{1}{T_P}$, this requires that player i's payoff is solely controlled by player $i-1$ and is close to the targeted payoffs $\underline{v}_i$ and $\bar{v}_i$;

3. $\frac{1-\delta}{\delta^{T^P}}$ converges to 0 faster than $\pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta)$ diverges and the sign of $\pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta)$ satisfies a proper condition:

$$\begin{cases} \lim_{\delta \to 1} \frac{1-\delta}{\delta^{T_P}} \sup_{x_{i-1}, h_{i-1}^{T_P+1}} \left|\pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta)\right| = 0, \\ \pi_i(G, h_{i-1}^{T_P+1} : \delta) \leq 0, \\ \pi_i(B, h_{i-1}^{T_P+1} : \delta) \geq 0. \end{cases} \tag{6}$$

   We call (6) the "feasibility constraint."

We explain why these conditions are sufficient. We see the infinitely repeated game as the repetition of T_P-period review phases. In each review phase, each player i has two possible states $\{G, B\} \ni x_i$ and player i with state x_i takes $\sigma_i(x_i)$ in the phase. (4) implies that both $\sigma_i(G)$ and $\sigma_i(B)$ are optimal regardless of the other players' states. (5) implies that player

i's ex ante value at the beginning of the phase is solely determined by player $i-1$'s state. That is, player $i-1$ is a controller of player i's payoff.

Here, $\pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta)$ represents the differences between player i's ex ante value given x_{i-1} at the beginning of the phase and the ex post value at the end of the phase after player $i-1$ observes $h_{i-1}^{T_P+1}$. $\pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta) = 0$ implies that the ex post value is the same as the ex ante value since player $i-1$ transits to the same state in the next phase with probability one. With $x_{i-1} = G$ (B, respectively), the smaller $\pi_i(G, h_{i-1}^{T_P+1} : \delta)$ (the larger $\pi_i(B, h_{i-1}^{T_P+1} : \delta)$, respectively), the more likely it is for player $i-1$ to transit to the opposite state B (G, respectively) in the next phase. The feasibility of this transition is guaranteed by (6).

The following lemma summarizes the discussion:

Lemma 1 *For Theorem 1, it suffices to show that, for any $v \in \text{int}(F^{\text{Nash}})$, for sufficiently large δ, there exist $\{\underline{v}_i, \bar{v}_i\}_{i\in I}$ with (3), T_P with $\lim_{\delta\to 1}\delta^{T_P} = 1$ and $\{\{\sigma_i(x_i)\}_{x_i\in\{G,B\}}\}_{i\in I}$ and $\{\{\pi_i(x_{i-1}, \cdot : \delta)\}_{x_{i-1}\in\{G,B\}}\}_{i\in I}$ such that (4), (5) and (6) are satisfied.*

Proof. See Section 10.1. ■

From now on, when we say player i's action plan, it means player i's behavioral mixed strategy $\sigma_i(x_i)$ within the current review phase (or, the finitely repeated game). On the other hand, when we say player i's strategy, it contains both $\sigma_i(x_i)$ and $\pi_{i+1}(x_i, \cdot : \delta)$ which determines player i's entire strategy in the infinitely repeated game.

Let us specify $\underline{v}_i$ and $\bar{v}_i$. This step is the same as Hörner and Olszewski (2006). Given $x \in \{G, B\}^N$, pick 2^N action profiles $\{a(x)\}_{x\in\{G,B\}^N}$. As we have mentioned, player $i-1$'s state x_{i-1} refers to player i's payoff and indicates whether this payoff is strictly above or below v_i no matter what the other players' states are. That is, player $i-1$'s state controls player i's payoff. Formally,

$$\max_{x:x_{i-1}=B} u_i(a(x)) < v_i < \min_{x:x_{i-1}=G} u_i(a(x)) \text{ for all } i \in I.$$

Take $\underline{v}_i$ and $\overline{v}_i$ such that

$$\max\left\{v_i^*, \max_{x:x_{i-1}=B} u_i(a(x))\right\} < \underline{v}_i < v_i < \overline{v}_i < \min_{x:x_{i-1}=G} u_i(a(x)). \tag{7}$$

Remember that v_i^* is a static Nash equilibrium payoff.

Action profiles that satisfy the desired inequalities may not exist. However, if Assumption 1 is satisfied, then there always exist an integer z and 2^z finite sequences $\{a_1(x), \ldots, a_z(x)\}_{x\in\{G,B\}^N}$ such that each vector $w_i(x)$, the average discounted payoff vector over the sequence $\{a_1(x), \ldots, a_z(x)\}_{x\in\{G,B\}^N}$, satisfies the appropriate inequalities provided δ is close enough to 1. The construction that follows must then be modified by replacing each action profile $a(x)$ by the finite sequence of action profiles $\{a_1(x), \ldots, a_z(x)\}_{x\in\{G,B\}^N}$. Details are omitted as in Hörner and Olszewski (2006).

Given $\rho > 0$ that will be determined in Section 7, given $a(x)$, for each i, we perturb $a_i(x)$ to $\alpha_i(x)$ so that player i takes all the actions in A_i with a positive probability no less than 2ρ:

$$\alpha_i(x) \equiv \left(1 - \textstyle\sum_{a_i \neq a_i(x)} 2\rho\right) a_i(x) + \textstyle\sum_{a_i \neq a_i(x)} 2\rho a_i.$$

Let $\{w(x)\}_{x\in\{G,B\}^N}$ be the corresponding payoff vectors under $\alpha(x)$:

$$w(x) \equiv u(\alpha(x)) \text{ with } x \in \{G, B\}^N.$$

With sufficiently small ρ, (7) implies

$$\max\left\{v_i^*, \max_{x:x_{i-1}=B} w_i(x)\right\} < \underline{v}_i < v_i < \overline{v}_i < \min_{x:x_{i-1}=G} w_i(x). \tag{8}$$

Below, we construct $\{\sigma_i(x_i)\}_{x_i,i}$ and $\{\pi_i(x_{i-1}, \cdot : \delta)\}_{x_{i-1},i}$ satisfying (4), (5) and (6) with $\bar{v}_i$ and $\underline{v}_i$ defined above in the finitely repeated game.

5 Overview of the Argument

This section provides an intuitive explanation for our construction. In this section, we focus on the two-player prisoners' dilemma and we assume v_i is arbitrarily close to the mutual cooperation payoff. This implies that we need to show the sufficient conditions with $\bar{v}_i$ close to $u_i(C_i, C_i)$:

$$\bar{v}_i \approx u_i(C_i, C_i), \tag{9}$$

and so we take $a_i(G, G) = C_i$. With two players, whenever we say players i and j, we assume players i and j are different: $i \neq j$.

Further, in this intuitive explanation, let us assume that $Y_i = \{g_i, b_i\}$, that is, player i has two possible signals, "good" and "bad," and that the good signal is more likely to occur with the opponent's cooperation: for all $a_i \in A_i$,

$$q_i(g_i \mid a_i, C_j) > q_i(g_i \mid a_i, D_j). \tag{10}$$

5.1 Structure of the Review Phase

In T_P-period finitely repeated games, at the beginning, each player i simultaneously announces a state $x_i \in \{G, B\}$ by cheap talk. $\sigma_i(x_i)$ tells the truth about x_i.[6] Now, the players have coordinated on the state profile x.

Based on this coordination, the players play the finitely repeated game for T_P periods. We see T_P periods as L repetitions of T-period review rounds, that is, $T_P = LT$. Here, we take

$$T = (1 - \delta)^{-\frac{1}{2}}$$

so that

$$T \to \infty \text{ and } \delta^{LT} \to 1 \text{ as } \delta \to 1 \tag{11}$$

for any finite L. (See Section 7 for the definition of L.) Intuitively, if the discount factor

[6] In the specification in Section 2.2, the players send the message at the end of each period. We see that the players end x_i at the end of the last period of the previous phase.

is large, then T is sufficiently long to aggregate information efficiently and, at the same time, the discounting over the finitely repeated game is negligible since δ^{LT} goes to unity. Throughout the paper, we neglect the integer problem since it is handled by replacing each variable s that should be an integer with $\min_{n\in\mathbb{N},n\geq s} n$.

5.2 Review Rounds

Let us now explain the review rounds. Look at the sufficient conditions in Section 4. (4) implies that player i wants to maximize

$$\frac{1-\delta}{1-\delta^{T_P}}\mathbb{E}\left[\sum_{t=1}^{T_P}\delta^{t-1}u_i\left(a_t\right)+\pi_i(x_j,h_j^{T_P+1}:\delta)\mid\sigma_i,\sigma_j(x)\right]$$

with $i-1=j$ in the two-player game. For sufficiently large δ, this is approximately equal to

$$\frac{1}{T_P}\mathbb{E}\left[\sum_{t=1}^{T_P}u_i\left(a_t\right)+\pi_i(x_j,h_j^{T_P+1})\mid\sigma_i,\sigma_j(x)\right]. \tag{12}$$

In this section, we assume that the players did not discount the future. Intuitively speaking, with a sufficiently high discount factor, we can replicate the situation where discount factor is unity by slightly adjusting the change of player i's continuation payoff.[7] Hence, in Section 5, we neglect discounting and heuristically assume that $\delta=1$ and that player i maximizes (12). See Condition 2 of Lemma 2 and Section 6.4 for the formal treatment of discounting.

On the other hand, (5) together with (9) implies that

$$\begin{aligned}&\frac{1-\delta}{1-\delta^{T_P}}\mathbb{E}\left[\sum_{t=1}^{T_P}\delta^{t-1}u_i\left(a_t\right)+\pi_i(x_j,h_j^{T_P+1}:\delta)\mid\sigma(G,G)\right]\\ \approx\;&\frac{1}{T_P}\mathbb{E}\left[\sum_{t=1}^{T_P}u_i\left(a_t\right)+\pi_i(x_j,h_j^{T_P+1})\mid\sigma(G,G)\right]\approx u_i(C,C).\end{aligned} \tag{13}$$

[7] Note, however, that known results in the case without discounting (for example, see Lehrer (1990)) cannot be extended to the case with discounting. It is the phase-belief-free property that allows us this adjustment.

Together with the feasibility constraint, this implies that we need to satisfy the following two requirements: player j incentivizes player i to take cooperation with a high probability, and if player i cooperates frequently, then the punishment $\pi_i(x_j, h_j^{T_P+1})$ should be close to zero in expectation.

To this end, player j aggregates information in each review round. Suppose now the players are in round l and player j takes $\alpha_j(l) \in \Delta(A_j)$ and expects player i to take $\alpha_i(l) \in \Delta(A_i)$ in each period of round l (as will be seen, the players take *i.i.d.* mixed actions within each review round). Since Assumption 3 implies that player j can statistically identify player i's action, player j can map her history in each period into a real number $\pi_i[\alpha(l)](y_j)$ so that

$$\mathbb{E}\left[u_i\left(a_i, \alpha_j\right) + \pi_i[\alpha(l)]\left(y_j\right) \mid a_i, \alpha_j(l)\right] \tag{14}$$

is independent of $a_i \in A_i$. Intuitively, conditional on $\alpha_j(l)$, after observing a "good" signal g_j which occurs more likely after player i's cooperation, player j gives a high point $\pi_i[\alpha(l)](y_j)$ while after observing a "bad" signal b_j which occurs more likely after player i's defection, player j gives a low point $\pi_i[\alpha(l)](y_j)$, so that the expected gain in points from cooperation cancels out the loss in instantaneous utilities. We normalize $\pi_i[\alpha(l)](y_j)$ by adding or subtracting a constant so that

$$\mathbb{E}\left[\pi_i[\alpha(l)]\left(y_j\right) \mid \alpha(l)\right] = 0. \tag{15}$$

Further, take $\bar{u}$ sufficiently large so that[8]

$$\bar{u} > \max_{j \in I, \alpha \in \Delta(A), y_j \in Y_j} \left|\pi_i[\alpha]\left(y_j\right)\right|. \tag{16}$$

Recall that we have L review rounds. For each round l, player j aggregates $\pi_i[\alpha(l)](y_{j,t})$

[8] Lemma 2 shows that the maximum is well defined.

and creates player j's score about player i:

$$X_j(l) = \sum\nolimits_{t:\ l\text{th review round}} \pi_i[\alpha(l)](y_{j,t}). \tag{17}$$

5.2.1 Conditional Independence

For a moment, assume that player i's signals were independent of player j's signals conditional on any action profile a, as in Matsushima (2004). In addition, we only explain how to construct an ε-sequential equilibrium in this subsubsection: an strategy profile consists an ε-sequential equilibrium if, for any player i, after any history h_i^t that happens with a positive probability, the gain of a deviation is bounded by ε. The reason is just for simple exposition: to convey the contrast between conditionally independent monitoring and conditionally dependent monitoring, it is enough to consider an ε-sequential equilibrium. Note that we will construct exact equilibrium that works for any correlation from next subsubsection.

To construct ε-sequential equilibrium with conditionally independent monitoring, we can assume that player i takes the same behavioral mixed strategy $\alpha_i(x)$ for all the rounds:

$$\alpha_i(l) = \alpha_i(x) \tag{18}$$

for all $l \in \{1, ..., L\}$. Especially, since we focus on $x = (G, G)$, $\alpha_i(x) = (1 - 2\rho) C_i + 2\rho D_i$ with a small $\rho > 0$, that is, player i takes C_i with high probability $1 - 2\rho$. Intuitively, player j, who also takes $\alpha_j(x)$ symmetrically defined, wants to incentivize player i to take $\alpha_i(x)$ by aggregating information over the review round and the expected punishment should be small if player i takes C_i frequently.

This can be done as follows. Define the reward function as

$$\pi_i(x_j, h_j^{T_P+1}) = \left\{-2\bar{u}T + \sum\nolimits_{l=1}^{L} X_j(l)\right\}_{-}, \tag{19}$$

where, in general, $\{X\}_{-}$ is equal to X if $X \leq 0$ and 0 otherwise.

Now let us check (12) and (13). From (15), the expected increase in the score in each period (that is, the expected point) is non-positive. Therefore, by the law of large numbers, given $\bar{u} > 0$ and $L \in \mathbb{N}$, for sufficiently large T, player i puts a high belief on the event that $\sum_{l=1}^{L} X_j(l) \leq 2\bar{u}T$. Since player i's signals are independent of player j's signals, player i cannot update any information about player j's score from player i's history.[9] Therefore, after any period τ and history h_i^τ, player i believes $\pi_i(x_j, h_j^{T_P+1}) = -2\bar{u}T + \sum_{l=1}^{L} X_j(l)$ with a high probability. Together with (14), player i believes that both cooperation and defection are optimal with a high probability. Hence, (12) is satisfied.

At the same time, since (18) implies $\alpha(l) = \alpha(x)$ and the expected value of $\pi_i[\alpha(l)](y_j)$ with $\alpha(l) = \alpha(x)$ is 0,

$$\begin{aligned} & \frac{1}{T_P}\mathbb{E}\left[\sum_{t=1}^{T_P} u_i(a_t) + \pi_i(x_j, h_j^{T_P+1}) \mid \sigma(G,G)\right] \\ \approx\ & \frac{1}{T_P}\left(\sum_{t=1}^{T_P} T_P u_i(\alpha(x)) - 2\bar{u}T\right) = u_i(\alpha(x)) - \frac{2}{L}\bar{u}. \end{aligned}$$

For sufficiently small ρ and L, this is close to $u_i(C,C)$ and so (13) is satisfied. Therefore, we are done.

5.2.2 Conditional Dependence

Now, assume that player i's signals and player j's signals can be arbitrarily correlated. Since player i with $\alpha_i(x)$ takes cooperation with a high probability, (15) implies that the expected score is close to 0 if player i always cooperate. Hence, to prevent an inefficient punishment, player j cannot punish player i after the score is excessively high (in the above example, more than $2\bar{u}T$). On the other hand, if the signals are correlated, then after some history of player i, judging from her own history and correlation, player i puts a high probability on the event that player j's score about player i has already been excessively high. Then, player i wants to start to defect.

[9] Precisely speaking, player i can update the realization of the score from the realization of her own mixture $\alpha_i(x)$ and learning the realization of player j's mixture from y_i. We omit the details of the proof since our construction for the general monitoring covers the conditionally independent monitoring.

More generally, there are correlations with which it is impossible to create a punishment schedule that is approximately efficient and that at the same time incentivizes player i to cooperate after any history.[10] Hence, we need to let player i's incentive to cooperate break down after some history. Symmetrically, player j also switches her own action after some history.

Reflective Learning Problem One may think that it solves the problem to define the equilibrium strategy so that player i defects after player i's expectation of player j's score about player i is much higher than the ex ante mean. If this were incentive compatible, then since such a history happens only rarely, the equilibrium strategy attains efficiency.

However, this creates the following problem: since player i switches her action based on player i's expectation of player j's score about player i, player i's action reveals player i's expectation of player j's score about player i. Since both "player i's expectation of player j's score about player i" and "player i's score about player j" are calculated from player i's history, player j, who wants to infer player i's score about player j, may have an incentive to learn "player i's expectation of player j's score about player i" by privately monitoring player i's action. If so, player j's decision of actions depends also on player j's expectation of player i's expectation of player j's score about player i. Proceeding one step further, player i's decision of actions depends on player i's expectation of player j's expectation of player i's expectation of player j's score about player i. This chain of "reflective learning" continues infinitely, and it is hard to analyze when the players have an incentive to cooperate.[11]

Cheap Talk to Coordinate the Continuation Play Without Reflecting Learning We want to construct an equilibrium that is immune to the reflective learning. To this end, the players use cheap talk to coordinate on the continuation play. Especially, we see the finitely repeated game as L repetition of T-period review rounds and at the beginning of each round, the players simultaneously send the histories within the previous round by cheap

[10] The formal proof of this claim is available upon request.

[11] Fong, Gossner, Hörner, and Sannikov (2010) take this approach and show that, under some open set of monitoring structure, the mutual cooperation is sustainable with a sufficiently high probability.

talk. Since the result of the cheap talk is common knowledge, there is no learning problem if the players have the incentive to tell the truth.

Now we will explain the equilibrium strategy. For that purpose, it is convenient to define a state in each round l, constructed from the messages of the players: $\lambda(l) \in \{\emptyset\} \cup I$. Intuitively, $\lambda(l) = \emptyset$ implies that no player i has announced that player i believes that player j's score has been excessively high until round $l-1$; $\lambda(l) = i \in \{1, 2\}$ implies that there is only one player i who has announced so. If both players announced so, we ignore the announcement and have $\lambda(l) = \emptyset$. The optimality of this ignorance will be verified later.

Let us first define player i's action plan in each round l. If $i \neq \lambda(l)$, then, intuitively, since player i believes that the score has been regular until round $l-1$, player i believes that there is enough room for the score to linearly increase with respect to the points without hitting zero until the end of round l. That is, player i is indifferent between C_i and D_i within round l. Given this indifference, we can let player i take one of the following three mixed action plans

$$\alpha_i(l) = \begin{cases} \alpha_i(x) \equiv (1-2\rho)\, C_i + 2\rho D_i, \\ \bar{\alpha}_i(x) \equiv (1-\rho)\, C_i + \rho D_i, \\ \underline{\alpha}_i(x) \equiv (1-3\rho)\, C_i + 3\rho D_i, \end{cases} \tag{20}$$

with probability $1-\eta$, $\eta/2$ and $\eta/2$, respectively, with small $\rho, \eta > 0$. Once player i decides $\alpha_i(l)$, player i takes an action according to $\alpha_i(l)$ *i.i.d.* within round l. Note that player i takes the same action plan $\alpha_i(x)$ as in the case with conditional independence with a high probability. However, with a small probability, player i puts more weight ($\bar{\alpha}_i(x)$) or less weight ($\underline{\alpha}_i(x)$) on cooperation.

On the other hand, if $i = \lambda(l)$, then player i believes that the score cannot increase linearly in the points without hitting zero and player i takes D_i with probability one: $\alpha_i(l) = D_i$. See ?? in Figure 1 for the illustration (we will explain the last column later).

Player j's action plan is symmetrically defined: if $j \neq \lambda(l)$, then

$$\alpha_j(l) = \begin{cases} \alpha_j(x) \equiv (1-2\rho) C_j + 2\rho D_j, \\ \bar{\alpha}_j(x) \equiv (1-\rho) C_j + \rho D_j, \\ \underline{\alpha}_j(x) \equiv (1-3\rho) C_j + 3\rho D_j, \end{cases} \tag{21}$$

with probability $1-\eta$, $\eta/2$ and $\eta/2$, respectively; if $j = \lambda(l)$, then $\alpha_j(l) = D_j$.

After round l, player i sends her history in round l, $\{a_{i,t}, y_{i,t}\}_{t:\text{round } l}$, to player j by cheap talk. Let $h_i^l \equiv \{a_{i,t}, y_{i,t}\}_{t:\text{round } l}$ be the true history and $\hat{h}_i^l \equiv \{\hat{a}_{i,t}, \hat{y}_{i,t}\}_{t:\text{round } l}$ be the reported history.

Hence, the strategies of the players are characterized as follows: for each l,

1. from period $(l-1)T+1$ (the initial period of round l) to period $lT-1$ (the second last period of round l), player i takes $\alpha_i(l)$ as explained above. As will be seen below, since $\lambda(l)$ is well-defined by player i's history at the beginning of round l, $\sigma_{i,t}^a$ is well-defined. Since $M_{i,t} = \emptyset$, $\sigma_{i,t}^m$ is redundant;

2. from period lT (the last period of round l), player i takes $\alpha_i(l)$ as explained above. Then, with $M_{i,t} = (A_i \times Y_i)^T$, $\sigma_{i,t}^m$ sends $\hat{h}_1^l = h_i^l = \{a_{i,t}, y_{i,t}\}_{t:\text{round } l}$.

To define player i's action plan, we are left to define the transition of $\lambda(l)$. The initial condition is $\lambda(1) \neq \emptyset$. For $l \geq 1$, if $\lambda(l) \neq \emptyset$, then $\lambda(l+1) = \lambda(l)$. Hence, let us concentrate on the case with $\lambda(l) = \emptyset$. Remember that $\lambda(l+1) = i$ means that player i believes that player j's score about player i has been regular until round l. Therefore, we will specify after what history h_i^l, player i believes that $X_j(l)$ was regular in round l.

Player i believes $X_j(l)$ is regular if both of the following two conditions are satisfied:

1. player i picks $\alpha_i(l) = \alpha_i(x)$;

2. the realized frequency of player i's action in round l is actually close to $\alpha_i(x)$.

Let us call these two conditions Conditions 1 and 2 for the belief of $X_j(l)$. Otherwise, player i believes that $X_j(l)$ was excessively high.

Given this, $\lambda(l+1)$ is determined as follows:

- if there is unique i such that $\hat{h}_i^l$ does not satisfy either Condition 1 or 2, then $\lambda(l+1) = i$;
- otherwise, $\lambda(l+1) = \emptyset$.

Now, since we have defined player i's action plan, let us define player j's reward function on player i. While taking the action, player j calculates $X_j(l)$. Player j uses her true action $\alpha_j(l)$ while she assumes that player i takes $\alpha_i(x)$ for round l:

$$X_j(l) = \sum\nolimits_{t:\ l\text{th review round}} \pi_i[\alpha(l)](y_{j,t}). \tag{22}$$

We say player j's score about player i is "regular" in round l if $X_j(l) \leq \frac{\bar{u}}{L}T$ and it is "excessively high" if $X_j(l) > \frac{\bar{u}}{L}T$. See Figure 2 for the illustration.

Intuitively, if $\lambda(l) \neq i$ (player i announced that she believes the score has been regular), then the reward is linear in the score (and so both C_i and D_i are optimal) while if $\lambda(l) = i$ (player i announced that she believes the score has been excessively high), then the reward is constant (and so player i wants to take defection):

$$-2\bar{u}T + \underbrace{\sum\nolimits_{l:\lambda(l)\neq i} X_j(l)}_{\substack{\text{player } j \text{ adds the score} \\ \text{for rounds with } \lambda(l)\neq i}} - \underbrace{\sum\nolimits_{l:i\lambda(l)=i} T(u_i(D_i,\alpha_j(l)) - u_i(\alpha_i(x),\alpha_j(l)))}_{\substack{\text{As will be seen, for rounds with } \lambda(l)=i, \\ \text{player } i \text{ takes } D_i \text{ instead of } \alpha_i(x). \\ \text{The marginal gain of } D_i \text{ is canceled out.}}}.$$

However, there are two events after which player j subtracts a large number $\bar{u}LT + \sum_l \left(\min_{\alpha_j} Tu_i(\alpha_i(x),\alpha_j) - Tu_i(\alpha_i(x),\alpha_j(l))\right)$:

1. there is round l where (i) $\lambda(l) = \emptyset$, (ii) the true score was excessively high in round l, that is, $X_j(l) > \frac{\bar{u}}{L}T$, and (iii) player i announced that player i believes that player j's score was regular, that is, $\hat{h}_i^l$ satisfied Conditions 1 and 2 for the belief of $X_j(l)$. Intuitively, this means player i made a mistake in the announcement;

2. there is round l where (i) $\lambda(l) = \emptyset$ and (ii) player j took $\alpha_j(l) \neq \alpha_j(x)$ or the actual frequency of player j's action was not close to $\alpha_j(x)$. In other words, $\lambda(l) = \emptyset$ and

player j's history h_j^l did not satisfy Conditions 1 and 2 for the belief of $X_i(l)$. Especially, if player j's announcement $\hat{h}_j^l$ does not satisfy either Condition 1 or 2 for the belief of $X_i(l)$ with $\lambda(l) = \emptyset$, then this condition is satisfied.

Let us call these conditions Conditions 1 and 2 for $\theta_j = B$. We say $\theta_j = B$ if at least one of these two conditions is satisfied and $\theta_j = G$ otherwise. Note that θ_j does not change if $\lambda(l) \neq \emptyset$. Note also that if $\lambda(l) = j$ for some l, then $\theta_j = B$.

In total, the reward is

$$\begin{aligned} \pi_i(x_j, h_j^{T_P+1}) &= -\mathbf{1}_{\{\theta_j = B\}} \left(\bar{u}LT + \sum\nolimits_l \left(\min_{\alpha_j} Tu_i(\alpha_i(x), \alpha_j) - Tu_i(\alpha_i(x), \alpha_j(l)) \right) \right) \\ & \quad -2\bar{u}T + \sum\nolimits_{l:\lambda(l)\neq i} X_j(l) \\ & \quad - \sum\nolimits_{l:i\lambda(l)=i} T(u_i(D_i, \alpha_j(l)) - u_i(\alpha_i(x), \alpha_j(l))). \end{aligned} \tag{23}$$

Intuitively,

$$\begin{cases} X_j(l) & \text{if } \lambda(l) \neq i, \\ \text{constant} & \text{if } \lambda(l) = i \end{cases} \tag{24}$$

is the reward in round l.

Note that this reward is always non-positive: the last line is always non-positive; the second line is non-positive if $\lambda(l+1) = i$ after $X_j(l) > \frac{\bar{u}}{L}T$; if $\lambda(l+1) \neq i$ after $X_j(l) > \frac{\bar{u}}{L}T$, then player i announced that she believes the score was regular (Conditions 1 and 2 for the belief of $X_j(l)$ are satisfied) after round l even though the score was excessively high, or player j announced that she believes the score was excessively high (Conditions 1 and 2 for the belief of $X_i(l)$ are not satisfied). Both imply $\theta_j = B$. From (16), $-\bar{u}LT$ is sufficiently small to make the reward non-positive since $\min_{\alpha_j} Tu_i(\alpha_i(x), \alpha_j) - Tu_i(\alpha_i(x), \alpha_j(l))$ is always non-positive.

Hence, we are left to check player i's incentive and efficiency. Consider the incentive first. If $\lambda(l) = i$, then (i) $\lambda(\tilde{l})$ with $\tilde{l} > l$ does not change and (ii) θ_j does not change either. Hence, (24) implies that D_i is optimal. If $\lambda(l) = j$, then again (i) $\lambda(\tilde{l})$ with $\tilde{l} > l$ does not change and (ii) θ_j does not change either. Hence, (24) implies that C_i and D_i are both optimal.

Since $\lambda(\tilde{l})$ with $\tilde{l} > l$ does not depend on $\hat{h}_i^l$, it is optimal for player i to tell the truth.

Hence, we concentrate on the case with $\lambda(l) = \emptyset$. We proceed by backward induction. In round L (the last round), the relevant part of (23) is $X_j(L)$. Hence, (14) and (22) imply that both C_i and D_i are optimal. The message $\hat{h}_i^L$ is irrelevant.

Consider round $L-1$. The relevant part of (23) is

$$X(L-1) - \mathbf{1}_{\{\theta_j = B\}} \left(\bar{u}LT + \sum_{l=1}^{L} \left(\min_{\alpha_j} Tu_i(\alpha_i(x), \alpha_j) - Tu_i(\alpha_i(x), \alpha_j(l)) \right) \right)$$
$$+ \begin{cases} X_j(L) & \text{if } \lambda(L) \neq i, \\ -T(u_i(D_i, \alpha_j(L)) - u_i(\alpha_i(x), \alpha_j(L))) & \text{if } \lambda(L) = i. \end{cases}$$

First, we consider player i's value in round L defined as

$$-\mathbf{1}_{\{\theta_j = B\}} \left(\min_{\alpha_j} Tu_i(\alpha_i(x), \alpha_j) - Tu_i(\alpha_i(x), \alpha_j(L)) \right)$$
$$+ \begin{cases} X_j(L) & \text{if } \lambda(L) \neq i, \\ -T(u_i(D_i, \alpha_j(L)) - u_i(\alpha_i(x), \alpha_j(L))) & \text{if } \lambda(L) = i. \end{cases}$$

- if $\theta_j = G$, then
  - if $\lambda(L) \neq i$, then from the discussion about round L, both C_i and D_i are optimal if $\lambda(L) \neq i$. Hence, we can calculate player i's value in round L with $\lambda(L) \neq i$, assuming player i takes $\alpha_i(x)$. In that case, the expectation of $X_j(L)$ is zero from (15) and (22). Hence, the value in round L is equal to $Tu_i(\alpha_i(x), \alpha_j(L))$;
  - if $\lambda(L) = i$, then player i will take D_i and the value is $Tu_i(\alpha_i(x), \alpha_j(L))$ since the deviation gain is canceled out by the reward function;

  Hence, regardless of $\lambda(L)$, player i's value in round L is $Tu_i(\alpha_i(x), \alpha_j(L))$. Since $\theta_j = G$, $\lambda(L) \neq j$ and so player j takes $\alpha_j(x)$, $\bar{\alpha}_j(x)$ and $\underline{\alpha}_j(x)$ with probability $1-\eta$, $\eta/2$ and $\eta/2$, with which the expected value of $Tu_i(\alpha_i(x), \alpha_j(L))$ is equal to $Tu_i(\alpha(x))$.

- if $\theta_j = B$, then by the same argument for the case with $\theta_j = G$, player i's payoff is $\min_{\alpha_j} Tu_i(\alpha_i(x), \alpha_j) - Tu_i(\alpha_i(x), \alpha_j(L)) + Tu_i(\alpha_i(x), \alpha_j(L)) = \min_{\alpha_j} Tu_i(\alpha_i(x), \alpha_j)$. Hence, regardless of $\lambda(L)$, player i's value in round L is $\min_{\alpha_j} Tu_i(\alpha_i(x), \alpha_j)$.

Therefore, we can conclude the relevant movement of the continuation payoff from round L is

$$\begin{aligned} & X(L-1) - \mathbf{1}_{\{\theta_j = B\}} \left(\bar{u}LT + \sum_{l=1}^{L-1} \left(\min_{\alpha_j} Tu_i(\alpha_i(x), \alpha_j) - Tu_i(\alpha_i(x), \alpha_j(l)) \right) \right) \\ & + \mathbf{1}_{\{\theta_j = B\}} \min_{\alpha_j} Tu_i(\alpha_i(x), \alpha_j) + \mathbf{1}_{\{\theta_j = G\}} Tu_i(\alpha(x)). \end{aligned} \tag{25}$$

Note that now we take the summation of $\min_{\alpha_j} Tu_i(\alpha_i(x), \alpha_j) - Tu_i(\alpha_i(x), \alpha_j(l))$ until round $L-1$ since $\min_{\alpha_j} Tu_i(\alpha_i(x), \alpha_j) - Tu_i(\alpha_i(x), \alpha_j(L))$ is included in $\min_{\alpha_j} Tu_i(\alpha_i(x), \alpha_j)$, the value in round L.

If we neglect the effect of player i's strategy on θ_j, then both C_i and D_i would be optimal by (14) and (22). Hence, if we adjust the reward function so that we can neglect this effect we are done.

First, we adjust the reward function so that we can actually neglect the effect of player i's strategy on θ_j by adding

$$\max_{\tilde{h}_i^{L-1}} \mathbb{E}\left[-\mathbf{1}_{\{\theta_j = B\}} \Delta_L \mid \hat{h}_i^{L-1}, \tilde{h}_i^{L-1} \right] - \mathbb{E}\left[-\mathbf{1}_{\{\theta_j = B\}} \Delta_L \mid \hat{h}_i^{L-1}, \hat{h}_i^{L-1} \right] \tag{26}$$

where Δ_L is the reduction of the continuation payoffs when $\theta_j = B$ happens

$$\begin{aligned} 0 \quad \leq \quad & \Delta_L \equiv \bar{u}LT + \sum_{l=1}^{L-1} \left(\min_{\alpha_j} Tu_i(\alpha_i(x), \alpha_j) - Tu_i(\alpha_i(x), \alpha_j(l)) \right) \\ & + Tu_i(\alpha(x)) - \min_{\alpha_j} Tu_i(\alpha_i(x), \alpha_j) \end{aligned}$$

and $\mathbb{E}\left[-\mathbf{1}_{\{\theta_j = B\}} \Delta_L \mid \hat{h}_i^{L-1}, \tilde{h}_i^{L-1} \right]$ is the expected value of this reduction conditional on that player i observed $\hat{h}_i^{L-1}$ and reports $\tilde{h}_i^{L-1}$.

Since Condition 2 for $\theta_j = B$ is solely determined by player j's mixture and independent of player i's strategy in round $L-1$, the effect of player i's strategy on θ_j is solely through

$\hat{h}_i^{L-1}$. Hence, if the truthtelling is optimal, then (26) cancels out the effect of player i's strategy on θ_j.

Second, player j incentivizes player i to tell the truth at the end of round $L-1$ even after taking (26) into account. Player j punishes player i based on $\hat{h}_i^{L-1}$ by

$$-\sum_{t:\text{ round } L-1} T^{-2}\left\|\mathbf{1}_{a_{j,t},y_{j,t}}-\mathbb{E}[\mathbf{1}_{a_{j,t},y_{j,t}}\mid a_{i,t},y_{i,t},\alpha_j(L-1)]\right\|^2. \tag{27}$$

Finally, the expected value of (27) before observing $y_{i,t}$ but after taking $a_{i,t}$ is different for different $a_{i,t}$'s. To cancel out this difference, player j adds $\pi_i^{\text{report}}[\alpha_{j,t}](y_{j,t})$ to the reward with $\pi_i^{\text{report}}[\alpha_{j,t}](y_{j,t})$ satisfying

$$\mathbb{E}\left[-T^{-2}\left\|\mathbf{1}_{a_{j,t},y_{j,t}}-\mathbb{E}[\mathbf{1}_{a_{j,t},y_{j,t}}\mid a_{i,t},y_{i,t},\alpha_{j,t}]\right\|^2+\pi_i^{\text{report}}[\alpha_{j,t}](y_{j,t})\mid a_{i,t},\alpha_{j,t}\right]=0 \tag{28}$$

for all $(a_{i,t},\alpha_{j,t})$. Assumption 3 guarantees the existence of such reward $\pi_i^{\text{report}}[\alpha_{j,t}](y_{j,t})$ and $\left|\pi_i^{\text{report}}[\alpha_{j,t}](y_{j,t})\right|=\Theta(T^{-2})$.

Now the reward is (23) plus (26), (27) and (28). Given (26), (27) and (28), if the truthtelling is optimal, then the effect of player i's strategy on θ_j is canceled out. Hence, we are left to show player i's incentive to tell the truth in the end of round $L-1$ and the entire reward is non-positive.

First, we show that (26) is small for any $\hat{h}_i^{L-1}$. Classify $\hat{h}_i^{L-1}$ into following four classes:

- if $\hat{h}_i^{L-1}$ does not satisfy Conditions 1 and 2 for the belief of $X_j(L-1)$, which minimizes the probability of $\theta_j=B$. Hence, (26) is zero;
- if $\hat{h}_i^{L-1}$ satisfies Conditions 1 and 2 for the belief of $X_j(L-1)$, then consider the following three cases for player i's signal frequency in periods when player i took C_i according to $\hat{h}_i^{L-1}$:
  - if the frequency was very close to the ex ante distribution given $(C_i,\alpha_j(x))$, then there are following two cases about player j's action plan $\alpha_j(L-1)$:

* $*$ if player j took $\alpha_j(L-1) \neq \alpha_j(x)$, then Condition 2 for $\theta_j = B$ is satisfied and regardless of $\hat{h}_i^{L-1}$, $\theta_j = B$ is determined;
* $*$ if player j took $\alpha_j(L-1) = \alpha_j(x)$, then player i's conditional expectation of $X_j(L-1)$ given $(C_i, \alpha_j(x))$ was also close to the ex ante mean of $X_j(L-1)$ under $(C_i, \alpha_j(x))$, which is close to zero for sufficiently small ρ.[12] Hence, by the large deviation theory, since the length of the round is T, player i put a belief no more than $\exp(-\Theta(T))$ on the event that $X_j(L-1) > \frac{\bar{u}}{L}T$;[13]

– if the frequency was skewed toward g_i, that is, if player i observed g_i more often than the ex ante frequency under $(C_i, \alpha_j(x))$, then by the large deviation theory, then player i put a belief no less than $1 - \exp(-\Theta(T))$ on the event that player j took $\bar{\alpha}_j(x)$ and $\theta_j = B$ is determined regardless of $\hat{h}_i^{L-1}$ since the frequency was skewed toward $q_i(C_i, C_j)$;

– if the frequency was skewed toward b_i, that is, if player i observed b_i more often than the ex ante frequency under $(C_i, \alpha_j(x))$, then by the same argument,[14] player i put a belief no less than $1 - \exp(-\Theta(T))$ on the event that player j took $\underline{\alpha}_j(x)$ and $\theta_j = B$ is determined regardless of $\hat{h}_i^{L-1}$;

Hence, in total, (26) is no less than $\exp(-\Theta(T))$.

Second, given (26) is small for all $\hat{h}_i^{L-1}$, the punishment (27) is sufficiently large to incentivize player i to tell the truth about h_i^{L-1}. Whenever player i's history $(a_{i,t}, y_{i,t})$ gives player i different belief about player j's history $(a_{j,t}, y_{j,t})$, (27) punishes player i by $\Theta(T^{-2})$

[12](15) and (17) implies that the ex ante mean of $X_j(L-1)$ under $(\alpha_i(x), \alpha_j(x))$ is zero. Since $\alpha_i(x)$ prescribes C_i with probability $1-2\rho$, for sufficiently small ρ, the ex ante mean of $X_j(L-1)$ under $(C_i, \alpha_j(x))$ is also close to zero.

[13]For a variable X_T which depends on T, we say $X_T = \exp(-\Theta(T^k))$ if and only if there exist $k_1, k_2 > 0$ such that $\exp(-k_1 T^k) \leq X_T \leq \exp(-k_2 T^k)$ for sufficiently large T.

[14]Since we assume $|Y_i| = |A_i|$, Assumption 3 implies that whenever player i's signal frequency is not close to the ex ante distribution under $(C_i, \alpha_j(x))$, it should be skewed toward either $q_i(C_i, C_j)$ or $q_i(C_i, D_j)$.

If $|Y_i| > |A_i|$, then it could be the case that although player i's signal frequency is not close to the ex ante distribution under $\alpha_j(x)$, it is skewed toward neither $q_i(C_i, C_j)$ nor $q_i(C_i, D_j)$. See Section 6.5.1 for how we take care of this case.

if player i tells a lie. Since (26) is bounded by $\exp(-\Theta(T))$, this punishment is sufficiently large. Note that (28) is sunk by the time when player i sends $\hat{h}_i^{L-1}$.

Therefore, we have verified player i's incentive to tell the truth in the end of round $L-1$. In addition, since (23) is non-positive and (26), (27) and (28) are all small (at least of order $\Theta(-T^2)$), by subtracting a small constant if necessary, the feasibility is satisfied.

Recursively, by backward induction, we can show that the equilibrium action plan is optimal for each round. Since the players take $\alpha(x)$ with a high probability for sufficiently small η and Conditions 1 and 2 for $\theta_j = B$ only happens with a small probability, efficiency is preserved.

6 Equilibrium Strategy

Now we define the equilibrium strategy for T_P-period finitely repeated game for the general N-player game. Remember that we see the finitely repeated game as L repetitions of T-period review rounds, where $L \in \mathbb{N}$ will be pinned down in Section 7 and $T = (1-\delta)^{-\frac{1}{2}}$ as in (11). Let $T(l)$ with $|T(l)| = T$ be the set of T periods in round l and $h_i^l = (a_{i,t}, y_{i,t})_{t \in T(l)}$ be player i's history in round l. Note that $T_P = LT$.

In Section 6.1, we define statistics useful for the equilibrium construction. In Section 6.2, we define the state variables that will be used to define the action plans and rewards. Given the states, Section 6.3 defines the action plan $\sigma_i(x_i)$ and Section 6.4 defines the reward function $\pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta)$. Section 6.5 finishes defining the strategy by determining the transition of the states defined in Section 6.2.

6.1 Statistics

We define one number $\bar{u} > 0$ and two statistics (functions of signals) useful for the equilibrium construction: $\pi_i[\alpha](y_{i-1})$ and $\pi_i^\delta(t, \alpha_{-i,t}, y_{i-1,t})$.

First, $\pi_i[\alpha](y_{i-1})$ is the point corresponding to $\pi_i[\alpha](y_j)$ briefly explained in Section 5.2. Specifically, for each $\alpha \in \Delta(A)$, we want to create a statistics (point) $\pi_i[\alpha](y_{i-1})$ such that

$\pi_i[\alpha] : Y_{i-1} \to (-\bar{u}, \bar{u})$ cancels out the differences in the instantaneous utilities for different a_i's:

$$u_i(a_i, \alpha_{-i}) + \mathbb{E}\left[\pi_i[\alpha](y_{i-1}) \mid a_i, \alpha_{-i}\right] \tag{29}$$

is independent of $a_i \in A_i$, as in (14). Further, we want to make sure that if $\alpha_{-i} = \alpha_{-i}(x)$, then the expected sum of the instantaneous utility and $\pi_i[\alpha(x)](y_{i-1})$ satisfies

$$u_i(a_i, \alpha_{-i}(x)) + \mathbb{E}\left[\pi_i[\alpha(x)](y_{i-1}) \mid a_i, \alpha_{-i}(x)\right] = u_i(\alpha(x)) \tag{30}$$

for all $a_i \in A_i$. In other words, we take

$$\mathbb{E}\left[\pi_i[\alpha(x)](y_{i-1}) \mid \alpha(x)\right] = 0. \tag{31}$$

This corresponds to (15) in Section 5.2. Taking $\bar{u}$ sufficiently large, we want to make sure that

$$2 \max_{i,a} |u_i(a)| + 2 \max_{i,\alpha} |\pi_i[\alpha](y_{i-1})| < \bar{u}. \tag{32}$$

We will prove that the maximum is well-defined in Section 10.2.

Second, we want to construct the point $\pi_i^\delta : \mathbb{N} \times \Delta(A_{-i}) \times Y_{i-1} \to \mathbb{R}$ such that the effect of discounting is canceled out:

$$\delta^{t-1} u_i\left(a_{i,t}, \alpha_{-i,t}\right) + \mathbb{E}\left[\pi_i^\delta\left(t, \alpha_{-i,t}, y_{i-1,t}\right) \mid a_{i,t}, \alpha_{-i,t}\right] = u_i\left(a_{i,t}, \alpha_{-i,t}\right) \tag{33}$$

for all $a_{i,t} \in A_i$, $a_{-i,t} \in \Delta(A_{-i})$ and $t \in \{1, ..., T_P\}$ and

$$\lim_{\delta \to 1} \frac{1-\delta}{1-\delta^{T_P}} \sum_{t=1}^{T_P} \sup_{\alpha_{-i,t}, y_{i-1,t}} \left|\pi_i^\delta\left(t, \alpha_{-i,t}, y_{i-1,t}\right)\right| = 0 \tag{34}$$

for all L with $T_P = LT$ and $T = (1-\delta)^{-\frac{1}{2}}$.

Since Assumption 3 implies that player $i-1$ can statistically infer player i's action given α_{-i}, the existence of such $\pi_i[\alpha](y_{i-1})$ and $\pi_i^\delta(t, \alpha_{-i,t}, y_{i-1,t})$ is guaranteed.

Lemma 2 *If Assumption 3 is satisfied, then there exists $\bar{u} > 0$ such that,*

1. *for each $i \in I$, $\alpha \in \Delta(A)$ and $\{\alpha(x)\}_{x \in \{G,B\}^N}$, there exists $\pi_i[\alpha] : Y_{i-1} \to (-\bar{u}, \bar{u})$ with (29), (30) and (32);*

2. *for each $i \in I$, there exists $\pi_i^\delta : \mathbb{N} \times \Delta(A_{-i}) \times Y_{i-1} \to \mathbb{R}$ such that, for all L with $T_P = LT$ and $T = (1-\delta)^{-\frac{1}{2}}$, (33) and (34) are satisfied.*

Proof. See Section 10.2. ■

In addition to these two statistics, we consider the following variables in round l. Let $f_i(a_i, y_i)$ be the frequency of an action-signal pair (a_i, y_i) in $T(l)$. Given $f_i(a_i, y_i)$,

$$f_i(a_i, Y_i) \equiv (f_i(a_i, y_i))_{y_i \in Y_i},$$
$$f_i(a_i) \equiv \textstyle\sum_{y_i} f_i(a_i, y_i),$$
$$f_i(Y_i \mid a_i) \equiv \frac{f_i(a_i, Y_i)}{f_i(a_i)}$$

are the vector of player i's signal frequency during the periods when player i takes a_i, the frequency of actions, and the vector of player i's conditional signal frequency given a_i, respectively.

Suppose players i takes $a_i \in A_i$ for more than $T/2$ periods in $T(l)$ and players $-(i,j)$ take $\alpha_{-(i,j)} \in \Delta(A_{-(i,j)})$. Then, by the law of large numbers, regardless of player j's action, $f_i(Y_i \mid a_i)$ is close to

$$\boldsymbol{Q}_i^j(a_i, \alpha_{-(i,j)}) \equiv \mathrm{aff}(\{q_i(a_i, a_j, \alpha_{-(i,j)})\}_{a_j \in A_j}) \cap \mathbb{R}_+^{|Y_i|}.$$

We can represent $\boldsymbol{Q}_i^j(a_i, \alpha_{-(i,j)})$ by the matrix expression

$$\boldsymbol{Q}_i^j(a_i, \alpha_{-(i,j)}) \equiv \{\mathbf{y}_i \in \mathbb{R}_+^{|Y_i|} : Q_i^j(a_i, \alpha_{-(i,j)})\mathbf{y}_i = \mathbf{q}_i^j(a_i, \alpha_{-(i,j)})\}.$$

Since all the signal frequencies should be on the simplex over Y_j, by affine transformation, we can assume that each element of $Q_i^j(a_i, \alpha_{-(i,j)})$ and $\mathbf{q}_i^j(a_i, \alpha_{-(i,j)})$ is in $(0,1)$.

Lemma 3 *For any $i, j \in I$ with $i \neq j$, $a_i \in A_i$ and $\alpha_{-(i,j)} \in \Delta(A_{-(i,j)})$, we can take $Q_i^j(a_i, \alpha_{-(i,j)})$ and $\mathbf{q}_i^j(a_i, \alpha_{-(i,j)})$ such that all the elements are in $(0, 1)$.*

Proof. See Section 10.3. ■

In general, for a random variable $z \in Z$, $\mathbf{1}_z \in \{0, 1\}^{|Z|}$ is a $|Z|$-dimensional random vector such that if z is realized, then the element corresponding to z is 1 and the others are 0. After taking $a_{i,t} = a_i$ and observing $y_{i,t}$, player i calculates $Q_i^j(a_i, \alpha_{-(i,j)})\mathbf{1}_{y_{i,t}}$. With D being the dimension of $Q_i^j(a_i, \alpha_{-(i,j)})\mathbf{1}_{y_{i,t}}$, player i draws D random variables from the uniform $[0, 1]$ independently. If the dth realization of these random variables is no less than the dth element of $Q_i^j(a_i, \alpha_{-(i,j)})\mathbf{1}_{y_{i,t}}$, we define the dth element of $\vec{Q}_i^j(a_i, \alpha_{-(i,j)})$ equal to 1. Otherwise, the dth element of $\vec{Q}_i^j(a_i, \alpha_{-(i,j)})$ is 0. By definition, the distribution of $\vec{Q}_i^j(a_i, \alpha_{-(i,j)})\mathbf{1}_{y_{i,t}}$ is independent of player j's action as long as players $-(i, j)$ take $\alpha_{-(i,j)}$. Let $f_i(\vec{Q}_i^j(a_i, \alpha_{-(i,j)}) \mid a_i)$ be the conditional frequency of $\vec{Q}_i^j(a_i, \alpha_{-(i,j)})$ in the periods when player i takes a_i in $T(l)$.

When we say $f_i(Y_i \mid a_i)$ is close to $\boldsymbol{Q}_i^j(a_i, \alpha_{-(i,j)})$, it means both of the following two conditions are satisfied:

-
$$\left\| Q_i^j(a_i, \alpha_{-(i,j)}) f_i(Y_i \mid a_i) - f_i(\vec{Q}_i^j(a_i, \alpha_{-(i,j)}) \mid a_i) \right\| < \frac{1}{K_1}; \tag{35}$$

-
$$\left\| f_i(\vec{Q}_i^j(a_i, \alpha_{-(i,j)}) \mid a_i) - \mathbf{q}_i^j(a_i, \alpha_{-(i,j)}) \right\| < \frac{1}{K_1}. \tag{36}$$

The following lemma guarantees that, for any $\varepsilon > 0$, for sufficiently large K_1, (35) and (36) imply

$$d(f_i(Y_i \mid a_i), \boldsymbol{Q}_i^j(a_i, \alpha_{-(i,j)})) < \varepsilon. \tag{37}$$

In this paper, we use Euclidean norm and Hausdorff metric.

Lemma 4 *For any $\varepsilon > 0$, there exists $\bar{K}_1$ such that, for all $K_1 > \bar{K}_1$, (35) and (36) imply (37).*

Proof. See Section 10.4. ■

Further, we want to make sure that the probability that $f_i(Y_i \mid a_i)$ is close to $\boldsymbol{Q}_i^j(a_i, \alpha_{-(i,j)})$ is independent of player j's strategy given that player i takes a_i for more than $T/2$ periods in $T(l)$ and that players $-(i,j)$ take $\alpha_{-(i,j)}$. (36) is independent since the distribution of $\vec{Q}_i^j(a_i, \alpha_{-(i,j)})\mathbf{1}_{y_{i,t}}$ is independent. For all the histories where player i takes a_i for more than $T/2$ periods in $T(l)$, conditional on player i's history in $T(l)$, (35) is satisfied with probability $1-\exp(-\Theta(T))$. Let $\bar{p} = 1-\exp(-\Theta(T))$ be the minimum of such a probability with respect to player i's histories satisfying the condition that player i takes a_i for more than $T/2$ periods in $T(l)$. If (35) happens with a larger probability p than $\bar{p}$ after some history, then player i draws a random variable from the uniform $[0,1]$. If this realization is no less than $p - \bar{p}$, then player i behaves as if (35) were not satisfied.

In total, when we say $f_i(Y_i \mid a_i)$ is close to $\boldsymbol{Q}_i^j(a_i, \alpha_{-(i,j)})$, then (36) and (35) are satisfied, taking this adjustment into account. Then, the probability that $f_i(Y_i \mid a_i)$ is close to $\boldsymbol{Q}_i^j(a_i, \alpha_{-(i,j)})$ is independent of player j's strategy.

6.2 States x_i, $\lambda(l)$ and θ_{i-1}

Now, we define three state variables useful to define the equilibrium strategy: x_i, $\lambda(l)$ and θ_{i-1}. The state $x_i \in \{G, B\}$ is determined at the beginning of the finitely repeated game and fixed. Since x is communicated by cheap talk at the beginning of the finitely repeated game truthfully, x becomes common knowledge. Hence, we use x_{-i} for the definition of player i's strategy.

As seen in Section 5.2, $\lambda(l) \in \{\emptyset\} \cup I \cup$punish is the state for round l. Intuitively, $\lambda(l) = i$ means that player $i-1$'s score about player i has been excessively high, and so if $\lambda(l) = i$, then player $i-1$'s reward on player i is constant and player i takes a static best response to players $-i$. In Section 5.2, we focus on the case with $x_{i-1} = G$ for all $i \in I$. If player $i-1$ has $x_{i-1} = B$, then instead of $\lambda(l) = i$, the players have $\lambda(l) =$ punish after player $i-1$'s score about player i has been excessively "low" and if $\lambda(l) =$ punish, all the players take static best response to each other, that is, the static Nash equilibrium α_i^*.

In addition, as seen in Section 5.2, after some events, player $i-1$ adds or subtracts a large number from the reward function. $\theta_{i-1} = B$ implies such an event happens while $\theta_{i-1} = G$ implies such an event does not happen.

6.3 Player i's Action Plan $\sigma_i(x_i)$

Now, we define player i's action plan $\sigma_i(x_i)$ given states x and $\lambda(l)$. See Section 6.4 for the definition of the reward function $\pi_i(x_{i-1}, h_{i-1}^{T_P} : \delta)$ and Section 6.5 for the transition of the states.

At the beginning of the finitely repeated game, player i tells the truth about x_i. If player i told a lie and her state is $\hat{x}_i$ when it is x_i, define $\sigma_i(x_i) = \sigma_i(\hat{x}_i)$, that is, player i's continuation action plan is as if her true state is $\hat{x}_i$.

In round l, player i with $\sigma_i(x_i)$ takes an *i.i.d.* action plan $\alpha_i(l)$, depending on $\lambda(l)$. To define the strategy, let $C_i = \{\mathbf{t}_i \in \mathbb{R}^{|A_i|} : \|\mathbf{t}_i\| = 1\}$ be the set of $|A_i|$-dimensional vectors with length 1 and $\bar{C}_i \subset C_i$ with $|\bar{C}_i| < \infty$ be the finite subset of C_i. See Lemma 5 for the formal definition of $\bar{C}_i$. Given $\bar{C}_i$ and $\rho > 0$ to be determined in Section 7, $\alpha_i(l)$ is determined as follows:

1. if $\lambda(l) \neq i$ and $\lambda(l) \neq$ punish, then with $\eta > 0$,
   (a) with probability $1 - \eta$, $\alpha_i(l) = \alpha_i(x)$;
   (b) with probability η, player i randomly draws $\mathbf{t}_i$ from $\bar{C}_i$ such that $\Pr(\mathbf{t}_i) = 1/|\bar{C}_i|$. If $\mathbf{t}_i$ is drawn, then player i takes
$$\alpha_i(x, \mathbf{t}_i) \equiv \left(1 - \sum\nolimits_{a_i \neq a_i(x)} (2\rho + \rho t_i(a_i))\right) a_i(x) + \left(\sum\nolimits_{a_i \neq a_i(x)} (2\rho + \rho t_i(a_i))\, a_i\right);$$
2. if $\lambda(l) = i$, then player i takes a static best response to $a_{-i}(x)$;
3. if $\lambda(l) \neq$ punish, then player i takes a static Nash equilibrium action α_i^*;

At the end of period lT (the last period of round l), each player i truthfully sends h_i^l simultaneously by cheap talk.

6.4 Reward Function

In this subsection, we explain player $i-1$'s reward function on player i, $\pi_i(x_{i-1}, h_{i-1}^{T_P} : \delta)$.

Reward Function As in (22), we call

$$X_{i-1}(l) \equiv \sum_{t \in T(l)} \pi_i[\alpha_i(x), \alpha_{-i}(l)](y_{i-1,t})$$

"player $i-1$'s score on player i."

The reward $\pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta)$ is written as

$$\begin{aligned}
\pi_i(x_{i-1}, h_{i-1}^{T_P+1} \quad : \quad \delta) = & \sum_{l=1}^{L} \sum_{t \in T(l)} \pi_i^{\delta}(t, \alpha_{-i,t}, y_{i-1,t}) \\
& + \text{sign}(x_{i-1}) \mathbf{1}_{\{\theta_{i-1}=B\}} \left\{ 3\bar{u}LT + \sum_{l=1}^{L} (\bar{\pi}_i(x, \alpha_{-i}(l), l) + X_{i-1}(l)) \right\} \\
& + \text{sign}(x_{i-1}) 2\bar{u}T + \mathbf{1}_{\{\theta_{i-1}=G\}} \left\{ \bar{\pi}_i(x, \lambda(l), l) + \sum_{l: \left\{ \begin{array}{c} \lambda(l) \neq i, \\ \lambda(l) \neq \text{punish} \end{array} \right.} X_{i-1}(l) \right\} \\
& + \sum_{l: \lambda(l) = \emptyset} \sum_{t \in T(l)} \left(-T^{-2} \left\| \mathbf{1}_{a_{-i,t}, y_{-i,t}} - \mathbb{E}[\mathbf{1}_{a_{-i,t}, y_{-i,t}} \mid \hat{a}_{i,t}, \hat{y}_{i,t}, \alpha_{-i,t}] \right\|^2 + \pi_i^{\text{report}}[\alpha_{-i,t}](y_{i-1,t}) \right) \\
& + \text{sign}(x_{i-1}) \sum_{l=1}^{L-1} v_i(\{\lambda(\tilde{l})\}_{\tilde{l}=1}^{l-1}, \{h_{-i}^{\tilde{l}}\}_{\tilde{l}=1}^{l}, \hat{h}_i^l) \qquad (38)
\end{aligned}$$

where $\text{sign}(x_{i-1}) = -1$ if $x_{i-1} = G$ and $\text{sign}(x_{i-1}) = 1$ if $x_{i-1} = B$.

Let us comment on the reward function line by line. The first line is to cancel out the effect of discounting. Hence, from now, we can assume $\delta = 1$.

The role of the second line is the same as $\bar{u}LT$ in (23). There are two possible events to induce $\theta_{i-1} = B$:

- $x_{i-1} = G$ and player $i-1$'s score is excessively high but player i announces that she believes it is regular, which corresponds to Condition 1 for $\theta_j = B$ in Section 5;

- there is another player $j \in -i$ such that $x_{j-1} = G$ and player j announces that she believes player $j-1$'s score is excessively high, which corresponds to Condition 2 for $\theta_j = B$ in Section 5;

In such a case, player $i-1$ subtracts or adds large number $3\bar{u}LT$ to satisfy the feasibility, depending on her state x_{i-1}. In addition, player $i-1$ uses the score to incentivize player i in the continuation play. Further, we have $\bar{\pi}_i(x, \alpha_{-i}(l), l)$. As will be seen in Section 8, once $\theta_{i-1} = B$ is induced, then player $i-1$ will make player i indifferent between any sequence $\{\lambda(l)\}_{l=1}^L$. $\bar{\pi}_i(x, \alpha_{-i}(l), l)$ cancels out the possible differences in player i's payoffs for different $\alpha_{-i}(l)$'s, which are determined by $\{\lambda(l)\}_{l=1}^L$.

The role of the third line is to incentivize player i by the score. As in $-2\bar{u}T$ in (23), $\text{sign}(x_{i-1})2\bar{u}T$ makes it rare for the score to be excessively high or low. Player i is incentivized to take the equilibrium action by the score. Similarly to $\bar{\pi}_i(x, \alpha_{-i}(l), l)$, $\bar{\pi}_i(x, \lambda(l), l)$ adjusts player i's incentive about the transition of $\{\lambda(l)\}_{l=1}^L$.

The fourth line incentivizes player i to tell the truth about the history at the end of each review round, as (27) in Section 5. As in (28), $\pi_i^{\text{report}}[\alpha_{-i,t}](y_{i-1,t})$ cancels out the differences in ex ante payoffs for different actions in terms of (27).

The last line deals with the fact that different histories of player i give different payoffs in the continuation play since the message of the histories affect the transition of $\{\lambda(l)\}_{l=1}^L$. As (26), we cancel out the effect of this differences on player i's incentives.

Finally, note that player $i-1$ uses the information owned by players $-(i-1,i)$ to calculate the reward (for example, $a_{-i,t}, y_{-i,t}$ in the fourth line). Player i yields this information from the messages by players $-(i-1,i)$. This does not affect the incentives of players $-(i-1,i)$ since this information is used only for the reward to player i.

6.5 Transition of the States

In this subsection, we explain the transition of the players' states. Since x is fixed in the phase, we consider $\lambda(l)$ and θ_{i-1}.

6.5.1 Transition of $\lambda(l+1) \in \{\emptyset\} \cup I \cup \text{punish}$

As mentioned in Section 5.2, $\lambda(l+1) = i$ implies that player i believes that player $i-1$'s score has been high. In addition, $\lambda(l+1) = \text{punish}$ implies that some player $i-1$ had an excessively low score on player i and triggered the punishment. Since player i does not have an incentive to tell that she believes that player $i-1$ has an excessively low score and that players $-i$ now need to punish player i, player $i-1$ announces the punishment.

The initial condition is $\lambda(1) = \emptyset$. Inductively, given $\lambda(l) \in \{\emptyset\} \cup I \cup \text{punish}$, $\lambda(l+1)$ is determined as follows: if $\lambda(l) \neq \emptyset$, then $\lambda(l+1) = \lambda(l)$. That is, once $\lambda(l) \neq \emptyset$ happens, it lasts until the end of the finitely repeated game. If $\lambda(l) = \emptyset$, then $\lambda(l+1) \in \{\emptyset\} \cup I \cup \text{punish}$ is determined as follows:

1. if there exists a unique i such that $x_{i-1} = G$ and "player i announces that player i believes that the score $X_{i-1}(l)$ is excessively high," then $\lambda(l+1) = i$;

2. otherwise, that is, if there is no i such that $x_{i-1} = G$ and "player i announces that player i believes that the score $X_{i-1}(l)$ is excessively high," then

   (a) if there exists i with $x_{i-1} = B$ such that "player $i-1$ triggers the punishment," then $\lambda(l) = \text{punish}$;

   (b) otherwise, $\lambda(l+1) = \emptyset$.

Let us call these conditions Conditions 1, 2-(a) and 2-(b) for $\lambda(l+1)$.

Now, we define when "player i announces that player i believes that the score $X_{i-1}(l)$ is excessively high" and when "player $i-1$ triggers the punishment," which are determined by player i's announcement $\hat{h}_i^l$ at the end of round l and player $i-1$'s announcement $\hat{h}_{i-1}^l$ at the end of round l, respectively.

When Player i Announces that Player i Believes that The Score $X_{i-1}(l)$ is Excessively High Since $\lambda(l+1)$ does not change after $\lambda(l) \neq \emptyset$, we concentrate on the case with $\lambda(l) = \emptyset$. This implies that, from Section 6.3, each player $j \in I$ takes

1. with probability $1-\eta$, $\alpha_j(l) = \alpha_j(x)$;

2. with probability η, player j randomly draws $\mathbf{t}_j$ from $\bar{C}_j$ such that $\Pr(\mathbf{t}_j) = 1/\left|\bar{C}_j\right|$. If $\mathbf{t}_j$ is drawn, then player j takes

$$\alpha_j(x, \mathbf{t}_j) \equiv \left(1 - \sum\nolimits_{a_j \neq a_j(x)} (2\rho + \rho t_j(a_j))\right) a_j(x) + \left(\sum\nolimits_{a_j \neq a_j(x)} (2\rho + \rho t_j(a_j))\, a_j\right).$$

If player i's message $\hat{h}_i^l$ satisfies at least one of the following two conditions, then we say that player i announces that $X_{i-1}(l)$ is excessively high. The first condition is that, as in Section 5, player i does not take $\alpha_i(l) \neq \alpha_i(x)$ or player i's action frequency is not close to the ex ante distribution $\alpha_i(x)$.

The second condition is that, for all $j \in -i$, player i's signal frequency while player i takes $a_i(x)$ in $T(l)$ is not close to the affine full of player i's signal frequencies with respect to player j's action.

Suppose that player i tells the truth and that neither of the two conditions is satisfied. Then, as mentioned in footnote **??**, conditional on that the first condition is not satisfied (and so player i takes $a_i(x)$ frequently), as long as the second condition fails (and so player i's signal frequency is close to the affine hull), player i's signal frequency is either close to the ex ante distribution under $\alpha_{-i}(x)$ (and so player i believes that the score is not excessively high), or implies that player i believes that $\alpha_j(l) \neq \alpha_j(x)$. In both cases, player i believes that there is no need to induce $\lambda(l+1) = i$.

In total, player i's message $\hat{h}_i^l$ satisfies at least one of the following two conditions, then we say that player i announces that $X_{i-1}(l)$ is excessively high:

1. player i takes $\alpha_i(l) \neq \alpha_i(x)$ or player i's action frequency is not close to the ex ante distribution $\alpha_i(x)$;

2. for all $j \in -i$, $f_i(Y_i \mid a_i(x))$ is not close to $\boldsymbol{Q}_i^j(a_i(x), \alpha_{-(i,j)}(x))$. See Section 6.1 for the definition. Note that, conditional on Condition 1 not being satisfied, together with with $\rho, \varepsilon < 1/4$, player i takes $a_i(x)$ for more than $T/2$ times in $T(l)$.

Let us call these two conditions Conditions 1 and 2 for the belief of $X_{i-1}(l)$.

When Player $i-1$ Triggers the Punishment Intuitively, player $i-1$ triggers the punishment if and only if player $i-1$ believes that player i takes $\alpha_i(l) \neq \alpha_i(x)$. Since $X_{i-1}(l)$ monitors player i, player $i-1$ triggers the punishment if $X_{i-1}(l)$ is small. Specifically, if player $i-1$'s message $\hat{h}_{i-1}^l$ satisfies all of the following three conditions, then we say that player $i-1$ triggers the punishment:

1. player $i-1$ takes $\alpha_{i-1}(l) = \alpha_{i-1}(x)$ and player $i-1$'s action frequency is close to the ex ante distribution $\alpha_{i-1}(x)$;

2. $f_{i-1}(Y_{i-1} \mid a_{i-1}(x))$ is close to $\boldsymbol{Q}_{i-1}^i(a_{i-1}(x), \alpha_{-(i-1,i)}(x))$;

3. $X_{i-1}(l) < -\frac{\bar{u}}{L}$.

Symmetrically to the discussion in Section 5, whenever player $i-1$'s history satisfies these conditions, player $i-1$ believes that player i take $\alpha_i(l) \neq \alpha_i(x)$, given players $-(i-1,i)$ take $\alpha_{-(i-1,i)}(x)$. Let us call these three conditions Conditions 1, 2 and 3 for player $i-1$'s punishment.

6.5.2 Transition of $\theta_{i-1} \in \{G, B\}$

As seen in Section 6.4, $\theta_{i-1} = B$ implies that once $\theta_{i-1} = B$ is induced, then we will make player i indifferent between any sequence $\{\lambda(l)\}_{l=1}^L$.

Here, we list the events after which player $i-1$ has $\theta_{i-1} = B$. If none of these events happens, then $\theta_{i-1} = G$:

1. there is round l where (i) $\lambda(l) = \emptyset$, (ii) $x_{i-1} = G$ and the true score was excessively high in round l, that is, $X_{i-1}(l) > \frac{\bar{u}}{L}T$, and (iii) player i announced that player i believes

that player $i-1$'s score was regular, that is, $\hat{h}_i^l$ does not satisfy neither Condition 1 nor 2 for the belief of $X_{i-1}(l)$. As in Section 5, this means player i made a mistake in the announcement;

2. there is round l where (i) $\lambda(l) = \emptyset$ or i and (ii) there exists player $j \in -i$ such that at least one of the following conditions is satisfied:

   (a) who took $\alpha_j(l) \neq \alpha_j(x)$ or the actual frequency of player j's action was not close to $\alpha_j(x)$ or

   (b) $f_j(Y_j \mid a_j(x))$ is not close to $\boldsymbol{Q}_j^i(a_j(x), \alpha_{-(i,j)}(x))$.

Let us call these conditions Conditions 1, 2-(a) and 2-(b) for θ_{i-1}. Since players $-i$ tell the truth,[15] from Section 6.5.1, whenever player $j \in -i$ announces that player j believes that player $j-1$'s score is excessively high, then either 2-(a) or 2-(b) above is satisfied. Hence, $\theta_{i-1} = G$ implies that $\lambda(l) = \emptyset$, i or punish. In addition, Condition 2-(a) for θ_{i-1} implies that if $\theta_{i-1} = G$, then players $-i$ take $\alpha_{-i}(l) = \alpha_{-i}(x)$ if $\lambda(l) = \emptyset$ or i and $\alpha_{-i}(l) = \alpha_{-i}^*$ if $\lambda(l) =$ punish.

6.6 Player i's Belief about $\{X_j(l), \alpha_j(l)\}_{j \in -i}$

We consider player i's belief about scores and actions by the other players $\{X_j(l), \alpha_j(l)\}_{j \in -i}$. If player i's true history h_i^l satisfies neither Condition 1 nor 2 for the belief of $X_{i-1}(l)$, that is, if player i does not announce that player $i-1$'s score is excessively high if she tells the truth, then player i puts a belief no less than $1 - \exp(-\Theta(T))$ on the event that, given $\lambda(l) = \emptyset$, either $|X_n(l)| \leq \frac{\bar{u}}{L}T$ for all $n \in -i$ or there exists $j \in -i$ with $\alpha_j(l) \neq \alpha_j(x)$:

Lemma 5 *For all $\bar{u}$ and L, there exists $\bar{\eta}$ such that, for all $\eta < \bar{\eta}$, there exist $\bar{\rho}$ and $\bar{\varepsilon}$ such that, for all $\rho < \bar{\rho}$ and $\varepsilon < \bar{\varepsilon}$, there exists $\{\bar{C}_n\}_{n \in I}$ such that for any history h_i^l satisfying neither Condition 1 nor 2 for the belief of $X_{i-1}(l)$, conditional on $\lambda(l) = \emptyset$, player i after*

[15] We consider player i's incentive here.

h_i^l puts a belief no less than $1-\exp(-\Theta(T))$ on the event that either $|X_n(l)| \leq \frac{\bar{u}}{L}T$ for all $n \in -i$ or there exists $j \in -i$ with $\alpha_j(l) \neq \alpha_j(x)$.

Proof. See Section 10.5. ■

There are two implications about player i's incentive to tell the truth about $\hat{h}_i^l$. Consider the case where player i's truthtelling strategy $\hat{h}_i^l = h_i^l$ does not announce that player i believes that the score is excessively high.

The first implication is about player i's belief about player $i-1$'s score. There are following two cases on which player i puts a high belief:

1. $X_{i-1}(l) \leq \frac{\bar{u}}{L}T$ and player i does not need to announce it, or

2. there exists $j \in -i$ with $\alpha_j(l) \neq \alpha_j(x)$. From Condition 2-(a) for θ_{i-1}, $\theta_{i-1} = B$ is determined and player i is indifferent between any message $\hat{h}_i^l$ since $\{\bar{\pi}_i(x, \alpha_{-i}(l), l)\}_l$ in (38) makes player i indifferent between any sequence $\{\lambda(l)\}_{l=1}^L$.

In both cases, given the fourth line of (38) which gives a slight incentive to tell the truth, player i has the incentive to tell the truth.

The second implication is about the punishment. $\{X_n(l)\}_{n\in -i}$ affects whether player n triggers the punishment. Suppose h_i^l satisfies neither Condition 1 nor 2 for the belief of $X_{i-1}(l)$. Then, there are following two cases on which player i puts a high belief:

1. $X_n(l) \geq -\frac{\bar{u}}{L}T$ with $n \in -i$ and no player $n \in -i$ triggers the punishment;

2. there exists $j \in -i$ with $\alpha_j(l) \neq \alpha_j(x)$. Again, player i is indifferent between any sequence $\{\lambda(l)\}_{l=1}^L$.

To see why this is important, suppose player i would put a high belief on $X_n(l) < -\frac{\bar{u}}{L}T$ for some $n \in -i$ but $\alpha_{-i}(l) = \alpha_{-i}(x)$. If $x_{i-1} = G$, then since player i's equilibrium payoff is originally high, when the players switch from $\alpha(x)$ to to α^*, player i's payoff becomes lower.[16] Since Condition 2 for $\lambda(l+1)$ says that, if player i told a lie and announced that

[16] On the other hand, if $x_{i-1} = B$, then since player i's equilibrium payoff is originally low, player i is indifferent between $\lambda(l+1) = \emptyset$ and $\lambda(l+1) = \text{punish}$. See the proof of Proposition 1.

player i believes that the score is excessively high, then player i could induce $\lambda(l+1) = i$ and $\lambda(l+1) \neq$ punish while if player i tells the truth, then $\lambda(l+1) =$ punish.[17] Hence, player i would have an incentive to tell a lie to prevent $\lambda(l+1) =$ punish.

Next, we consider player $i-1$'s incentive to trigger the punishment whenever Conditions 1, 2 and 3 for player $i-1$'s punishment are satisfied.

Lemma 6 *For all $\bar{u}$ and L, there exists $\bar{\eta}$ such that, for all $\eta < \bar{\eta}$, there exist $\bar{\rho}$ and $\bar{\varepsilon}$ such that, for all $\rho < \bar{\rho}$ and $\varepsilon < \bar{\varepsilon}$, there exists $\{\bar{C}_n\}_{n \in I}$ such that for any history h_{i-1}^l satisfying Conditions 1, 2 and 3 for player $i-1$'s punishment, conditional on $\lambda(l) = \emptyset$, player $i-1$ after h_{i-1}^l puts a belief no less than $1 - \exp(-\Theta(T))$ on the event that there exists $j \in -(i-1)$ with $\alpha_j(l) \neq \alpha_j(x)$.*

Proof. See Section 10.6. ■

This lemma implies that, together with the conditions for θ_{i-2} (with $i-1$ replaced with $i-2$) implies that player $i-1$ right before sending $\hat{h}_{i-1}^l$ at the end of around l puts a belief no less than $1 - \exp(-\Theta(T))$ on the event that $\theta_{i-2} = B$, that is, player $i-1$ is indifferent between any $\lambda(l+1)$. Hence, given the fourth line of (38), player $i-1$ has an incentive to tell the truth about h_{i-1}^l and induce $\lambda(l+1) =$ punishment.

7 Variables

In this section, we show that all the variables can be taken so that all the requirements that have been imposed are satisfied: $\bar{u}$, L, η, ρ and ε. First, $\bar{u}$ is determined in Lemma 2.

Second, fix L so that

$$\max\left\{v_i^*, \max_{x:x_{i-1}=B} u_i(a(x))\right\} + 2\frac{\bar{u}}{L} < \underline{v}_i < \bar{v}_i < \min_{x:x_{i-1}=G} u_i\left(a\left(x\right)\right) - 2\frac{\bar{u}}{L}.$$

This is possible because of (7).

[17] Again, if there is player $j \in -i$ with $x_{j-1} = G$ and h_j^l satisfying either Condition 1 or 2 for the belief of $X_{j-1}(l)$, then $\theta_{i-1} = B$ and player i is indifferent between any message $\hat{h}_i^l$.

Third, given $\bar{u}$ and L, fix $\bar{\eta}$ so that (i) Lemma 5 holds, that (ii) for all $\eta < \bar{\eta}$, we have

$$
\begin{aligned}
& (1 - LN\eta) \max\left\{v_i^*, \max_{x:x_{i-1}=B} u_i(a(x))\right\} + 2\frac{\bar{u}}{L} + LN\eta 3\bar{u} \\
< \ & \underline{v}_i < \bar{v}_i < (1 - LN\eta) \min_{x:x_{i-1}=G} u_i\left(a\left(x\right)\right) - 2\frac{\bar{u}}{L} - LN\eta 3\bar{u}.
\end{aligned}
$$

Fourth, fix $\eta < \bar{\eta}$. Then, we can take $\bar{\rho}$ and $\bar{\varepsilon}$ so that Lemmas 5 and 6 hold. Take $\varepsilon < \bar{\varepsilon}$ and $\rho < \bar{\rho}$ so that

$$
\begin{aligned}
& (1 - LN\eta) \max\left\{v_i^*, \max_{x:x_{i-1}=B} u_i(a(x))\right\} + 2\frac{\bar{u}}{L} + LN\eta 3\bar{u} \\
< \ & \underline{v}_i < \bar{v}_i < (1 - LN\eta) \min_{x:x_{i-1}=G} u_i\left(\alpha\left(x\right)\right) - 2\frac{\bar{u}}{L} - LN\eta 3\bar{u}.
\end{aligned} \tag{39}
$$

Take $\{\bar{C}_n\}_{n\in I}$ so that Lemmas 5 and 6 hold.

Finally, take K_1 sufficiently large so that Lemma 4 holds.

Since $T_P = LT$ and $T = (1-\delta)^{-\frac{1}{2}}$, we have $\lim_{\delta\to 1} \delta^{T_P} = 1$. Therefore, discounting for the payoffs in the next review phase goes to zero.

8 Optimality of $\sigma_i(x_i)$

We have defined $\sigma_i(x_i)$ and π_i^{main} except for $\bar{\pi}_i(x, \lambda(l), l)$ and $\bar{\pi}_i(x, \alpha_{-i}(l), l)$. In this section, based on Lemmas 5 and 6, we show that if we properly define $\bar{\pi}_i(x, \lambda(l), l)$ and $\bar{\pi}_i(x, \alpha_{-i}(l), l)$, then $\sigma_i(x_i)$ and π_i^{main} satisfy (4), (5) and (6), which finishes the proof of 1. The intuition is the same as Section 5.

Proposition 1 *For sufficiently large δ, there exist $\bar{\pi}_i(x, \lambda(l), l)$ and $\bar{\pi}_i(x, \alpha_{-i}(l), l)$ such that (4), (5) and (6) are satisfied.*

Proof. See Section 10.7. ■

9 Concluding Remarks

We have shown the Nash-threat folk theorem with communication for a general game. There are two possible extensions from the current results.

The first is to dispense with cheap talk. That is, the players communicate via actions rather than via cheap talk. In such an extension, there is a following difficulty. When player $i-1$ tries to send the message that player $i-1$'s state is $x_{i-1} = B$, player i, whose value is low when $x_{i-1} = B$, wants to manipulate the signal distributions of players $-(i-1, i)$ by deviation in order to prevent players $-i$ from coordinating on the state unfavorable to player i. To deal with this problem, we need the pairwise full rank condition: no matter what player i does, player $j \in -(i-1, i)$ can statistically distinguish player $i-1$'s actions. Note that the current paper only assumes the individual full rank (Assumption 3).

The second is to consider the minimax-threat folk theorem. In such a case, since the action profile to minimax player i can be different from that to minimax player $j \neq i$, if something suspicious happens, the players need to figure out whether they should punish player i or player j, whereas in the Nash-threat folk theorem, the players can punish all of them by switching to the static Nash equilibrium. For this reason, we again need the pairwise full rank condition to statistically distinguish player i's deviations and player j's deviations.

See Sugaya (2012b), Sugaya (2012c) and Sugaya (2012d) for how to formally prove the minimax-threat folk theorem without cheap talk.

10 Appendix

10.1 Proof of Lemma 1

To see why this is enough for Theorem 1, define the strategy in the infinitely repeated game as follows: define

$$\begin{aligned} p(G, h_{i-1}^{T_P+1} &: \delta) \equiv 1 + \frac{1-\delta}{\delta^{T_P}} \frac{\pi_i(G, h_{i-1}^{T_P+1} : \delta)}{\bar{v}_i - \underline{v}_i}, \\ p(B, h_{i-1}^{T_P+1} &: \delta) \equiv \frac{1-\delta}{\delta^{T_P}} \frac{\pi_i(B, h_{i-1}^{T_P+1} : \delta)}{\bar{v}_i - \underline{v}_i}. \end{aligned} \tag{40}$$

If (6) is satisfied, then for sufficiently large δ, $p(G, h_{i-1}^{T_P+1} : \delta)$, $p(B, h_{i-1}^{T_P+1} : \delta) \in [0, 1]$ for all $h_{i-1}^{T_P+1}$. We see the repeated game as the repetition of T_P-period "review phases." In each phase, player i has a state $x_i \in \{G, B\}$. Within the phase, player i with state x_i plays according to $\sigma_i(x_i)$ in the current phase. After observing $h_i^{T_P+1}$ in the current phase, the state in the next phase is equal to G with probability $p(x_i, h_i^{T_P+1} : \delta)$ and B with the remaining probability.

Player $i-1$'s initial state is equal to G with probability p_v^{i-1} and B with probability $1 - p_v^{i-1}$ such that

$$p_v^{i-1} \bar{v}_i + (1 - p_v^{i-1}) \underline{v}_i = v_i.$$

Then, since

$$\begin{aligned} & (1-\delta) \sum_{t=1}^{T_P} \delta^{t-1} u_i(a_t) + \delta^{T_P} \left[p(G, h_{i-1}^{T_P+1} : \delta) \bar{v}_i + (1 - p(G, h_{i-1}^{T_P+1} : \delta)) \underline{v}_i \right] \\ = \; & \left(1 - \delta^{T_P}\right) \frac{1-\delta}{1-\delta^{T_P}} \left\{ \sum_{t=1}^{T_P} \delta^{t-1} u_i(a_t) + \pi_i(G, h_{i-1}^{T_P+1} : \delta) \right\} + \delta^{T_P} \bar{v}_i \end{aligned}$$

and

$$
\begin{aligned}
&(1-\delta)\sum_{t=1}^{T_P}\delta^{t-1}u_i\left(a_t\right)+\delta^{T_P}\left[p(B,h_{i-1}^{T_P+1}:\delta)\bar{v}_i+(1-p(B,h_{i-1}^{T_P+1}:\delta))\underline{v}_i\right] \\
=\ &\left(1-\delta^{T_P}\right)\frac{1-\delta}{1-\delta^{T_P}}\left\{\sum_{t=1}^{T_P}\delta^{t-1}u_i\left(a_t\right)+\pi_i(B,h_{i-1}^{T_P+1}:\delta)\right\}+\delta^{T_P}\underline{v}_i,
\end{aligned}
$$

(4) and (5) imply that, for sufficiently large discount factor δ,

1. conditional on the opponent's state, the above strategy in the infinitely repeated game is optimal;
2. if player $i-1$ is in the state G, then player i's payoff from the infinitely repeated game is $\bar{v}_i$ and if player $i-1$ is in the state B, then player i's payoff is $\underline{v}_i$;
3. the payoff in the initial period is $p_v^{i-1}\bar{v}_i+(1-p_v^{i-1})\underline{v}_i=v_i$ as desired.

10.2 Proof of Lemma 2

Construction of $\pi_i[\alpha]$ By linear independence of $\left(q_{i-1}\left(a_i,a_{-i}\right)\right)_{a_i\in A_i}$ for all $a_{-i}\in A_{-i}$ (Assumption 3), for all $\alpha\in\Delta(A)$, there exists $\pi_i[\alpha]:Y_{i-1}\to\mathbb{R}$ such that

$$
u_i(a_i,\alpha_{-i})+\mathbb{E}\left[\pi_i[\alpha](y_{i-1})\mid a_i,\alpha_{-i}\right]=0.
$$

Without loss, we assume that $\pi_i[\alpha](y_{i-1})$ is upper hemi-continuous with respect to α. Since $\Delta\left(A\right)\ni\alpha$ is compact, there exists $\bar{u}$ such that $\pi_i[\alpha]:Y_{i-1}\to(-\bar{u},\bar{u})$ for all $\alpha\in\Delta\left(A\right)$. Re-taking $\bar{u}$ if necessary, we can add or subtract a constant so that (30) is satisfied. In addition, since $\max_{a\in A}|u_i(a)|$ is bounded, we can make sure that (32) are satisfied, again re-taking $\bar{u}$ if necessary.

Construction of $\pi_i^{\delta}\left(t,a_{-i,t},y_{i-1,t}\right)$ Again, by linear independence of $\left(q_{i-1}\left(a_i,a_{-i,t}\right)\right)_{a_i\in A_i}$, we can construct $\pi_i^{\delta}\left(t,a_{-i,t},y_{i-1,t}\right)$ with (33). Since $\left(1-\delta^{t-1}\right)u_i\left(a_t\right)$ converges to 0 as δ goes to unity for all $t \in \{1, ..., T_P\}$ with $T_P = \Theta(T)$ and $T = (1-\delta)^{-\frac{1}{2}}$, we have

$$\lim_{\delta \to 1} \sup_{t \in \{1,...,T_P\}, a_{-i,t}, y_{i-1,t}} \left| \pi_i^{\delta}(t, a_{-i,t}, y_{i-1,t}) \right| = 0,$$

which implies (34).

10.3 Proof of Lemma 3

Fix $a_i, \alpha_{-(i,j)}$ arbitrarily and we omit $(a_i, \alpha_{-(i,j)})$. Let M be the maximum absolute value of the elements of Q_i^j. Since $\mathrm{aff}\left(\left\{q_i(a_i, a_j, \alpha_{-(i,j)})\right\}_{a_j \in A_j}\right) \subset \mathrm{aff}\left(\{\mathbf{1}_{y_i}\}_{y_i \in Y_i}\right)$, we can assume that the first row of Q_i^j is parallel to $(1, ..., 1)$ and that the first element of $\mathbf{q}_i^j$ is 1. Define

$$\begin{aligned} M^* &\equiv \frac{1}{2M+2} E + \frac{M+1}{2M+2} \begin{pmatrix} 1, 0, ..., 0 \\ \vdots \\ 1, 0, ..., 0 \end{pmatrix} \\ \tilde{Q}_i^j &\equiv M^* Q_i^j, \\ \tilde{\mathbf{q}}_i^j &\equiv M^* \mathbf{q}_i^j, \end{aligned}$$

that is, the (l, n) element of $\tilde{Q}_i^j$ is $\frac{\left(Q_i^j\right)_{l,n} + M + 1}{2M+2} \in (0,1)$ and the lth element of $\tilde{\mathbf{q}}_i^j$ is $\frac{\left(\mathbf{q}_i^j\right)_l + 1}{2M+2} \in (0,1)$.

Since M^* is invertible, $Q_i^j \mathbf{y}_i - \mathbf{q}_i^j = 0$ is equivalent to $\tilde{Q}_i^j \mathbf{y}_i - \tilde{\mathbf{q}}_i^j = 0$. Hence, we have

$$\boldsymbol{Q}_i^j = \left\{ \mathbf{y}_i \in \mathbb{R}_+^{|Y_i|} : Q_i^j \mathbf{y}_i = \mathbf{q}_i^j \right\} = \left\{ \mathbf{y}_i \in \mathbb{R}_+^{|Y_i|} : \tilde{Q}_i^j \mathbf{y}_i = \tilde{\mathbf{q}}_i^j \right\} \equiv \tilde{\boldsymbol{Q}}_i^j.$$

10.4 Proof of Lemma 4

Define $\bar{K}_1$

$$2/\bar{K}_1 \equiv \min_{f_i(Y_i|a_i)} \left\| Q_i^j(a_i, \alpha_{-(i,j)}) f_i(Y_i \mid a_i) - \mathbf{q}_i^j(a_i, \alpha_{-(i,j)}) \right\|$$

subject to $d(f_i(Y_i \mid a_i), \boldsymbol{Q}_i^j(a_i, \alpha_{-(i,j)})) \geq \varepsilon$ and $f_i(Y_i \mid a_i)$ being included in the simplex on Y_j. Since the objective function is continuous and the set of $f_i(Y_i \mid a_i)$ satisfying the constraints is compact, the minimum is well defined. Since $d(f_i(Y_i \mid a_i), \boldsymbol{Q}_i^j(a_i, \alpha_{-(i,j)})) \geq \varepsilon$ implies $\left\| Q_i^j(a_i, \alpha_{-(i,j)}) f_i(Y_i \mid a_i) - \mathbf{q}_i^j(a_i, \alpha_{-(i,j)}) \right\| > 0$, $\bar{K}_1 < \infty$.

Since the triangle inequality guarantees that (35) and (36) imply $\left\| Q_i^j(a_i, \alpha_{-(i,j)}) f_i(Y_i \mid a_i) - \mathbf{q}_i^j(a_i, \alpha_{-(i,j)}\right.$ $2/K_1$, which means $d(f_i(Y_i \mid a_i), \boldsymbol{Q}_i^j(a_i, \alpha_{-(i,j)})) < \varepsilon$.

10.5 Proof of Lemma 5

The belief of player i about player j's action α_j given $\lambda(l) = \emptyset$ and $\alpha_{-(i,j)} = \alpha_{-(i,j)}(x)$ is calculated by

$$
\begin{aligned}
& \log \frac{\Pr(\alpha_j \mid \{a_{i,t}, y_{i,t}\}_{t \in T(l)}, \alpha_{-(i,j)}(x))}{\Pr(\alpha_j(x) \mid \{a_{i,t}, y_{i,t}\}_{t \in T(l)}, \alpha_{-(i,j)}(x))} \\
= & \log \prod_{a_i, y_i} \frac{(q_i(y_i \mid a_i, \alpha_j, \alpha_{-(i,j)}(x)))^{f_i(a_i,y_i)T}}{(q_i(y_i \mid a_i, \alpha_j(x), \alpha_{-(i,j)}(x)))^{f_i(a_i,y_i)T}} + \log \frac{\eta}{2(1-\eta)} \\
= & T \sum_{a_i, y_i} f_i(a_i, y_i) \left(\log q_i(y_i \mid a_i, \alpha_j, \alpha_{-(i,j)}(x)) - \log q_i(y_i \mid a_i, \alpha_j(x), \alpha_{-(i,j)}(x)) \right) \\
& + \log \frac{\eta}{2(1-\eta)} \qquad (41)
\end{aligned}
$$

where $f_i(a_i, y_i)$ is the frequency of (a_i, y_i) in periods $T(l)$.

Imagine player j takes

$$
\alpha_j(x, \boldsymbol{\lambda}_j) \equiv \left(1 - \sum\nolimits_{a_j \neq a_j(x)} (2\rho + \lambda_j(a_j)) \right) a_j(x) + \left(\sum\nolimits_{a_j \neq a_j(x)} (2\rho + \lambda_j(a_j)) \, a_j \right)
$$

for some $\boldsymbol{\lambda}_j \in [\rho, \rho]^{|A_j|-1}$ and consider

$$
\mathcal{L}_i^j(x, \boldsymbol{\lambda}_j) \equiv \sum_{a_i, y_i \in Y_i(a_i)} f_i(a_i, y_i) \log q_i(y_i \mid a_i, \alpha_j(x, \boldsymbol{\lambda}_j), \alpha_{-(i,j)}(x)) \qquad (42)
$$

with

$$\nabla \mathcal{L}_i^j(x, \boldsymbol{\lambda}_j) \equiv \left(\sum_{a_i, y_i} f_i(a_i, y_i) \frac{q_i(y_i \mid a_i, a_j, \alpha_{-(i,j)}(x)) - q_i(y_i \mid a_i, a_j(x), \alpha_{-(i,j)}(x))}{q_i(y_i \mid a_i, \alpha_j(x, \boldsymbol{\lambda}_j), \alpha_{-(i,j)}(x))} \right)_{a_j \neq a_j(x)}. \tag{43}$$

We need to show the following two arguments given that h_i^l satisfies neither Condition 1 nor 2 for the belief of $X_{i-1}(l)$: first, if there exists j with $\left\|\nabla \mathcal{L}_i^j(x, \mathbf{0})\right\| > \eta$, then player i believes that $\alpha_j(l) \neq \alpha_j(x)$ with a high probability. Second, if $\left\|\nabla \mathcal{L}_i^j(x, \mathbf{0})\right\| \leq \eta$ for all j, then player i believes that $X_{i-1}(l)$ is close to the ex ante value 0.

The event that h_i^l satisfies neither Condition 1 nor 2 for the belief of $X_{i-1}(l)$ is equivalent to satisfying both of the following two conditions:

1. player i takes $\alpha_i(l) = \alpha_i(x)$ and player i's action frequency is close to the ex ante distribution $\alpha_i(x)$;

2. there exists $j \in -i$ such that $f_i(Y_i \mid a_i(x))$ is close to $\boldsymbol{Q}_i^j(a_i(x), \alpha_{-(i,j)}(x))$.

In the following part of this subsection, when we say Conditions 1 and 2, we refer to these conditions above rather than Conditions 1 and 2 for the belief of $X_{i-1}(l)$.

Proof of the First Part We show that, if there exists j with $\left\|\nabla \mathcal{L}_i^j(x, \mathbf{0})\right\| > \eta$, then player i puts a belief no less than $1 - \exp(-\Theta(T))$ on the event that $\alpha_j(l) \neq \alpha_j(x)$ given $\alpha_{-(i,j)}(l) = \alpha_{-(i,j)}(x)$.

By (41), it suffices to show that there exists $\bar{\eta} > 0$ such that, for any $\eta < \bar{\eta}$, there exists $\bar{\rho} > 0$ such that for all $\rho < \bar{\rho}$, there exist $\bar{C}_j$ and $k > 0$ such that if $\left\|\nabla \mathcal{L}_i^j(x, \mathbf{0})\right\| > \eta$, then there exists $\mathbf{t}_j \in \bar{C}_j$ with

$$\mathcal{L}_i^j(x, \boldsymbol{\lambda}_j) - \mathcal{L}_i^j(x, \mathbf{0}) \geq kT,$$

where $\boldsymbol{\lambda}_j$ is equal to $\rho \mathbf{t}_j$.

Take Taylor expansion of $\mathcal{L}_i^j(x, \boldsymbol{\lambda}_j)$ with $\boldsymbol{\lambda}_j = \rho \mathbf{t}_j$ around $\mathbf{0}$:

$$\begin{aligned}
& \sum_{a_i, y_i} f_i(a_i, y_i) \log q_i(y_i \mid a_i, \alpha_j(\boldsymbol{\lambda})) \\
= & \sum_{a_i, y_i} f_i(a_i, y_i) \log q_i(y_i \mid a_i, \alpha_j(x)) + \nabla \mathcal{L}_i^j(x, \mathbf{0}) \cdot \rho \mathbf{t}_j + \frac{1}{2}(\rho \mathbf{t}_j^\top) H_i^j(x, \tilde{\rho} \mathbf{t}_j)(\rho \mathbf{t}_j)
\end{aligned}$$

for some $\tilde{\rho} \in [0, \rho]$. Here, $H_i^j(x, \tilde{\rho} \mathbf{t}_j)$ is Hessian matrix for the second derivative of $\mathcal{L}_i^j(x, \boldsymbol{\lambda}_j)$ with respect to $\boldsymbol{\lambda}_j$.

Consider the lower bound of the second term:

$$\begin{aligned}
& \min_{\nabla \mathcal{L}_i^j(x,\mathbf{0}): \left\| \nabla \mathcal{L}_i^j(x,\mathbf{0}) \right\| \geq \eta} \max_{\mathbf{t}_j \in \bar{C}_j} \nabla \mathcal{L}_i^j(x, \mathbf{0}) \cdot \rho \mathbf{t}_j \\
\geq & \rho \left(\min_{\nabla \mathcal{L}_i^j(x,\mathbf{0}): \left\| \nabla \mathcal{L}_i^j(x,\mathbf{0}) \right\| \geq \eta} \max_{\mathbf{t}_j \in C_j} \nabla \mathcal{L}_i^j(x, \mathbf{0}) \cdot \rho \mathbf{t}_j - \max_{\left\| \nabla \mathcal{L}_i^j(x,\mathbf{0}) \right\|} \left\| \nabla \mathcal{L}_i^j(x, \mathbf{0}) \right\| \max_{\mathbf{t}_j \in C_j} \min_{\bar{\mathbf{t}} \in \bar{C}_j} \left\| \mathbf{t}_j - \bar{\mathbf{t}}_j \right\| \right) \\
= & \rho \left(\eta - \max_{\left\| \nabla \mathcal{L}_i^j(x,\mathbf{0}) \right\|} \left\| \nabla \mathcal{L}_i^j(x, \mathbf{0}) \right\| \max_{\mathbf{t}_j \in C_j} \min_{\bar{\mathbf{t}} \in \bar{C}_j} \left\| \mathbf{t}_j - \bar{\mathbf{t}}_j \right\| \right).
\end{aligned}$$

Since $\left\| \nabla \mathcal{L}_i^j(x, \mathbf{0}) \right\|$ is bounded for all $\rho \in [0, 1]$ and the signal distribution $(f_i(a_i, y_i))_{a_i, y_i}$, if we take $\bar{C}_j$ sufficiently dense, we have

$$\min_{\nabla \mathcal{L}_i^j(x,\mathbf{0}): \left\| \nabla \mathcal{L}_i^j(x,\mathbf{0}) \right\| \geq \eta} \max_{\mathbf{t}_j \in \bar{C}_j} \nabla \mathcal{L}_i^j(x, \mathbf{0}) \cdot \rho \mathbf{t}_j > \frac{1}{2} \eta \rho$$

for all $\rho \in [0, 1]$.

Next, consider the upper bound of the third term: Assumption 2 guarantees that there exists K_2 such that, for all $\mathbf{x} \in \mathbb{R}^{|A_j|-1}$, $\max_{\tilde{\rho} \in [0,1], \mathbf{t}_j \in C_j, \{a_{i,t}, y_{i,t}\}_{t \in T(l)}} \mathbf{x}^\top H_i^j(x, \tilde{\rho} \mathbf{t}_j) \mathbf{x} \leq K_2 \left\| \mathbf{x} \right\|^2$.

Therefore, for $\rho \leq \bar{\rho} \equiv \eta / 4K_2$ and $k \equiv \frac{\eta}{4} \rho$, we have $\mathcal{L}_i^j(x, \boldsymbol{\lambda}_j) - \mathcal{L}_i^j(x, \mathbf{0}) \geq kT$, as desired.

Proof of the Second Part Hence, we will show that if $\left\| \nabla \mathcal{L}_i^j(x, \mathbf{0}) \right\| \leq \eta$ for all j and Conditions 1 and 2 above are satisfied, then player i believes that $X_n(l)$ is close to the ex ante value 0 for all $n \in -i$ given $\alpha_{-i}(x)$.

Fix j so that $f_i(Y_i \mid a_i(x))$ is close to $\boldsymbol{Q}_i^j(a_i(x), \alpha_{-(i,j)}(x))$. Since $\left\|\nabla \mathcal{L}_i^j(x, \mathbf{0})\right\| \leq \eta$, we have

$$\left\| \left(\sum_{a_i, y_i} f_i(a_i, y_i) \frac{q_i(y_i \mid a_i, a_j, \alpha_{-(i,j)}(x)) - q_i(y_i \mid a_i, a_j(x), \alpha_{-(i,j)}(x))}{q_i(y_i \mid a_i, \alpha_{-i}(x))} \right)_{a_j \neq a_j(x)} \right\| \leq \eta.$$

By Assumption 2, for sufficiently small ρ and ε, this implies that

$$\left\| \left(\sum_{y_i} f_i(y_i \mid a_i(x)) \frac{\Delta_i^x(y_i, a_j)}{q_i(y_i \mid a_i(x), \alpha_{-i}(x))} \right)_{a_j \neq a_j(x)} \right\| \leq 2\eta \tag{44}$$

with

$$\Delta_i^x(y_i, a_j) \equiv q_i(y_i \mid a_i(x), a_j, \alpha_{-(i,j)}(x)) - q_i(y_i \mid a_i(x), a_j(x), \alpha_{-(i,j)}(x)).$$

In addition, since Condition 2 implies (37), there exist $\mathbf{t} \in \mathbb{R}^{|A_j|-1}$ and $\boldsymbol{\varepsilon} \in \mathbb{R}^{|Y_i|}$ with $\|\boldsymbol{\varepsilon}\| < \varepsilon$ such that

$$f_i(y_i \mid a_i(x)) = \sum_{\tilde{a}_j \in A^*(x)} t(\tilde{a}_j) \Delta_i^x(y_i, \tilde{a}_j) + q_i(a_i(x), \alpha_{-i}(x)) + \boldsymbol{\varepsilon}, \tag{45}$$

where $A_j^*(x)$ is the set of $a_j \neq a_j(x)$ so that for all $a_j \in A_j^*(x)$, $(\Delta_i^x(y_i, a_j))_{y_j \in Y_j}$ is linearly independent.

By Assumption 2, for sufficiently small ε, (44) and (45) imply

$$-3\eta \leq \sum_{y_i} \sum_{\tilde{a}_j \in A^*(x)} \frac{t(\tilde{a}_j) \Delta_i^x(y_i, \tilde{a}_j) \Delta_i^x(y_i, a_j)}{q_i(y_i \mid a_i(x), \alpha_{-i}(x))} \leq 3\eta$$

for all $a_j \neq a_j(x)$.

Multiplying $t(a_j)$ and adding them up with respect to $a_j \in A_j^*(x)$ yield

$$-3\eta \sum_{\tilde{a}_j \in A^*(x)} |t(a_j)| \leq \sum_{y_i} \frac{\left(\sum_{\tilde{a}_j \in A^*(x)} t(a_j) \Delta_i^x(y_i, a_j) \right)^2}{q_i(y_i \mid a_i(x), \alpha_{-i}(x))} \leq 3\eta \sum_{\tilde{a}_j \in A^*(x)} |t(a_j)|.$$

Since there exists K_3 such that $|t(a_j)| \leq K_3$ for all the signal distributions satisfying (45), this implies

$$\left\| \sum_{\tilde{a}_j \in A^*(x)} t(a_j)(q_i(a_i(x), a_j) - q_i(a_i(x), a_j(x))) \right\| \leq 3\eta K_4$$

for some K_4. Since $(\Delta_i^x(y_i, a_j))_{y_j \in Y_j}$ is linearly independent, there exists K_5 such that $|t(a_j)| \leq 3\eta K_5$ for all $\tilde{a}_j \in A^*(x)$, that is, $f_i(Y_i \mid a_i(x))$ is close to the ex ante distribution $q_i(a_i(x), \alpha_{-i}(x))$.

For sufficiently small η, ε and ρ, since player i takes $a_i(x)$ sufficiently often and $f_i(Y_i \mid a_i(x))$ is close to the ex ante distribution $q_i(a_i(x), \alpha_{-i}(x))$, player i's expectation of $|X_n(l)|$ is no more than $\frac{1}{2}\frac{\bar{u}}{L}T$ from (31) for all $n \in -i$. By the central limit theorem, this means player i believes that $|X_n(l)| \leq \frac{\bar{u}}{L}T$ with probability $1 - \exp(-\Theta(T))$, as desired.

10.6 Proof of Lemma 6

Since player $i-1$ triggers the punishment if she tells the truth about h_{i-1}^l, h_{i-1}^l satisfies the following three conditions:

1. player $i-1$ takes $\alpha_{i-1}(l) = \alpha_{i-1}(x)$ and player $i-1$'s action frequency is close to the ex ante distribution $\alpha_{i-1}(x)$;

2. $f_{i-1}(Y_{i-1} \mid a_{i-1}(x))$ is close to $\boldsymbol{Q}_{i-1}^i(a_{i-1}(x), \alpha_{-(i-1,i)}(x))$;

3. $X_{i-1}(l) < -\frac{\bar{u}}{L}$.

By the same proof as in Lemma 5, if $\left\|\nabla \mathcal{L}_{i-1}^i(x, \mathbf{0})\right\| \leq \eta$, then Conditions 1 and 2 imply that $f_{i-1}(Y_{i-1} \mid a_{i-1}(x))$ is close to the ex ante distribution $q_{i-1}(a_{i-1}(x), \alpha_{-(i-1)}(x))$, which contradicts Condition 3 for sufficiently small ρ. Hence, $\left\|\nabla \mathcal{L}_{i-1}^i(x, \mathbf{0})\right\| > \eta$ and so player $i-1$ believes that player i took $\alpha_i(l) \neq \alpha_i(x)$.

10.7 Proof of Proposition 1

For (6), it suffices to have

$$|\bar{\pi}_i(x, \lambda(l), l)|, |\bar{\pi}_i(x, \alpha_{-i}(l), l)| \leq \bar{u}LT, \tag{46}$$

$$\bar{\pi}_i(x, \lambda(l), l), \bar{\pi}_i(x, \alpha_{-i}(l), l) \begin{cases} \leq 0 & \text{if } x_{i-1} = G, \\ \geq 0 & \text{if } x_{i-1} = B, \end{cases} \tag{47}$$

and

$$\left| v_i(\{\lambda(\tilde{l})\}_{\tilde{l}=1}^{l-1}, \{h_{-i}^{\tilde{l}}\}_{\tilde{l}=1}^{l}, \hat{h}_i^l) \right| \leq \exp(-\Theta(T)) \tag{48}$$

for all $x \in \{G, B\}^N$, $\{\lambda(l)\}_{l=1}^L \in \{G, B\}^L$, $l \in \{1, ..., L\}$ and $\{h_{-i}^l, \hat{h}_i^l\}_{l=1}^L$. To see why (46), (47) and (48) are sufficient, notice the following: first, (46) and (48) with $T = (1-\delta)^{-\frac{1}{2}}$ implies that

$$\lim_{\delta \to 1} \frac{1-\delta}{\delta^{T_P}} \sup_{x_{i-1}, h_{i-1}^{T_P+1}} \left| \pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta) \right| = 0,$$

as desired.

Second, for $x_{i-1} = G$, $\pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta)$ defined in (38) is always non-positive: if $\theta_{i-1} = B$, then $\pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta) \leq 0$ since $\text{sign}(x_{i-1}) \mathbf{1}_{\{\theta_{i-1}=B\}} 3\bar{u}LT$ is sufficiently small. If $\theta_{i-1} = G$, then either $\lambda(l) = \emptyset$, i or punish by Condition 2 in Section 6.5.2. Since $\theta_{i-1} = G$, whenever $X_{i-1}(l) > \frac{\bar{u}}{L}T$ happens, $\lambda(l+1) = i$ is induced by Condition 1 for θ_{i-1}. Hence,

$$\text{sign}(x_{i-1}) 2\bar{u}T + \sum_{l=1}^{L} X_{i-1}(l) \leq -2\bar{u}T + (L-1)\frac{\bar{u}}{L}T + \bar{u}T \leq -\frac{\bar{u}}{L}T,$$

which means $\pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta)$ is always non-positive.

Third, for $x_{i-1} = B$, $\pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta)$ defined in (38) is always non-negative: if $\theta_{i-1} = B$, then $\pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta) \geq 0$ since $\text{sign}(x_{i-1}) \mathbf{1}_{\{\theta_{i-1}=B\}} 3\bar{u}LT$ is sufficiently large. If $\theta_{i-1} = G$, then either $\lambda(l) = \emptyset$ or punish by Condition 2 for θ_{i-1}. From Section 6.5.1, whenever

$X_{i-1}(l) < -\frac{\bar{u}}{L}T$ happens, $\lambda(l+1) = \text{punish}$ is induced. Hence,

$$\text{sign}(x_{i-1})2\bar{u}T + \sum_{l=1}^{L} X_{i-1}(l) \geq 2\bar{u}T - (L-1)\frac{\bar{u}}{L}T - \bar{u}T \geq \frac{\bar{u}}{L}T,$$

which means $\pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta)$ is always non-negative.

Next, we will verify the optimality of $\sigma_i(x_i)$ and derive player i's payoffs by backward induction.

When player i sends h_i^L in round L (the last round), the only relevant part of the reward is

$$\sum_{t \in T(L)} -T^{-2} \left\| \mathbf{1}_{a_{-i,t}, y_{-i,t}} - \mathbb{E}[\mathbf{1}_{a_{-i,t}, y_{-i,t}} \mid \hat{a}_{i,t}, \hat{y}_{i,t}, \alpha_{-i,t}] \right\|^2, \tag{49}$$

which makes it optimal to tell the truth about h_i^L.

Given the truthtelling incentive, $\pi_i^{\text{report}}[\alpha_{-i,t}](y_{i-1,t})$ cancels out the difference in (49) for different actions by (28). Therefore, the fourth line of (38) does not affect player i's incentive in round L.

In round L, there are following cases:

1. if $\theta_{i-1} = B$, then the second line of (38) makes any action optimal;

2. if $\theta_{i-1} = G$, then Section 6.5.2 implies that $\lambda(L) = \emptyset$, i or punish, and $\alpha_{-i}(L) = \alpha_{-i}(x)$ if $\lambda(L) = \emptyset$ or i and $\alpha_{-i}(L) = \alpha_{-i}^*$ if $\lambda(L) = \text{punish}$. Hence,

    (a) if $\lambda(L) = \emptyset$, then the third line of (38) makes any action optimal;

    (b) if $\lambda(L) = i$, since (38) is not sensitive to player $i-1$'s history in round L and $\alpha_{-i}(L) = \alpha_{-i}(x)$, it is optimal for player i to take the static best response to $\alpha_{-i}(x)$;

    (c) if $\lambda(L) = \text{punish}$, since (38) is not sensitive to player $i-1$'s history in round L and $\alpha_{-i}(L) = \alpha_{-i}^*$, it is optimal for player i to take the static best response to α_{-i}^*, that is, α_i^*.

Hence, $\sigma_i(x_i)$ is optimal.

Now, consider player i's payoff from round L. Given $\theta_{i-1} = B$, we define $\bar{\pi}_i(x, \alpha_{-i}(L), L)$ so that player i's payoff in round L, defined as

$$\frac{1}{T}\mathbb{E}\left[\sum_{t \in T(l)} u_i(a_t) + \bar{\pi}_i(x, \alpha_{-i}(L), L) + X_{i-1}(L) \mid \alpha_{-i}(L)\right],$$

is equal to 0. That is, player i's value is independent of $\alpha_{-i}(L)$. By Lemma 2, $|\bar{\pi}_i(x, \alpha_{-i}(L), L)| \leq \bar{u}T$.

Given $\theta_{i-1} = G$,

1. if $\lambda(L) = \emptyset$, then

$$\frac{1}{T}\mathbb{E}\left[\sum_{t \in T(l)} u_i(a_t) + X_{i-1}(L) \mid \alpha_{-i}(L)\right] = u_i(\alpha(x)) \begin{cases} \geq \min_{x:x_{i-1}=G} u_i(\alpha(x)) & \text{if } x_{i-1} = G \\ \leq \max\left\{v_i^*, \max_{x:x_{i-1}=B} u_i(a(x))\right\} & \text{if } x_{i-1} = B \end{cases}$$

since player i takes $\alpha_i(x)$ and $\alpha_{-i}(L) = \alpha_{-i}(x)$;

2. if $\lambda(L) = i$, then

$$\frac{1}{T}\mathbb{E}\left[\sum_{t \in T(l)} u_i(a_t) \mid \alpha_{-i}(L)\right] \geq u_i(\alpha(x)) \geq \min_{x:x_{i-1}=G} u_i(\alpha(x))$$

since player i takes the static best response to $\alpha_{-i}(x)$ and $\alpha_{-i}(L) = \alpha_{-i}(x)$. Note that, from Section 6.5.1, $\lambda(L) = i$ implies $x_{i-1} = G$;

3. if $\lambda(L) = \text{punish}$, then

$$\frac{1}{T}\mathbb{E}\left[\sum_{t \in T(l)} u_i(a_t) \mid \alpha_{-i}(L)\right] = v_i^* \leq \max\left\{v_i^*, \max_{x:x_{i-1}=B} u_i(\alpha(x))\right\}.$$

Therefore, there exists $\bar{\pi}_i(x, \lambda_j(L), L)$ with (46) and (47) such that player i's payoff in round L is equal to

$$\begin{cases} \min_{x:x_{i-1}=G} u_i(\alpha(x)) & \text{if } x_{i-1}=G \text{ and } \lambda(L)\neq \text{punish}, \\ \max\left\{v_i^*, \max_{x:x_{i-1}=B} u_i(\alpha(x))\right\} & \text{if } x_{i-1}=B \text{ or } \lambda(L)=\text{punish} \end{cases}$$

for all $\lambda(L)$.

In total, player i's payoff in round L satisfies

$$\begin{cases} 0 & \text{if } \theta_{i-1}=B, \\ \min_{x:x_{i-1}=G} u_i(\alpha(x)) & \text{if } \theta_{i-1}=G,\ x_{i-1}=G \text{ and } \lambda(L)\neq\text{punish}, \\ \max\left\{v_i^*, \max_{x:x_{i-1}=B} u_i(\alpha(x))\right\} & \text{if } \theta_{i-1}=G \text{ and ``} x_{i-1}=B \text{ or and } \lambda(L)=\text{punish.''} \end{cases} \tag{50}$$

Again, all the cases with $\lambda(L)=j\in -i$ is included in the cases with $\theta_{i-1}=B$.

Given this value, let us define $v_i(\{\lambda(\tilde{l})\}_{\tilde{l}=1}^{L-2},\{h_{-i}^{\tilde{l}}\}_{\tilde{l}=1}^{L-1},\hat{h}_i^{L-1})$: with $V_i^L(\{\lambda(\tilde{l})\}_{\tilde{l}=1}^{L-2},\{h_{-i}^{\tilde{l}}\}_{\tilde{l}=1}^{L-1},\hat{h}_i^{L-1})$ be player i's payoff in round L given $\{\lambda(\tilde{l})\}_{\tilde{l}=1}^{L-2}$, $\{h_{-i}^{\tilde{l}}\}_{\tilde{l}=1}^{L-1}$ and player i's message $\hat{h}_i^{L-1}$,

$$\begin{aligned} v_i(\{\lambda(\tilde{l})\}_{\tilde{l}=1}^{L-2},\{h_{-i}^{\tilde{l}}\}_{\tilde{l}=1}^{L-1},\hat{h}_i^{L-1}) &= \max_{\tilde{h}_i^{L-1}} \mathbb{E}\left[V_i^L(\{\lambda(\tilde{l})\}_{\tilde{l}=1}^{L-2},\{h_{-i}^{\tilde{l}}\}_{\tilde{l}=1}^{L-1},\tilde{h}_i^{L-1}) \mid \hat{h}_i^{L-1}\right] \\ &\quad -\mathbb{E}\left[V_i^L(\{\lambda(\tilde{l})\}_{\tilde{l}=1}^{L-2},\{h_{-i}^{\tilde{l}}\}_{\tilde{l}=1}^{L-1},\hat{h}_i^{L-1}) \mid \hat{h}_i^{L-1}\right]. \end{aligned}$$

That is, imagine $\hat{h}_i^{L-1}$ is the true history of player i. $v_i(\{\lambda(\tilde{l})\}_{\tilde{l}=1}^{L-2},\{h_{-i}^{\tilde{l}}\}_{\tilde{l}=1}^{L-1},\hat{h}_i^{L-1})$ gives player i's payoff in round L under the most profitable message $\tilde{h}_i^{L-1}$. We will show that (48) is satisfied. Consider the following two cases:

1. if there is $j\in -i$ who announces that player j believes that $X_{j-1}(L)$ is excessively high, then from Section 6.5.2, $\theta_{i-1}=B$ is determined and from (50), player i's continuation payoff in round L is 0 regardless of $\hat{h}_i^{L-1}$. Hence, $v_i(\{\lambda(\tilde{l})\}_{\tilde{l}=1}^{L-2},\{h_{-i}^{\tilde{l}}\}_{\tilde{l}=1}^{L-1},\hat{h}_i^{L-1})=0$;

2. if there is no $j\in -i$ who announces that player j believes that $X_{j-1}(L)$ is excessively high, then

   (a) suppose that $x_{i-1}=G$ and the message of player i's history, $\hat{h}_i^{L-1}$, means that player i believes that $X_{i-1}(L)$ is excessively high. In such a case, there are following three considerations:

i. given $\theta_{i-1} = B$, player i's continuation payoff in round L is independent of $\tilde{h}_i^{L-1}$.

   Given $\theta_{i-1} = G$, player i's continuation payoff in round L depends on $\tilde{h}_i^{L-1}$ in the following way: if $\lambda(L) \neq$ punish, she gets $\min_{x:x_{i-1}=G} u_i(\alpha(x))$ while if $\lambda(L) =$ punish, then player i gets $\max\left\{v_i^*, \max_{x:x_{i-1}=B} u_i(a(x))\right\}$. From Section 4, player i's payoff is maximized by not triggering the punishment. From Section 6.5.1, $\hat{h}_i^{L-1}$ minimizes the probability of $\lambda(L) =$ punish;

ii. now consider the transition of θ_{i-1}. By (2) and (50), player i's payoff is higher with $\theta_{i-1} = G$ than with $\theta_{i-1} = B$. From Section 6.5.2, probability of $\theta_{i-1} = B$ is minimized by sending $\hat{h}_i^{L-1}$;

Therefore, in 2-(a), $v_i(\{\lambda(\tilde{l})\}_{\tilde{l}=1}^{L-2}, \{h_{-i}^{\tilde{l}}\}_{\tilde{l}=1}^{L-1}, \hat{h}_i^{L-1}) = 0$;

(b) suppose that $x_{i-1} = G$ and the message of player i's history, $\hat{h}_i^{L-1}$, means that player i does not believe that $X_{i-1}(L)$ is excessively high and that player i does not trigger the punishment. In such a case, there are following two considerations:

i. given θ_{i-1}, player i's continuation payoff in round L depends on $\tilde{h}_i^{L-1}$ as in 2-(a)-i. However, by Lemma 5, player i puts the belief no less than $1 - \exp(-\Theta(T))$ on the event that there is no player among $-i$ who triggers the punishment or $\theta_{i-1} = B$ is determined and player i's payoff in round L is 0 regardless of $\tilde{h}_i^{L-1}$ (if $\hat{h}_i^{L-1}$ is the true history). Hence, player i believes that there is no $\tilde{h}_i^{L-1}$ which increases the continuation payoff by more than $\exp(-\Theta(T))$ compared to $\hat{h}_i^{L-1}$;

ii. now consider the transition of θ_{i-1}. Again, player i's payoff is higher with $\theta_{i-1} = G$ than with $\theta_{i-1} = B$. However, from Section 6.5.2 and Lemma 5, player i puts the belief no less than $1 - \exp(-\Theta(T))$ on the event that Condition 1 for $\theta_{i-1} = B$ is not the case or $\theta_{i-1} = B$ is already determined independently of $\hat{h}_i^{L-1}$. Hence, player i believes that there is no $\tilde{h}_i^{L-1}$ which increases the continuation payoff by more than $\exp(-\Theta(T))$ compared to $\hat{h}_i^{L-1}$;

Therefore, in 2-(b), $\left|v_i(\{\lambda(\tilde{l})\}_{\tilde{l}=1}^{L-2}, \{h_{-i}^{\tilde{l}}\}_{\tilde{l}=1}^{L-1}, \hat{h}_i^{L-1})\right| \leq \exp(-\Theta(T))$;

(c) suppose that $x_{i-1} = G$ and the message of player i's history, $\hat{h}_i^{L-1}$, means that player i does not believe that $X_{i-1}(L)$ is excessively high and that player i triggers the punishment. In such a case, there are following two considerations:

i. given θ_{i-1}, player i's continuation payoff in round L depends on $\tilde{h}_i^{L-1}$ as in 2-(a)-i. However, by Lemma 6, player i puts the belief no less than $1 - \exp(-\Theta(T))$ on the event that $\theta_{i-1} = B$ is determined and player i's payoff in round L is 0 regardless of $\tilde{h}_i^{L-1}$. Hence, player i believes that there is no $\tilde{h}_i^{L-1}$ which increases the continuation payoff by more than $\exp(-\Theta(T))$ compared to $\hat{h}_i^{L-1}$;

ii. now consider the transition of θ_{i-1}. As stated above, player i puts the belief no less than $1-\exp(-\Theta(T))$ on the event that $\theta_{i-1} = B$ is already determined independently of $\tilde{h}_i^{L-1}$. Hence, player i believes that there is no $\tilde{h}_i^{L-1}$ which increases the continuation payoff by more than $\exp(-\Theta(T))$ compared to $\hat{h}_i^{L-1}$;

Therefore, in 2-(c), $\left|v_i(\{\lambda(\tilde{l})\}_{\tilde{l}=1}^{L-2}, \{h_{-i}^{\tilde{l}}\}_{\tilde{l}=1}^{L-1}, \hat{h}_i^{L-1})\right| \leq \exp(-\Theta(T))$;

(d) suppose that $x_{i-1} = B$. In such a case, there are following two considerations:

i. given θ_{i-1}, player i's continuation payoff in round L is independent of $\tilde{h}_i^{L-1}$ by (50);

ii. now consider the transition of θ_{i-1}. From Section 6.5.2, the transition of $\theta_{i-1} = B$ is independent of $\tilde{h}_i^{L-1}$ if $x_{i-1} = B$;

Therefore, in 2-(d), $v_i(\{\lambda(\tilde{l})\}_{\tilde{l}=1}^{L-2}, \{h_{-i}^{\tilde{l}}\}_{\tilde{l}=1}^{L-1}, \hat{h}_i^{L-1}) = 0$.

In total, we have (48). This implies the following three facts:

1.

$$\sum_{t\in T(L-1)} -T^{-2}\left\|\mathbf{1}_{a_{-i,t},y_{-i,t}}-\mathbb{E}[\mathbf{1}_{a_{-i,t},y_{-i,t}}\mid \hat{a}_{i,t},\hat{y}_{i,t},\alpha_{-i,t}]\right\|^2 \tag{51}$$

is sufficiently large to incentivize player i to tell the truth about h_i^L;

2. given the truthtelling incentive, $\pi_i^{\text{report}}[\alpha_{-i,t}](y_{i-1,t})$ cancels out the difference in the (49) for different actions by (28). Therefore, the fourth line of (38) does not affect player i's incentive in round $L-1$.

3. given the truthtelling incentive, $v_i(\{\lambda(\tilde{l})\}_{\tilde{l}=1}^{L-2}, \{h_{-i}^{\tilde{l}}\}_{\tilde{l}=1}^{L-1}, \hat{h}_i^{L-1})$ cancels out differences in player i's expected payoffs in round L for different histories of player i from the perspective of player i in round $L-1$. Therefore, player i in round $L-1$ wants to maximize

$$\frac{1}{T}\mathbb{E}\left[\sum_{t\in T(L-1)} u_i(a_t)+\bar{\pi}_i(x,\alpha_{-i}(L-1),L-1)+X_{i-1}(L-1)\mid \alpha_{-i}(L-1)\right]. \tag{52}$$

Hence, as in round L, we can define $\bar{\pi}_i(x,\lambda_j(L-1),L-1)$ and $\bar{\pi}_i(x,\alpha_{-i}(L-1),L-1)$ with (46) and (47) so that player i's payoff in round $L-1$, defined as (52), is equal to

$$\begin{cases} 0 & \text{if } \theta_{i-1}=B, \\ \min_{x:x_{i-1}=G} u_i(\alpha(x)) & \text{if } \theta_{i-1}=G,\ x_{i-1}=G \text{ and } \lambda(L-1)\neq \text{punish}, \\ \max\left\{v_i^*, \max_{x:x_{i-1}=B} u_i(a(x))\right\} & \text{if } \theta_{i-1}=G \text{ and ``} x_{i-1}=B \text{ or } \lambda(L-1)=\text{punish.''} \end{cases}$$

By the same argument, there exists $v_i(\{\lambda(\tilde{l})\}_{\tilde{l}=1}^{L-3}, \{h_{-i}^{\tilde{l}}\}_{\tilde{l}=1}^{L-2}, \hat{h}_i^{L-2})$ with (48) such that it is optimal for player i to tell the truth about h_i^{L-2} and player i in round $L-2$ wants to maximize (52) with $L-1$ replaced with $L-2$.

Recursively, for each l, we can make sure that $\sigma_i(x)$ is optimal and player i's payoff in round l is equal to

$$\begin{cases} 0 & \text{if } \theta_{i-1} = B, \\ \min_{x:x_{i-1}=G} u_i(\alpha(x)) & \text{if } \theta_{i-1} = G,\ x_{i-1} = G \text{ and } \lambda(l) \neq \text{punish}, \\ \max\left\{v_i^*, \max_{x:x_{i-1}=B} u_i(a(x))\right\} & \text{if } \theta_{i-1} = G \text{ and ``} x_{i-1} = B \text{ or } \lambda(l) = \text{punish.''} \end{cases}$$

From Section 6.5.2 and the central limit theorem, $\theta_{i-1} = B$ happens with probability no more than $NL\eta$. Given $\theta_{i-1} = G$, by the central limit theorem, $\lambda(l) =$ punish happens only with probability $\exp(-\Theta(T))$. Therefore, from (39), we can further modify $\bar{\pi}_i(x, \lambda(1), 1)$ with (47) and (46) such that $\sigma_i(x_i)$ gives $\bar{v}_i$ ($\underline{v}_i$, respectively) if $x_{i-1} = G$ (B, respectively) without affecting the incentives. Hence, we are done.

References

Aoyagi, M. (2002): "Collusion in dynamic Bertrand oligopoly with correlated private signals and communication," *Journal of Economic Theory*, 102(1), 229–248.

Bhaskar, V., and I. Obara (2002): "Belief-based equilibria in the repeated prisoners' dilemma with private monitoring," *Journal of Economic Theory*, 102(1), 40–69.

Compte, O. (1998): "Communication in repeated games with imperfect private monitoring," *Econometrica*, 66(3), 597–626.

Deb, J. (2011): "Cooperation and community responsibility: A folk theorem for repeated random matching games," *mimeo.*

Ely, J., J. Hörner, and W. Olszewski (2005): "Belief-free equilibria in repeated games," *Econometrica*, 73(2), 377–415.

Ely, J., and J. Välimäki (2002): "A robust folk theorem for the prisoner's dilemma," *Journal of Economic Theory*, 102(1), 84–105.

Fong, K., O. Gossner, J. Hörner, and Y. Sannikov (2010): "Efficiency in a repeated prisoners' dilemma with imperfect private monitoring," *mimeo.*

Fudenberg, D., and D. Levine (2007): "The Nash-threats folk theorem with communication and approximate common knowledge in two player games," *Journal of Economic Theory*, 132(1), 461–473.

Fudenberg, D., D. Levine, and E. Maskin (1994): "The folk theorem with imperfect public information," *Econometrica*, 62(5), 997–1039.

Fudenberg, D., and E. Maskin (1986): "The folk theorem in repeated games with discounting or with incomplete information," *Econometrica*, 53(3), 533–554.

Hörner, J., and W. Olszewski (2006): "The folk theorem for games with private almost-perfect monitoring," *Econometrica*, 74(6), 1499–1544.

——— (2009): "How robust is the folk theorem?," *The Quarterly Journal of Economics*, 124(4), 1773–1814.

Kandori, M. (2011): "Weakly belief-free equilibria in repeated games with private monitoring," *Econometrica*, 79(3), 877–892.

Kandori, M., and H. Matsushima (1998): "Private observation, communication and collusion," *Econometrica*, 66(3), 627–652.

Kandori, M., and I. Obara (2006): "Efficiency in repeated games revisited: The role of private strategies," *Econometrica*, 74(2), 499–519.

——— (2010): "Towards a belief-based theory of repeated games with private monitoring: An application of POMDP," *mimeo.*

Lehrer, E. (1990): "Nash equilibria of n-player repeated games with semi-standard information," *International Journal of Game Theory*, 19(2), 191–217.

Matsushima, H. (2004): "Repeated games with private monitoring: Two players," *Econometrica*, 72(3), 823–852.

Miyagawa, E., Y. Miyahara, and T. Sekiguchi (2008): "The folk theorem for repeated games with observation costs," *Journal of Economic Theory*, 139(1), 192–221.

Obara, I. (2009): "Folk theorem with communication," *Journal of Economic Theory*, 144(1), 120–134.

Phelan, C., and A. Skrzypacz (2012): "Beliefs and private monitoring," *The Review of Economic Studies*.

Piccione, M. (2002): "The repeated prisoner's dilemma with imperfect private monitoring," *Journal of Economic Theory*, 102(1), 70–83.

Radner, R., R. Myerson, and E. Maskin (1986): "An example of a repeated partnership game with discounting and with uniformly inefficient equilibria," *Review of Economic Studies*, 53(1), 59–69.

Sekiguchi, T. (1997): "Efficiency in repeated prisoner's dilemma with private monitoring," *Journal of Economic Theory*, 76(2), 345–361.

Sugaya, T. (2012a): "belief-free review-strategy equilibrium without conditional independence," *mimeo*.

——— (2012b): "Folk theorem in repeated games with private monitoring," *mimeo*.

——— (2012c): "Folk theorem in repeated games with private monitoring: multiple players," *mimeo*.

——— (2012d): "Folk theorem in repeated games with private monitoring: two players," *mimeo*.

Takahashi, S. (2010): "Community enforcement when players observe partners' past play," *Journal of Economic Theory*, 145(1), 42–62.

Yamamoto, Y. (2007): "Efficiency results in N player games with imperfect private monitoring," *Journal of Economic Theory*, 135(1), 382–413.

——— (2009): "A limit characterization of belief-free equilibrium payoffs in repeated games," *Journal of Economic Theory*, 144(2), 802–824.

——— (2012): "Characterizing belief-free review-strategy equilibrium payoffs under conditional independence," *mimeo*.