Columbia University

Department of Economics
Discussion Paper Series

The Folk Theorem for Repeated Games with Observation Costs

Eiichi Miyagawa
Yasuyuki Miyahara
Tadashi Sekiguchi

Discussion Paper No.: 0405-12

Department of Economics
Columbia University
New York, NY 10027

December 2004

The Folk Theorem for Repeated Games with Observation Costs*

Eiichi Miyagawa† Columbia University
Yasuyuki Miyahara‡ Kobe University
Tadashi Sekiguchi§ Kyoto University

November 16, 2004

Abstract

This paper studies repeated games with private monitoring where players make optimal decisions with respect to costly monitoring activities, just as they do with respect to stage-game actions. We consider the case where each player can observe other players' current-period actions accurately only if he incurs a certain level of disutility. In every period, players decide whether to monitor other players and whom to monitor. We show that the folk theorem holds for any finite stage game that satisfies the standard full dimensionality condition and for any level of observation costs. The theorem also holds under general structures of costless private signals and does not require explicit communication among the players. Therefore, tacit collusion can attain efficient outcomes in general repeated games with private monitoring if perfect private monitoring is merely feasible, however costly it may be.

JEL Classification: C72, C73, D43, D82, K21, L13, L40.

Keywords: Repeated games, private monitoring, costly monitoring, tacit collusion, communication, secret price cuts.

*We are grateful to Masaki Aoyagi, Kyle Bagwell, Olivier Compte, Prajit Dutta, Nicolas Sahuguet, Tomas Sjöström, Jun Xue, and seminar participants at Penn State University and University of Texas at Austin for their helpful comments. We gratefully acknowledge financial supports from the Zengin Foundation for Studies on Economics and Finance, and Sekiguchi also thanks supports from the Grant-in-Aids for Scientific Research (14730014) and the 21st Century COE Research.

†Department of Economics, Columbia University, 420 West 118th Street, New York, NY 10027, U.S.A.; em437@columbia.edu.

‡Graduate School of Business Administration, Kobe University, 2-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan; miyahara@kobe-u.ac.jp.

§Institute of Economic Research, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan; sekiguchi@kier.kyoto-u.ac.jp.

1 Introduction

The theory of infinitely repeated games has demonstrated that a group of agents with long-term relationships can sustain a large set of outcomes that cannot be sustained in static situations. A major result in this literature is the folk theorem, which states that any feasible and individually rational payoff vector can be sustained if players are sufficiently patient. Since the folk theorem has been proved for virtually all stage games, it provides a general insight. On the other hand, whether the theorem holds depends critically on what players know about each other's past actions. A seminal version of the folk theorem by Fudenberg and Maskin (1986) assumes *perfect monitoring*: players obtain accurate information about other players' past actions. The result has been extended to the case of *imperfect public monitoring*, where all players receive the same noisy information (Abreu, Pearce, and Stacchetti, 1990; Fudenberg, Levine, and Maskin, 1994; Fudenberg and Levine, 1994).[1] More recent studies, which will be reviewed below, deal with the case of imperfect *private* monitoring, where players may receive different noisy information, but extending the folk theorem to this case has been difficult.

While the literature has examined various information structures, these studies share a common assumption: players have no choice of the quality of information they obtain. Players' monitoring activities are determined exogenously and players do not deal with any trade-off between the quality of information and costs of information acquisition. The goal of this paper is to study whether the folk theorem holds if players are confronted with the trade-off and make optimal decisions with respect to monitoring, just as they do for their stage-game actions.

More specifically, we consider the case where each player can obtain precise information about the other players' current-period actions if he pays a certain level of utility costs, referred to as *observation costs*. If a player does not pay the observation costs, he only receives a noisy private signal of the other players' actions. An economic example is repeated Bertrand competition where each firm chooses a price and learns the realized level of its own sales, and it is possible but costly to observe other firms' prices.

We also assume that monitoring activities can be done stealthily: an agent's monitoring activities are unobservable and do not even generate noisy signals to other players. This assumption makes it difficult to create incentives to carry out costly monitoring, as we will discuss below. This assumption also implies that what players observe from monitoring activities are their private information. Since all information is acquired privately, our class of repeated games belongs to that with private monitoring.

[1]This is the case considered in the influential paper by Green and Porter (1984), where all firms observe the market price, which is only a noisy indicator of quantities chosen by firms.

The presence of observation costs implies that a player monitors other players only if the benefit of monitoring is expected to exceed its costs. Since we assume that monitoring activities are unobservable and do not affect other players' future actions directly, the benefit of monitoring is purely the value of additional information that the player obtains from monitoring. For example, if a player is expected to play a certain action with certainty, other players expect to gain nothing from observing the player's action.

Because of this feature, it is not straightforward to extend existing constructions of equilibrium strategies to our class of repeated games. As an illustration, consider a grim-trigger strategy, under which players switch to the repetition of a static Nash equilibrium if they observe a deviation. Under the strategy, since players are believed to cooperate in the first period, the previous argument implies that players have no incentive to do costly monitoring in the first period. Therefore, deviations in the first period are not observed, and the only way to deter deviations in this period is to start punishments on the basis of their free private signals. However, since the signals are private information, punishments cannot be coordinated and it is difficult to construct an equilibrium along this line without any observation activity in general environments, as we know from the literature of imperfect private monitoring. Furthermore, we can take advantage of players' monitoring abilities by using strategies that induce monitoring activities.

To construct equilibria that induce monitoring activities, we use strategies in which players randomize. However, randomization alone does not solve the problem. As an illustration, consider a repeated prisoners' dilemma and suppose that an equilibrium strategy is such that each player puts a positive probability on defection at any history (as in Ely and Välimäki (2002)). In such an equilibrium, even though players may be randomizing, no monitoring takes place at all. Indeed, there is no gain from monitoring since defection is guaranteed to be an optimal action at any history (since a positive probability is assigned) and monitoring decisions have no direct influence on the other players' future actions. This argument shows that, to induce monitoring, the equilibrium strategy has to be such that one's optimal action depends on others' past actions. For prisoners' dilemma, this means that there needs to be a history at which a player strictly prefers to cooperate given the continuation strategy of other players. But, if a player cooperates with certainty and it is known, other players will not monitor the player and the problem discussed above persists.

Nevertheless, we show that the folk theorem does hold in our class of repeated games. A few features of the folk theorem are as follows. First, the (minmax) folk theorem holds for all n-player finite stage games that satisfy the standard full dimensionality condition. Second, the theorem holds under a rather weak assumption on the probability structure of free noisy private signals. Specifically, the assumption requires that there be no one whose action has no influence at all over the other players' private signals. That is, the action of each player should have a non-zero influence on the probability distribution of at least another player's private signal under at least one action profile. The nature of the influence is immaterial. Third, the folk theorem holds for *any* level of observation costs. That is, payoff vectors arbitrarily close to the Pareto frontier can be supported even if observation costs are arbitrarily large. This is the case since there exist equilibria in which players monitor each other only periodically and the expected per-period monitoring cost is small. The level of observation cost does matter, of course, since it affects the critical level of discount factor in the folk theorem.

There are a few papers that study repeated games with observation costs. The first study on this class of repeated games is Ben-Porath and Kahneman (2003), who prove the folk theorem under the assumption that players can communicate explicitly along the repeated game.[2] Communication is indeed possible in many situations, but there are important situations in which communication is prohibited (e.g., by antitrust laws). Thus it is worth examining what can be attained without communication, and our result shows that communication is in fact not necessary for the folk theorem.[3] Our earlier paper (Miyagawa, Miyahara, and Sekiguchi, 2003) studies the same class of repeated games without communication but deals only with the case when observation costs are sufficiently small.[4]

The literature has found difficulty with analyzing repeated games with imperfect private monitoring, since it is difficult to design punishments that can be initiated in a coordinated fashion. Early positive results in this context are limited to the prisoners' dilemma with almost perfect monitoring (e.g., Sekiguchi, 1997; Bhaskar and Obara, 2002; Piccione, 2002; Ely and Välimäki, 2002). Recent studies obtain more general results, but they are still limited to either (a) the prisoners' dilemma and its variants (Matsushima, 2004; Yamamoto, 2003), (b) the case of almost perfect monitoring (Mailath and Morris, 2002; Hörner and Olszewski, 2004), or (c) a subclass of equilibria that is not large enough to generate a general folk theorem (Ely, Hörner, and Olszewski, 2003).

Given the difficulty, some papers, since Matsushima (1991), consider the case when the players can communicate explicitly to exchange their private information (see Ben-Porath and Kahneman, 1996, 2003; Kandori and Matsushima, 1998; Compte, 1998;

[2]That is, after each period, each player can announce a message publicly.

[3]There are also a few technical differences. First, while we assume that players make monitoring decisions after choosing actions and observing public randomization, Ben-Porath and Kahneman consider the case where players choose actions and monitoring decisions at the same time. Second, while our result requires free noisy signals, Ben-Porath and Kahneman's holds without any free noisy signal.

[4]Another related paper is Ahn and Suominen (2001). In a model of random matching, they assume that players can invest in their monitoring ability, which determines the probability that they see their neighbors' actions. The major difference is that in their paper the investment decisions are made only at the beginning of the repeated game, while we assume that players make an observation decision every period.

Aoyagi, 2002). Communication has been found to greatly facilitate the analysis of repeated games with private monitoring since it allows players to coordinate their continuation actions. Indeed, Compte (1998) and Kandori and Matsushima (1998) prove folk theorems with communication for general stage games under certain conditions on private signals.

The present paper contributes to this literature by showing that if it is *feasible* to observe other players' actions without error, i.e., if the cost of perfect (private) monitoring is finite, then the folk theorem holds without communication in general repeated games with private monitoring. That is, by modeling players' monitoring decisions explicitly, we find that a sufficient condition for the folk theorem is that perfect monitoring is an *option* for each player, and it does not matter how costly the option is. The availability of this option allows players to coordinate their continuation actions when they need, and the cost of the option is not important since the option needs to be executed only infrequently.

In this sense, costly perfect monitoring can effectively replace explicit communication as a coordination device. Indeed, the feasibility of perfect monitoring enables players to communicate implicitly using their stage-game actions. However, implicit communication via stage-game actions has a few distinguishing features. First, it is not "cheap talk" since it is costly for both senders and receivers. Second, since receiving information is not only costly but optional, implicit communication works only if players have incentives to receive information. Third, once a player deviates by not receiving information (i.e., not observing other players), then he loses track of other players' continuation actions and this fact is not noticed by other players. Therefore the continuation play is no longer an equilibrium, which makes it difficult to utilize dynamic programming in our class of repeated games. Fourth, in the case of communication via stage-game actions, the message space is constrained by the action set, which may contain only two actions. Finally, implicit communication via stage-game actions may be safer than explicit communication in the presence of antitrust laws.

The present paper also contributes to the literature of imperfect *public* monitoring since the signal structure in our model subsumes imperfect public monitoring as a special case. For this class of repeated games, Fudenberg, Levine, and Maskin (1994) prove the folk theorem without communication under certain distinguishability assumptions on the signal structure. As Radner, Myerson, and Maskin (1986) show, there exist reasonable signal structures for which the folk theorem without communication fails. Kandori (2003) shows that, if communication is allowed, the folk theorem holds for a larger class of signal structures. The present paper shows that the folk theorem holds for an even larger class of signal structures (virtually any signal structure), even without communication, if the cost of perfect monitoring is finite. In particular, while the assumptions in Fudenberg, Levine, and Maskin (1994) and Kandori (2003) place restrictions on the numbers of signals and actions, our assumptions place no restrictions.

The remainder of the paper is organized as follows. The next section describes the model. Section 3 states the result and gives a detailed sketch of the proof. The formal proof is relegated to the Appendix. Section 4 concludes.

2 Model

We consider a repeated game, where a set of players play the same game repeatedly over periods $t = 1, 2, \ldots$. Let $N = \{1, 2, \ldots, n\}$ denote a finite set of players, where $n \geq 2$, and let A_i be a finite set of actions that player i can choose in each period, where $|A_i| \geq 2$. Let $A \equiv A_1 \times \cdots \times A_n$ denote the set of action profiles.

Given a set K, let $\Delta(K)$ denote the set of probability distributions over K. Thus $\mathcal{A}_i \equiv \Delta(A_i)$ denotes the set of mixed actions of player i, $\mathcal{A} \equiv \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$ denotes the set of mixed action profiles, and $\Delta(A)$ denotes the set of correlated action profiles.

At each period, after all players choose actions, each player i observes a signal ω_i costlessly and privately. The set of signals that player i might receive is given by a finite set Ω_i. A signal profile $\omega = (\omega_1, \ldots, \omega_n) \in \Omega_1 \times \cdots \times \Omega_n$ is realized with probability $P(\omega \,|\, a)$ given an action profile a. Let $P_i(\omega_i \,|\, a)$ denote the marginal distribution of ω_i given a. We assume the following on $P_i(\,\cdot\,|\,\cdot\,)$.

Assumption 1. For all $i \in N$, all $\omega_i \in \Omega_i$, and all $a \in A$,

$$P_i(\omega_i \,|\, a) > 0.$$

Assumption 2. There exists no player $i \in N$ such that for all pairs $\{a_i^1, a_i^2\} \subseteq A_i$, all $a_{-i} \in A_{-i}$, and all $r \in N \setminus \{i\}$,

$$P_r(\,\cdot\,|\, a_i^1, a_{-i}) = P_r(\,\cdot\,|\, a_i^2, a_{-i}).$$

Assumption 1 states that any $\omega_i \in \Omega_i$ is realized with a positive probability given any action profile. Since the full-support condition is required only for individual signal spaces, there may exist some (ω, a) such that $P(\omega \,|\, a) = 0$. Assumption 2 states that there exists no player who has no influence at all on any other player's signal. This means that for each player, there exists a pair of actions that induce different probability distributions of at least one player's signal under at least one action profile. It should be noted that Assumptions 1 and 2 allow the case of *public* signals, the case in which $\omega_1 = \cdots = \omega_n$ always holds.

The stage-game payoff for player i is given by $\pi_i(a_i, \omega_i)$, which depends on his own action a_i and the realized private signal ω_i. Since the payoff depends on what the player already knows, it gives no additional information (about the other players' actions or signals).[5] One special case is when the realized stage-game payoff is the sole information contained in the free signal, in which case the function $\pi_i(a_i, \cdot)\colon \Omega_i \to \mathbb{R}$ is one-to-one for each a_i.

Given an action profile $a \in A$, the expected stage-game payoff for player i is

$$u_i(a) \equiv \sum_{\omega_i \in \Omega_i} \pi_i(a_i, \omega_i) P_i(\omega_i \mid a).$$

We write $u(a) = (u_i(a))_{i \in N}$. For a mixed action profile $\alpha \in \mathcal{A}$, we abuse notation and write $u(\alpha) = (u_i(\alpha))_{i \in N}$ to denote the expected payoff profile under α. Similarly, for a correlated action profile $\rho \in \Delta(A)$, we write $u_i(\rho) = \sum_{a \in A} \rho(a) u_i(a)$ and $u(\rho) = (u_i(\rho))_{i \in N}$.

Monitoring activities take place at the end of each period. After all players choose actions and receive signals, each player chooses the set of players to observe. Let $\lambda_i\colon 2^{N \setminus \{i\}} \to \mathbb{R}_+$ be the observation cost function for player i. If player i chooses $J \subseteq N \setminus \{i\}$, he incurs observation costs $\lambda_i(J)$ and obtains completely accurate information about the realized action profile $(a_j)_{j \in J}$ in the present period. We assume that $\lambda_i(\emptyset) = 0$, $\lambda_i(J) \geq 0$ for all J, and $\lambda_i(J) \leq \lambda_i(J')$ if $J \subseteq J'$.[6]

We assume that what players observe from their observation activities are their private information. To make our problem more difficult, we also assume that observation activities are completely stealthy. This means first that observation activities are not observable at all to other players. That is, whether player i observes another player j in a given period (let alone what i observes) is unobservable to any player $k \neq i$ even if k observes i in the period (even if $k = j$). Second, players do not even receive any noisy information about other players' observation activities. These assumptions imply that one's observation decision itself does not affect other players' future actions at all and therefore deviations with respect to monitoring cannot be punished directly. This feature makes it difficult to create monitoring incentives.

We now turn to the definition of repeated-game strategy. We assume that there exists a public randomization device (e.g., public lotteries, last digits of the Dow Jones, etc), which generates a sequence of independent random variables $(X_1, Y_1, X_2, Y_2, X_3, \dots)$ that are all uniformly distributed over $[0, 1]$. Random variable X_t $(t = 1, 2, \dots)$ is realized at the beginning of period t before players choose actions, while Y_t $(t = 1, 2, \dots)$ is realized in the middle of period t just before players make monitoring decisions. The realizations of the random variables ("sunspots") are irrelevant to payoffs and observable publicly and costlessly.

[5] If the realized payoff does give additional information, we can redefine signals to include payoff information. That is, we can redefine a signal as a pair (ω_i, π_i) of the original signal and the realized payoff. If this pair satisfies the full-support condition, Assumption 1 is preserved and the payoff gives no additional information.

[6] This monotonicity condition is assumed only for simplicity and can be dispensed with. See the remarks at the end of Section 3.

The sequence of events within a given period t is given as follows. First, players observe the realization of public random variable X_t. Second, players simultaneously choose an action $a_i \in A_i$. Third, each player i observes a signal ω_i privately, which determines $\pi_i(a_i, \omega_i)$. Fourth, players observe the realization of the middle-of-period public random variable Y_t. Fifth, each player i chooses whom to monitor, $J_i \subseteq N \setminus \{i\}$. Finally, i observes the realized action profile $(a_j)_{j \in J_i}$ and incurs a disutility of $\lambda_i(J_i)$.

Player i's (private) history at the beginning of period $t \geq 2$ is a sequence of realizations of public random variables, his own actions, realizations of his private signals, and his observations about the other players' actions, all up to (including) period $t-1$. Formally, it is a sequence

$$\begin{aligned} h_i^t &= [x_k, a_{i,k}, \omega_{i,k}, y_k, (a_{j,k})_{j \neq i}]_{k=1}^{t-1} \\ &\in \Big[[0,1] \times A_i \times \Omega_i \times [0,1] \times \prod_{j \in N \setminus \{i\}} (A_j \cup \{\phi\})\Big]^{t-1}. \end{aligned}$$

In this sequence, $x_k \in [0,1]$ is the realized value of random variable X_k, $a_{i,k} \in A_i$ is player i's action in period k, $\omega_{i,k} \in \Omega_i$ is the realized private signal of i in period k, $y_k \in [0,1]$ is the realized value of random variable Y_k, and $a_{j,k} \in A_j \cup \{\phi\}$ is i's observation about player j's action in period k, where $a_{j,k} = \phi$ means that i did not observe j in period k and therefore has no observation.

For all $t = 1, 2, \ldots$, let H_i^t denote the set of all (private) histories for player i at period t (H_i^1 is an arbitrary singleton). A strategy of player i is a pair of functions $\sigma_i = (\sigma_i^a, \sigma_i^m)$ such that

$$\begin{aligned} \sigma_i^a &: \bigcup_{t=1}^{\infty} (H_i^t \times [0,1]) \to \Delta(A_i), \\ \sigma_i^m &: \bigcup_{t=1}^{\infty} (H_i^t \times [0,1] \times A_i \times \Omega_i \times [0,1]) \to \Delta(2^{N \setminus \{i\}}). \end{aligned}$$

A strategy profile $\sigma = (\sigma_1, \ldots, \sigma_n)$ generates a probability distribution over sequences $(a_t, (J_{i,t})_{i \in N})_{t=1}^{\infty}$, where $a_t \in A$ is the action profile in period t and $J_{i,t} \subseteq N \setminus \{i\}$ is the set of players that i observes in period t. Given the sequence, player i's overall payoff is

$$(1-\delta) \sum_{t=1}^{\infty} \delta^{t-1} \big[u_i(a_t) - \lambda_i(J_{i,t})\big],$$

where $\delta \in (0,1)$ is a discount factor common to all players. Players maximize the expected overall payoff. We are interested in sequential equilibria of the repeated game when the discount factor is close to one.

3 Result

Player i's minmax payoff is defined by

$$\underline{u}_i \equiv \min_{\alpha_{-i} \in \mathcal{A}_{-i}} \max_{a_i \in A_i} u_i(a_i, \alpha_{-i}),$$

where $\mathcal{A}_{-i} \equiv \prod_{j \neq i} \mathcal{A}_j$. Let

$$V \equiv \text{convex hull of } \{u(a) : a \in A\},$$
$$V^* \equiv \{v \in V : v_i \geq \underline{u}_i \text{ for all } i \in N\}.$$

Note that $\underline{u}_i$, V, and V^* are all defined independently of the observation cost functions $\lambda_1, \ldots, \lambda_n$.

Our result is the following.

Theorem. *For any $v^* \in \operatorname{int} V^*$, there exists $\underline{\delta} \in (0,1)$ such that, for any $\delta \in [\underline{\delta}, 1)$, there exists a sequential equilibrium whose payoff profile is v^*.*[7]

Proof. See Appendix.

The proof is constructive: for a given payoff profile $v^* \in \operatorname{int} V^*$, we construct a specific strategy profile σ that is a sequential equilibrium and yields v^* if the discount factor is close to one. In the remainder of this section, we give an informal but detailed description of the strategy profile for a given target payoff profile $v^* \in \operatorname{int} V^*$ and discuss why it is an equilibrium.

To simplify the exposition, we assume, in this section, that there exists a static Nash equilibrium $\alpha^{\text{NE}} \in \mathcal{A}$ that attains the minmax values: $u(\alpha^{\text{NE}}) = \underline{u}$. This allows us to use the perpetual repetition of the Nash equilibrium as punishments. In the general case, we construct punishment phases by adapting the construction of Fudenberg and Maskin (1986); see the remarks at the end of this section.

We also assume, to simplify the following exposition, that there exists a subset $A' \subseteq A$ such that (i) for all $a \in A'$, no player plays a best response, and (ii) the convex hull of $\{u(a) : a \in A'\}$ contains an open ball around v^*. Condition (i) implies that for all $a \in A'$, each player i has a short-run better response $d_i^a \in A_i$, which defines the *minor deviation* for i when a is the cooperation action profile. For a given $a \in A'$, consider a mixed action profile α^a given by

$$\alpha_i^a \equiv (1 - \eta^a) \cdot a_i + \eta^a \cdot d_i^a \qquad \text{for all } i \in N,$$

where $\eta^a \in (0,1)$. We choose a small number for η^a so that d_i^a remains a short-run

[7] As usual, the statement is meaningful only if the standard full dimensionality condition by Fudenberg and Maskin (1986) is satisfied; namely, int $V^* \neq \emptyset$.

better response than a_i against α^a_{-i}. For the construction without the simplifying assumption, see the remarks at the end of this section.

The equilibrium play is characterized by three types of periods: *cooperation* periods, *examination* periods, and *report* periods. We begin with describing those periods and the rule that governs the transition among these periods.

Cooperation Periods. Cooperation periods are parameterized by $\rho \in \Delta(A')$ and denoted by $Coop(\rho)$. In $Coop(\rho)$, the public random variable realized at the beginning of the period chooses a pure action profile $a \in A'$ with probability $\rho(a)$ and then players play the mixed action profile α^a. So, each player i randomizes between a_i (cooperation) and d^a_i (minor deviation), placing a small probability on the minor deviation.

Players then use the middle-of-period public randomization to coordinate their monitoring decisions. Specifically, with probability $1-\mu$, where $0 < \mu < 1$, no one observes any player, and the next period continues to be $Coop(\rho)$ with the same ρ. On the other hand, with probability μ, all players observe each other. In this case, the play in the next period, say period $t+1$, is determined as follows. The public randomization at the beginning of period $t+1$ selects a player i randomly. If this player played the minor deviation (d^a_i) in period t, then the punishment stage (i.e., the repetition of the static Nash equilibrium α^{NE}) starts in period $t+1$ with probability $\xi^a_i \in (0,1)$. The probability ξ^a_i is determined so that player i is indifferent between cooperation and minor deviation, i.e., the short-run gain from the minor deviation equals the long-run loss from the possibility of proceeding to the static Nash equilibrium (the equation will be given below). On the other hand, if player i played a *major deviation* (i.e., any $a'_i \notin \{a_i, d^a_i\}$) in period t, then a punishment starts with probability one.[8] If no punishment starts, then period $t+1$ is an *examination* period.

Examination Periods. In an examination period, the public randomization at the beginning of the period selects a pair of players (j,k) such that $j \neq k$. Player k is then given a test to show that he indeed observed other players. Specifically, player k is asked to "state" whether player j cooperated or not in the previous period. To state it without using explicit communication, player k chooses a stage-game action according to a predetermined rule that associates actions with answers. Specifically, the prescribed action for player k (i.e., the right answer) is denoted by $b^{right}_k \in A_k$ and defined as follows. By Assumption 2, it can be shown that there exist a completely mixed action profile β_{-k}, a player $r \neq k$ (*referee* for k), and a partition $\{\Omega^1_r, \Omega^2_r\}$ of Ω_r, such that

$$\operatorname*{Argmax}_{b_k \in A_k} P_r(\Omega^1_r \mid b_k, \beta_{-k}) \bigcap \operatorname*{Argmax}_{b_k \in A_k} P_r(\Omega^2_r \mid b_k, \beta_{-k}) = \emptyset. \tag{1}$$

Indeed, a violation of (1) means that there exists an action b_k that maximizes and

[8] Actually, for a rather technical reason, the strategy used in the proof forgives even major deviations with a positive probability; See Appendix for details.

minimizes $P_r(\Omega_r^1 \mid b_k, \beta_{-k})$, which means that the probability is constant in b_k. If this is the case for all $r \neq k$, all Ω_r^1, and all β_{-k}, then Assumption 2 is violated.

Let those sets of maximizers in (1) be denoted B_k^1 and B_k^2, respectively. Let b_k^1 and b_k^2 be pure actions such that

$$b_k^1 \in \operatorname*{Argmax}_{b_k \in B_k^1} \; u_k(b_k, \beta_{-k}), \qquad b_k^2 \in \operatorname*{Argmax}_{b_k \in B_k^2} \; u_k(b_k, \beta_{-k}). \tag{2}$$

Then we define

$$b_k^{right} = \begin{cases} b_k^1 & \text{if } j \text{ cooperated in the previous period,} \\ b_k^2 & \text{otherwise.} \end{cases}$$

This is the action prescribed to player k.

Meanwhile, the other players $i \neq k$ play mixed actions β_{-k}. Regardless of the middle-of-period public randomization, all players observe all $i \neq k$. Since b_k^{right} is a pure action, other players have no incentive to observe k.

The state transition is determined by public randomization at the beginning of the next period. With probability 1/2, the next period is again an examination period, with a newly chosen pair (j', k') such that $j' \neq k$ and $k' \neq j'$. The constraint $j' \neq k$ is imposed since player k did not randomize and thus his realized action cannot serve as a question in the next examination.[9] With the remaining probability, the next period is a *report* period.

Before we proceed to describe the report period, let us explain what the examination period does and why we proceed to a report period. An important feature of the examination period is that, since k and j are chosen randomly, anyone who did not observe all randomizing players in the previous period may face uncertainty about which action to play. This poses a problem for the player if playing a wrong action in this period causes a severe long-run loss.

However, it is not straightforward to ensure that playing a wrong action in the examination period causes a long-run loss. The problem is that, as mentioned above, no one monitors player k in this period, since k plays a pure action b_k^{right} and, in equilibrium, all players (including k) know what b_k^{right} is. An easy fix is to make b_k^{right} a mixed action, but this is not possible if there are only two actions. Indeed, if player k has only two actions and b_k^{right} assigns a mixed action in at least one case, then one of the actions is always optimal and he does not suffer from uncertainty.

Therefore, to deter player k from deviating in the examination period, other players initiate punishments on the basis of their free noisy signals. Since k's action affects the free signal of his referee (player r) defined above, his signal can be used to judge whether or not player k played correctly and then start punishments if the signal is not

[9] In the two-person case, we necessarily have $k' = k$.

favorable. However, there remains a problem, which is that the referee's signal is his private information and does not allow players to start punishments in a coordinated fashion. This is why we proceed to a report period, in which the referee "reports" his private signal to others by means of his stage-game action, as we now describe.

Report Periods. For each player i, choose a pair of actions $\{c_i', c_i''\} \subseteq A_i$ arbitrarily in advance. Let k be the player who was under examination in the last period, and let $r \neq k$ be the referee for k (specified in (1)). Then, in this period, the referee r either "approves" or "disapproves" of player k's answer by choosing c_r' or c_r'', respectively. Specifically, the referee plays the following mixed action:

$$\begin{aligned}
&0.9 \cdot c_r' + 0.1 \cdot c_r'' && \text{if } b_k^{right} = b_k^1 \text{ and } \omega_r \in \Omega_r^1, \\
&0.9 \cdot c_r' + 0.1 \cdot c_r'' && \text{if } b_k^{right} = b_k^2 \text{ and } \omega_r \in \Omega_r^2, \\
&0.1 \cdot c_r' + 0.9 \cdot c_r'' && \text{if } b_k^{right} = b_k^1 \text{ and } \omega_r \notin \Omega_r^1, \\
&0.1 \cdot c_r' + 0.9 \cdot c_r'' && \text{if } b_k^{right} = b_k^2 \text{ and } \omega_r \notin \Omega_r^2,
\end{aligned}$$

where b_k^{right} is the prescribed pure action for player k in the last period and ω_r is the signal that the referee received in the last period. Meanwhile, all the other players $i \neq r$ (including k) randomize between c_i' and c_i'' with equal probability. All players monitor all other players.

To make sense of the referee's action, recall that since $b_k^\ell \in B_k^\ell$ $(\ell = 1, 2)$, the probability that the referee receives a signal $\omega_r \in \Omega_r^\ell$ is maximized if player k plays b_k^ℓ. If the prescribed action for player k in the examination period was b_k^ℓ and the referee received a signal $\omega_r \in \Omega_r^\ell$, then the referee basically approves of k's answer by playing c_r'. The referee actually randomizes, giving his approval only with probability 0.9. Symmetrically, if the referee receives a signal $\omega_r \notin \Omega_r^\ell$, he disapproves of k's answer with probability 0.9.

This construction of the referee's action implies that, when $b_k^{right} = b_k^\ell$ $(\ell = 1, 2)$, player k maximizes the probability of getting the referee's approval if and only if he plays an action in B_k^ℓ. Since B_k^1 and B_k^2 are disjoint, it follows that if player k is uncertain of b_k^{right}, he cannot avoid playing a wrong action, and the expected probability of getting the referee's approval is strictly smaller than in the case where he knows b_k^{right} for certain.

We put randomization in the referee's action to ensure that the other players have incentives to monitor the referee.[10] Once players observe the referee's action, they can

[10] Even if the referee does not randomize, his action appears random to other players since his action depends on his private signal. However, if the correlation of signals across players is such that the realization of certain signals ω_i for a player $i \neq r$ rules out certain signals ω_r for the referee (which indeed occurs if the signals are public, i.e., $\omega_1 = \cdots = \omega_n$), then a player may be able to infer the referee's action without costly monitoring. This introduces an unnecessary complication to the proof, which is why we introduced the randomization. The exact way in which the referee's action trembles is immaterial.

decide whether to punish player k in a coordinated fashion, as we will describe.

The state transition depends on public randomization at the beginning of the next period. With probability 1/2, the next period is again an examination period with a newly selected pair (j, k). With the remaining probability, the next period is a cooperation period with a newly selected $\rho \in \Delta(A')$. The selection of ρ is now described in detail.

Selection of ρ. The selection has two goals. The first is to offset the difference between the target payoffs v^* and the realized payoffs during the previous two periods, which are an examination period and a report period. By doing this, we can make each player's continuation value from any examination period equal to v^* regardless of the realized actions in the examination period and the subsequent report period. This in turn makes the players, except the one under examination, indifferent about their actions during those periods and willing to randomize as prescribed. The second goal of the selection rule for ρ is to punish the player under examination who did not get his referee's approval.

To give a detailed description of the selection of ρ, we begin with its first goal. That is, we look for a selection rule for $\rho \in \Delta(A')$ that makes the continuation value from any examination period equal to v^*.

By construction, if period t is a cooperation period and period $t-1$ is a report period, then period $t-2$ is an examination period. Let k be the player under examination, b_k^{right} be his prescribed action in the examination period, b_{-k}^{obs} be the observed action profile of the other players in the examination period, and c^{obs} be the observed action profile in the report period. Let $Coop_i(\rho)$ be the continuation value from a cooperation period with ρ. Then let $\rho' \in \Delta(A')$ be a distribution that satisfies the following equation for all $i \in N$:

$$\begin{aligned} v_i^* &= (1-\delta)\Big[u_i(b_k^{right}, b_{-k}^{obs}) - \lambda_i(N \setminus \{k, i\})\Big] \\ &\quad + \frac{1}{2}\delta v_i^* + \frac{1}{2}\delta(1-\delta)\Big[u_i(c^{obs}) - \lambda_i(N \setminus \{i\})\Big] \\ &\quad + \frac{1}{4}\delta^2 v_i^* + \frac{1}{4}\delta^2\, Coop_i(\rho'). \end{aligned} \tag{3}$$

If a distribution ρ' that satisfies this equation is chosen for any given $(k, b_k^{right}, b_{-k}^{obs}, c^{obs})$, then the equation implies that the continuation value from any examination period is indeed v_i^* for all players. Rearranging (3) yields

$$\begin{aligned} 0 &= \Big[v_i^* - u_i(b_k^{right}, b_{-k}^{obs}) + \lambda_i(N \setminus \{k, i\})\Big] \\ &\quad + \Big[v_i^* - u_i(c^{obs}) + \lambda_i(N \setminus \{i\})\Big]\frac{\delta}{2} \\ &\quad + \Big[v_i^* - Coop_i(\rho')\Big]\frac{\delta^2}{4(1-\delta)}. \end{aligned} \tag{4}$$

To identify ρ' that satisfies (4), we need to compute the value $Coop_i(\rho')$. Since $Coop_i(\rho')$ is the continuation value from a cooperation period, it satisfies the following equation:

$$\begin{aligned}Coop_i(\rho') = {} & (1-\delta)\sum_{a\in A'}\rho'(a)u_i(a_i,\alpha_{-i}^a) + (1-\mu)\delta\, Coop_i(\rho') \\ & - \mu(1-\delta)\lambda_i(N\setminus\{i\}) \\ & + \mu\delta\Big[\sum_{a\in A'}\rho'(a)\frac{1}{n}\sum_{j\neq i}\eta^a\xi_j^a\Big]\underline{u}_i \\ & + \mu\delta\Big[1-\sum_{a\in A'}\rho'(a)\frac{1}{n}\sum_{j\neq i}\eta^a\xi_j^a\Big]v_i^*.\end{aligned} \tag{5}$$

This equation is written based on the case where player i plays a_i for any given $a \in \operatorname{supp}(\rho')$. At equilibrium, player i is indifferent between a_i and d_i^a in the cooperation period, which is ensured if the probability ξ_i^a is chosen to satisfy

$$(1-\delta)\big[u_i(d_i^a,\alpha_{-i}^a) - u_i(a_i,\alpha_{-i}^a)\big] = \mu\delta(1/n)\xi_i^a(v_i^* - \underline{u}_i). \tag{6}$$

The left-hand side denotes the short-run gains from the minor deviation, while the right-hand side denotes the long-run losses from possible transition to the punishment stage.

Substituting (6) into (5) to eliminate ξ_j^a, we obtain the following simple expression for $Coop_i(\rho')$:

$$Coop_i(\rho') = \frac{(1-\delta)\hat{u}_i(\rho') + \mu\delta v_i^*}{1-\delta+\mu\delta} \tag{7}$$

where the payoff functions $\hat{u}_i$ are defined by

$$\begin{aligned}\hat{u}_i(a) \equiv {} & u_i(a_i,\alpha_{-i}^a) - \mu\lambda_i(N\setminus\{i\}) \\ & - \eta^a\sum_{j\neq i}\big[u_j(d_j^a,\alpha_{-j}^a) - u_j(a_j,\alpha_{-j}^a)\big]\frac{v_i^* - \underline{u}_i}{v_j^* - \underline{u}_j}.\end{aligned} \tag{8}$$

The function $\hat{u}_i$, which we referred to as the *virtual payoff function* of player i, represents his stage-game payoff in a cooperation period when a is selected as the cooperative action profile, after we take into account expected monitoring costs, small probabilities with which minor deviations are played, and expected losses from possible transition to punishments. By choosing small numbers for μ and η^a, we can make the virtual payoff function $\hat{u}_i$ arbitrarily close to the true payoff function u_i.

Substituting (7) into (4), we obtain that for all $i \in N$,

$$\begin{aligned} 0 = &\left[v_i^* - u_i(b_k^{right}, b_{-k}^{obs}) + \lambda_i(N \setminus \{k, i\})\right] \\ &+ \left[v_i^* - u_i(c^{obs}) + \lambda_i(N \setminus \{i\})\right]\frac{\delta}{2} \\ &+ \left[v_i^* - \hat{u}_i(\rho')\right]\frac{\delta^2}{4(1-\delta+\mu\delta)}. \end{aligned} \tag{9}$$

There exists $\rho' \in \Delta(A')$ that satisfies this equation for all i, if μ and η^a are close to 0 and δ is close to 1. Indeed, if δ is close to 1 and μ is close to 0, the last line in (9) can be arbitrarily large even if $\hat{u}_i(\rho')$ is close to v_i^*. Since v^* is an interior point of the convex hull of $\{u(a) : a \in A'\}$ and $\hat{u}_i$ is close to u_i, a distribution $\rho' \in \Delta(A')$ that yields the equality for all i exists regardless of the values of $(k, b_k^{right}, b_{-k}^{obs}, c^{obs})$.

The preceding argument shows that if ρ' is always chosen to satisfy (9), then the continuation value from any examination period equals v_i^* for all players. Further, (3) shows that this continuation value is the same regardless of the randomizing players' realized actions in the examination and report periods (i.e., (b_{-k}^{obs}, c^{obs})). This implies that all these randomizing players are completely indifferent about their actions.

The choice of ρ' that satisfies (9), however, does not give a right incentive to player k (i.e., the player under examination) in the examination period, since (9) depends only on what player k is prescribed to do (i.e., b_k^{right}) and not what he does. This is where our second goal comes in. To deal with player k's incentive, we modify ρ' slightly to punish or reward player k depending on the "report" of the referee. Specifically, let AP_k be defined by

$$AP_k = \begin{cases} 0.9P_r(\Omega_r^1 \mid b_k^1, \beta_{-k}) + 0.1P_r(\Omega_r^2 \mid b_k^1, \beta_{-k}) & \text{if } b_k^{right} = b_k^1, \\ 0.9P_r(\Omega_r^2 \mid b_k^2, \beta_{-k}) + 0.1P_r(\Omega_r^1 \mid b_k^2, \beta_{-k}) & \text{if } b_k^{right} = b_k^2. \end{cases}$$

This is the expected probability that player k earns his referee's approval given that k plays as prescribed. Let $\varepsilon > 0$ be a small number. Then, finally, let $\rho \in \Delta(A')$ be such that

$$\hat{u}_i(\rho) = \hat{u}_i(\rho') \qquad \text{for all } i \neq k, \tag{10}$$

$$\hat{u}_k(\rho) = \begin{cases} \hat{u}_k(\rho') + \varepsilon(1 - AP_k) & \text{if } c_r^{obs} = c_r', \\ \hat{u}_k(\rho') - \varepsilon AP_k & \text{otherwise.} \end{cases} \tag{11}$$

This is the ρ that is used in the new cooperation period. If $\varepsilon > 0$ is sufficiently small, there exists $\rho \in \Delta(A')$ that satisfies these equations. The equations mean that player k receives a "bonus" of $\varepsilon(1 - AP_k)$ if he earns his referee's approval, and pays a "penalty" of εAP_k otherwise. Since the referee's approval is given with probability

AP_k in equilibrium, the expected net bonus is zero, and hence the continuation value from an examination period remains unchanged and equal to v_k^*.

Initial Play. The initial period is set as a cooperation period $Coop(\rho^*)$ where $\rho^* \in \Delta(A')$ is chosen to satisfy

$$\hat{u}(\rho^*) = v^*. \tag{12}$$

Such a ρ^* exists if μ and η^a are close to 0 and δ is close to 1. Then (7) implies that v^* is indeed the payoff profile for the entire repeated game under the strategy profile.

Incentives. We now discuss why players have incentives to follow the strategy described above. First, as discussed above, players have incentives to randomize between cooperation and minor deviation in cooperation periods since the probabilities ξ_i^a of punishing minor deviants (see (6)) are determined precisely to make minor deviation indifferent to cooperation. Second, in report periods, all players (including the referee) are completely indifferent about their actions, because of the way ρ is chosen in the subsequent cooperation period. Similarly, in examination periods, all players except the one under examination are indifferent about their actions. It remains to verify the incentive of the player under examination in examination periods and the incentive for each player to observe other players when required.

We first discuss the incentive of the player under examination (player k) to play the prescribed action b_k^{right} in an examination period. Let $\ell \in \{1,2\}$ be such that $b_k^\ell = b_k^{right}$. Since b_k^ℓ is a short-run optimal choice within B_k^ℓ (see (2)), it suffices to check whether he has an incentive to play an action $b_k \notin B_k^\ell$. The short-run gain from playing an action $b_k \notin B_k^\ell$ is at most

$$(1-\delta) \max_{a,a' \in A} |u_k(a) - u_k(a')|.$$

On the other hand, playing $b_k \notin B_k^\ell$ necessarily lowers the probability that the referee receives a signal in Ω_r^ℓ by some $L > 0$ (see (1)). If the referee does not receive a signal in Ω_r^ℓ and if the next period is a report period (which occurs with probability 1/2), then the probability that the referee chooses c_r' (i.e., approves of k's answer) goes down from 0.9 to 0.1. This has three effects on player k's payoffs. First, there is a direct effect on k's stage-game payoff in the report period. Second, there is an effect on ρ' since ρ' depends on c_r^{obs} (see (3)). However, (3) implies that these two effects are canceled out by each other. Finally, there is an effect on ρ through the last term of (11). If the referee gives his disapproval (playing c_r''), then $\hat{u}_k(\rho)$ goes down by ε, which means, by (7), that $Coop_k(\rho)$ goes down by $(1-\delta)\varepsilon/(1-\delta+\mu\delta)$. This effect matters if the report period is followed immediately by a cooperation period, which occurs with probability 1/2. Altogether, the long-run loss from playing $b_k \notin B_k^\ell$ is at least

$$L\frac{1}{2}(0.9-0.1)\frac{1}{2}\delta^2\frac{(1-\delta)\varepsilon}{1-\delta+\mu\delta}.$$

A sufficient condition for this to exceed the short-run gain is

$$\max_{a,a'\in A}|u_k(a)-u_k(a')|<\frac{0.2L\delta^2\varepsilon}{1-\delta+\mu\delta}.$$

This is satisfied if δ is close to 1 and μ is close to 0. Recall that μ is the probability that observation is prescribed in cooperation periods, which also determines how long a cooperative phase with the same ρ is expected to continue. If μ is small, a cooperative phase with the same ρ is expected to continue for a large number of periods and therefore a slight mofidication of ρ has a significant long-run effect. This gives player k strong incentives to answer correctly in the examination period since doing so maximizes the probability that ρ is modified favorably.

We now turn to the incentive to observe other players. Suppose that player i did not observe player $j \neq i$ when he was prescribed to. Since η^a and $(\xi_h^a)_{h\in N}$ are small, it follows that with a probability $p \geq 1/2$, the next period is an examination period. In the examination period, player i is chosen to be examined with probability $1/n$. Furthermore, at least with probability $1/(n-1)$, the action b_i^{right} depends on player j's previous action. In this (worst) case, player i is uncertain of b_i^{right}. Therefore, there is a positive probability that his action turns out to be wrong: $b_i \notin B_i^\ell$ when $b_i^{right} = b_i^\ell$ $(\ell \in \{1,2\})$. This probability is bounded below by some $F>0$. As before, if this happens, the probability that his referee receives a signal in Ω_r^ℓ goes down by at least $L>0$. Repeating the previous argument, we conclude that the long-run loss to player i is at least

$$\frac{1}{2n(n-1)}FL\frac{1}{2}(0.9-0.1)\frac{1}{2}\delta^3\frac{(1-\delta)\varepsilon}{1-\delta+\mu\delta}.$$

A sufficient condition for this to exceed the short-run gain from not monitoring j is that

$$\lambda_i(S\cup\{j\})-\lambda_i(S)<\frac{0.1FL\delta^3\varepsilon}{n(n-1)(1-\delta+\mu\delta)}$$

holds for all $S\subseteq N\setminus\{i,j\}$. This is the case if δ is close to 1 and μ is close to 0.

Remarks. The above exposition assumes that there is a rich set A' of action profiles where no one is playing a best response. In the general construction given in Appendix, we first show that for any pure action profile $a\in A$, there exists a mixed action profile α^a where (i) each player i for whom a_i is not a best response to α^a randomizes between a_i and one of the better responses $d_i^a\in A_i$, playing the better response with a small probability, and (ii) each player i for whom a_i is a best response to α^a plays it with probability one. Given this, any $a \in A$ can be used as a cooperation action profile. When a pure action profile a is chosen by ρ, players play the mixed action profile α^a, and if observation is prescribed, players observe only those who randomize in α^a.

Another simplifying assumption made in the above exposition is that the reversion to a static Nash equilibrium can be used as punishments. In the general construction in Appendix, we adapt the punishment scheme of Fudenberg and Maskin (1986) to our setting. Specifically, we introduce *minmax* periods, in which players play a perturbed minmax action profile where all players randomize except for the minmaxed player. The perturbation is introduced to create monitoring incentives. As in cooperation periods, observation is prescribed with a positive probability, in which case the play returns to an examination period. However, in the continuation strategy, the target payoff profile v^* (and the selection rule for ρ) is modified for the following two objectives. One is to punish the minmaxed player and reward the other players. The other is to make the minmaxing players indifferent about their actions in the minmax period, which is possible since these players are observed when the play exits from the minmax period.

We also assumed that the observation cost functions are monotonic: $\lambda_i(J) \leq \lambda_i(J')$ if $J \subseteq J'$. However, non-monotonic cost functions (which are relevant if it is easier to monitor multiple players at the same time) can be accommodated easily. It suffices to modify the strategy so that, if J is the set of randomizing players in the current period and monitoring is instructed, player i monitors J' that solves $\min_{J' \supseteq J} \lambda_i(J')$ and then ignores any observed deviation by players in $J' \setminus J$.

Finally, we note that the set V^* is only a subset of feasible and individually rational payoff vectors for our class of repeated games. To see this, note that the set of feasible payoff vectors in our context is

$$\bar{V} \equiv \{(v_i - p_i \lambda_i(N \setminus \{i\}))_{i \in N} : v \in V \text{ and } p \in [0,1]^N\},$$

which is a superset of V since V deals only with the case where $p_i = 0$ for all i. While any $v \in V$ is feasible, players can also decrease their payoffs by paying observation costs, and the reduced payoff vector may not be in V. Our proof relies on a strategy profile that works only if the frequency of monitoring is close to zero, and it is not straightforward to modify the strategy to accommodate payoff profiles in $\bar{V} \setminus V$.[11]

[11]Another reason for the difference between V^* and the set of feasible and individually rational payoff vectors is that the minmax value $\underline{u}_i$ is defined under the assumption that the other players randomize independently. Since actions and signals are private information, the other players can actually make their actions appear correlated to the player being punished, and the minmax value in correlated actions may be lower than that in mixed actions. For the idea of using private signals to induce correlations in repeated games, see Lehrer (1991).

4 Conclusion

This paper extends the folk theorem to repeated games in which monitoring activities are modeled explicitly and players make monitoring decisions optimally as they do for their stage-game actions. Players face the trade-off between the quality of information and the cost of information acquisition. Equilibrium strategies constructed in the standard theory do not work in this context since they are not designed to motivate players to monitor each other.

The present paper finds that, for the standard folk theorem to extend to the generalized class of repeated games, it suffices that the cost of observing other players' actions without error is finite. According to this result, the folk theorem under perfect monitoring extends, with virtually no change, as long as perfect monitoring is an option for each player, even if it is a very costly option. In particular, patient players can attain any payoff vector that is arbitrarily close to the Pareto frontier, and therefore the presence of monitoring costs causes no efficiency loss in the case of extreme patience.

Since the model allows all monitoring activities to be done privately, the present paper contributes to the literature of repeated games with private monitoring by giving a folk theorem without explicit communication for general stage games and general structures of private signals.

A Appendix: Proof

A.1 Preliminaries

We use the sup metric for Euclidean spaces: for all $v, w \in \mathbb{R}^\ell$, $\|v-w\| \equiv \max_{i\in\{1,\ldots,\ell\}}|v_i - w_i|$. For all $\varepsilon > 0$ and all $v \in \mathbb{R}^N$, let $\bar{N}_\varepsilon(v)$ denote the closed ε-neighborhood of v.

Let $v^* \in \operatorname{int} V^*$ be an arbitrarily chosen target payoff profile. Then there exists $\varepsilon > 0$ such that

$$\bar{N}_{4\varepsilon}(v^*) \subseteq V^*, \tag{13}$$

$$3\varepsilon < \min_{i\in N}\big[v_i^* - \underline{u}_i\big]. \tag{14}$$

Let $D \in \mathbb{R}$ be defined by

$$D \equiv \max_{i\in N}\Big[\lambda_i(N \setminus \{i\}) + \max_{a,a'\in A}|u_i(a) - u_i(a')|\Big] > 0.$$

Then there exist $q > 0$ and $\bar{\eta} \in (0, 1/2)$ such that

$$qD < \varepsilon(1-q), \tag{15}$$

$$2\bar{\eta}D(n-1)(1+\frac{D}{q\varepsilon}) < \varepsilon. \tag{16}$$

The following lemma defines a mixed action profile α^a for each $a \in A$. In the equilibrium we will construct, players play this mixed action profile in the cooperative stage when a is chosen as the cooperative action profile.

Lemma. *For all $a \in A$, there exists a mixed action profile α^a such that for all $i \in N$:*

(i) Either $\alpha_i^a = a_i$ or

$$\alpha_i^a = (1-\eta_i^a)\cdot a_i + \eta_i^a \cdot d_i^a \tag{17}$$

where $d_i^a \neq a_i$ and $0 < \eta_i^a \leq \bar{\eta}$.

(ii) If $\alpha_i^a = a_i$, then a_i is a best response to α_{-i}^a.

(iii) If $\alpha_i^a \neq a_i$, i.e., α_i^a is given by (17), then $u_i(d_i^a, \alpha_{-i}^a) > u_i(a_i, \alpha_{-i}^a)$.

Proof. Fix a pure action profile $a \in A$. For a given mixed action profile $\alpha \in \mathcal{A}$, define a set $Dev(\alpha)$ by

$$Dev(\alpha) \equiv \{i \in N : \max_{a_i'\in A_i} u_i(a_i', \alpha_{-i}) > u_i(a_i, \alpha_{-i})\}.$$

This is the set of players for whom a_i is not a best response to α_{-i}. Let $D^0 \equiv Dev(a)$. If $D^0 = \emptyset$, then we are done by setting $\alpha^a = a$. So suppose $D^0 \neq \emptyset$. Then for all $i \in D^0$, there exists an action $d_i^a \in A_i$ such that $u_i(d_i^a, a_{-i}) > u_i(a)$. Thus for any mixed action profile $\hat{\alpha}$ such that $\|\hat{\alpha} - a\|$ is sufficiently small, we have $u_i(d_i^a, \hat{\alpha}_{-i}) > u_i(a_i, \hat{\alpha}_{-i})$ for all $i \in D^0$. Hence, there exists $\eta_i^a \in (0, \bar{\eta}]$ for all $i \in D^0$, such that the mixed action profile α^1 defined by

$$\alpha_i^1 \equiv \begin{cases} (1-\eta_i^a) \cdot a_i + \eta_i^a \cdot d_i^a & \text{if } i \in D^0 \\ a_i & \text{if } i \notin D^0 \end{cases}$$

satisfies $u_i(d_i^a, \alpha_{-i}^1) > u_i(a_i, \alpha_{-i}^1)$ for all $i \in D^0$. Thus $D^0 \subseteq Dev(\alpha^1) \equiv D^1$. If $D^1 = D^0$, we are done by setting $\alpha^a = \alpha^1$. So suppose otherwise. Then for all $i \in D^1 \setminus D^0$, there exists an action $d_i^a \in A_i$ such that $u_i(d_i^a, \alpha_{-i}^1) > u_i(a_i, \alpha_{-i}^1)$. Therefore, for any mixed action profile $\hat{\alpha}$ such that $\|\hat{\alpha} - \alpha^1\|$ is sufficiently small, $u_i(d_i^a, \hat{\alpha}_{-i}) > u_i(a_i, \hat{\alpha}_{-i})$ for all $i \in D^1$. Hence, there exist $\eta_i^a \in (0, \bar{\eta}]$ for all $i \in D^1 \setminus D^0$, such that the mixed action profile α^2 defined by

$$\alpha_i^2 \equiv \begin{cases} (1-\eta_i^a) \cdot a_i + \eta_i^a \cdot d_i^a & \text{if } i \in D^1 \\ a_i & \text{if } i \notin D^1 \end{cases}$$

satisfies $u_i(d_i^a, \alpha_{-i}^2) > u_i(a_i, \alpha_{-i}^2)$ for all $i \in D^1$. Thus $D^1 \subseteq Dev(\alpha^2) \equiv D^2$. Since the number of players is finite, repeating this procedure yields k such that $D^{k+1} = D^k$. We then set $\alpha^a = \alpha^{k+1}$. Q.E.D.

For all $a \in A$, let D^a be the set of randomizing players in α^a. Define

$$\underline{\eta}^a \equiv \min_{i \in D^a} \eta_i^a \qquad (\underline{\eta}^a = +\infty \text{ if } D^a = \emptyset),$$
$$\underline{\eta}^1 \equiv \min_{a \in A} \underline{\eta}^a > 0. \tag{18}$$

We now modify the minmax profiles slightly in such a way that all players except for the minmaxed player randomize over all actions. Specifically, for all $i \in N$, there exists a mixed action profile $m^i \in \mathcal{A}$ such that

$$\text{supp}(m_j^i) = A_j \quad \text{for all } j \neq i, \tag{19}$$
$$u_i(m^i) = \max_{a_i \in A_i} u_i(a_i, m_{-i}^i) < \underline{u}_i + \varepsilon, \tag{20}$$
$$m_i^i \in A_i, \tag{21}$$

where "supp" denotes the support of the probability distribution. By (19),

$$\underline{\eta}^2 \equiv \min_{i \in N} \min_{j \neq i} \min_{a_j \in A_j} m_j^i(a_j) > 0. \tag{22}$$

Assumption 2 implies that for all $k \in N$, there exist a completely mixed action profile β^k_{-k}, a player $r(k) \in N \setminus \{k\}$, and a partition of $\Omega_{r(k)}$, $\{\Omega^{1,k}_{r(k)}, \Omega^{2,k}_{r(k)}\}$, such that

$$\operatorname*{Argmax}_{b_k \in A_k} P_{r(k)}(\Omega^{1,k}_{r(k)} \mid b_k, \beta^k_{-k}) \bigcap \operatorname*{Argmax}_{b_k \in A_k} P_{r(k)}(\Omega^{2,k}_{r(k)} \mid b_k, \beta^k_{-k}) = \emptyset. \tag{23}$$

The player $r(k)$ will be called the *referee* for k. Let B^1_k and B^2_k be defined by

$$B^1_k \equiv \operatorname*{Argmax}_{b_k \in A_k} P_{r(k)}(\Omega^{1,k}_{r(k)} \mid b_k, \beta^k_{-k}), \qquad B^2_k \equiv \operatorname*{Argmax}_{b_k \in A_k} P_{r(k)}(\Omega^{2,k}_{r(k)} \mid b_k, \beta^k_{-k}).$$

By (23), $B^1_k \cap B^2_k = \emptyset$. Let b^1_k and b^2_k be such that

$$b^1_k \in \operatorname*{Argmax}_{b_k \in B^1_k} u_k(b_k, \beta^k_{-k}), \qquad b^2_k \in \operatorname*{Argmax}_{b_k \in B^2_k} u_k(b_k, \beta^k_{-k}).$$

The definitions of B^1_k and B^2_k imply

$$\begin{aligned} L^1_k &\equiv P_{r(k)}(\Omega^{1,k}_{r(k)} \mid b^1_k, \beta^k_{-k}) - \max_{b_k \notin B^1_k} P_{r(k)}(\Omega^{1,k}_{r(k)} \mid b_k, \beta^k_{-k}) > 0, \\ L^2_k &\equiv P_{r(k)}(\Omega^{2,k}_{r(k)} \mid b^2_k, \beta^k_{-k}) - \max_{b_k \notin B^2_k} P_{r(k)}(\Omega^{2,k}_{r(k)} \mid b_k, \beta^k_{-k}) > 0. \end{aligned} \tag{24}$$

Let $L \in \mathbb{R}$ be defined by

$$L \equiv \min_{k \in N} \min\{L^1_k, L^2_k\} > 0. \tag{25}$$

Since β^k_{-k} is completely mixed, we have

$$\underline{\eta}^3 \equiv \min_{k \in N} \min_{j \neq k} \min_{a_j \in A_j} \beta^k_j(a_j) > 0. \tag{26}$$

Let $\underline{\eta} \in \mathbb{R}$ be defined by

$$\underline{\eta} \equiv \min\{\underline{\eta}^1, \underline{\eta}^2, \underline{\eta}^3, 0.1\} > 0. \tag{27}$$

Let $\underline{p} \in \mathbb{R}$ be defined by

$$\underline{p} \equiv \min_{i \in N} \min_{\omega_i \in \Omega_i} \min_{a \in A} P_i(\omega_i \mid a) > 0,$$

where the inequality holds by Assumption 1. Let $(\mu, \underline{\delta}) \in (0,1)^2$ be sufficiently close to $(0,1)$ so that for all $\delta \in [\underline{\delta}, 1)$,

$$\hat{\mu} \equiv \frac{1-\delta}{\delta}\frac{1-q}{q} < 1, \tag{28}$$

$$2n(1-\delta)D < \mu\delta q\varepsilon, \tag{29}$$

$$2\mu D < \varepsilon, \tag{30}$$

$$6n(n-1)D(1-\delta+\mu\delta) < 0.1\delta^3\underline{\eta}\underline{p}L\varepsilon. \tag{31}$$

In what follows, we fix $\delta \geq \underline{\delta}$.

For all $j \in N$, let $W^j \in \mathbb{R}^N$ be defined by

$$W_i^j \equiv \begin{cases} (1-q)(v_i^* - \varepsilon) + q\big[u_i(m^i) - \hat{\mu}\lambda_i(N \setminus \{i\})\big] & \text{if } i = j, \\ (1-q)(v_i^* + \varepsilon) + q\Big[\min\limits_{\hat{a} \in \operatorname{supp}(m^j)} u_i(\hat{a}) - \hat{\mu}\lambda_i(N \setminus \{i,j\})\Big] & \text{if } i \neq j. \end{cases}$$

For all $v \in \bar{N}_\varepsilon(v^*)$, all $a \in A$, and all $i \in D^a$, let $\xi_i^v(a) \in \mathbb{R}$ be defined by

$$(1-\delta)\big[u_i(d_i^a, \alpha_{-i}^a) - u_i(a_i, \alpha_{-i}^a)\big] = \frac{1}{|D^a|}\xi_i^v(a)\mu\delta(v_i - W_i^i). \tag{32}$$

To see that $\xi_i^v(a)$ is a probability (i.e., falls in between 0 and 1), note that

$$v_i - W_i^i \geq v_i^* - \varepsilon - W_i^i > q(v_i^* - u_i(m^i) - \varepsilon) > q(v_i^* - \underline{u}_i - 2\varepsilon) > q\varepsilon \tag{33}$$

by (14) and (20). This implies $0 < \xi_i^v(a) \leq n(1-\delta)D/(\mu\delta q\varepsilon) < 1/2$ by (29).

For all $v \in \bar{N}_\varepsilon(v^*)$, we define a *virtual payoff function* $u^v \colon A \to \mathbb{R}^N$ as

$$\begin{aligned} u_i^v(a) \equiv{}& u_i(a_i, \alpha_{-i}^a) - \mu\lambda_i(D^a \setminus \{i\}) \\ &- \sum_{j \in D^a \setminus \{i\}} \eta_j^a \Big[u_j(d_j^a, \alpha_{-j}^a) - u_j(a_j, \alpha_{-j}^a)\Big]\frac{v_i - W_i^j}{v_j - W_j^j} \end{aligned} \tag{34}$$

for all i. For all $v \in \bar{N}_\varepsilon(v^*)$ and all $a \in A$,

$$\begin{aligned} |u_i^v(a) - u_i(a)| &\leq |u_i(a_i, \alpha_{-i}^a) - u_i(a)| + \mu D + (n-1)\bar{\eta}D\frac{D}{q\varepsilon} \\ &\leq D\Big[(n-1)\bar{\eta} + \mu + \frac{(n-1)\bar{\eta}D}{q\varepsilon}\Big] \\ &< \varepsilon \end{aligned}$$

by (16) and (30). This implies that there exists $\rho^* \in \Delta(A)$ such that

$$u^{v^*}(\rho^*) = v^*. \tag{35}$$

For all $h \in N$ and all $a \in \mathrm{supp}(m^h)$, let $V^h(a) \in \mathbb{R}^N$ be defined by

$$V_i^h(a) \equiv \begin{cases} v_i^* - \varepsilon & \text{if } i = h, \\ v_i^* + \varepsilon - \frac{q}{1-q}\big[u_i(a) - \min\limits_{\hat{a} \in \mathrm{supp}(m^h)} u_i(\hat{a})\big] & \text{if } i \neq h. \end{cases} \tag{36}$$

Then for all $i \neq h$, $v_i^* + \varepsilon \geq V_i^h(a) > v_i^*$ by (15). This implies

$$V^h(a) \in \bar{N}_\varepsilon(v^*).$$

Thus if we define $V^{**} \subseteq \mathbb{R}^N$ by

$$V^{**} \equiv \{v^*\} \cup \{V^h(a) : h \in N \text{ and } a \in \mathrm{supp}(m^h)\},$$

then $V^{**} \subseteq \bar{N}_\varepsilon(v^*)$.

A.2 Strategy

We now construct a strategy profile that yields the target payoff profile v^* and is a sequential equilibrium under $\delta \geq \underline{\delta}$. The strategy has four types of states: *cooperation*, *examination*, *report*, and *minmax*.

Cooperation States. A cooperation state, denoted $\mathit{Coop}(v, \rho)$, is indexed by a payoff profile $v \in V^{**}$ and a distribution $\rho \in \Delta(A)$. In particular, the initial period is in state $\mathit{Coop}(v^*, \rho^*)$, where ρ^* is given by (35). In each period of state $\mathit{Coop}(v, \rho)$, the public randomization at the beginning of the period selects each $a \in A$ with probability $\rho(a)$ as the cooperative action profile of the period. Suppose ρ selects a. Then players play the mixed action profile α^a. The observation activity is determined by the public randomization in the middle of the period. With probability $1 - \mu$, players do not observe any player. In this case, the state remains $\mathit{Coop}(v, \rho)$ in the next period.

With probability μ, on the other hand, each player i is prescribed to observe all players in $D^a \setminus \{i\}$. If $D^a = \emptyset$, then the public randomization at the beginning of the next period selects a pair of players (j, k) such that $j \neq k$ with equal probability, and the next period is in $\mathit{Exam}(v, (j, k), a_j, a_j)$. If $D^a \neq \emptyset$, let $(a_i^{obs})_{i \in D^a}$ denote the realized action profile of players in D^a. The public randomization at the beginning of the next period selects a pair of players (j, k) such that $j \in D^a$ and $k \neq j$ with equal probability. If $a_j^{obs} = d_j^a$, then the state changes to $\mathit{Minmax}(j)$ with probability $\xi_j^v(a)$. If $a_j^{obs} \notin \{a_j, d_j^a\}$, then the state changes to $\mathit{Minmax}(j)$ with probability $1/2$. In all these cases (where $D^a \neq \emptyset$), if the state does not change to $\mathit{Minmax}(j)$, it changes to $\mathit{Exam}(v, (j, k), a_j, a_j^{obs})$.

Examination States. An examination state, denoted $\mathit{Exam}(v, (j, k), a_j, a_j^{obs})$, is indexed by a payoff profile $v \in V^{**}$, a pair of players (j, k) such that $j \neq k$, $a_j \in A_j$, and

$a_j^{obs} \in A_j$. In this state, player k is prescribed to play a pure action b_k^{right} determined by

$$b_k^{right} = \begin{cases} b_k^1 & \text{if } a_j^{obs} = a_j, \\ b_k^2 & \text{otherwise.} \end{cases}$$

The other players play a mixed action profile β_{-k}^k. Regardless of the public randomization in the middle of the period, each player i observes $N \setminus \{i, k\}$. Let $b_{-k}^{obs} \in A_{-k}$ denote the realized action profile of players $N \setminus \{k\}$. The state transition depends on the public randomization at the beginning of the next period. With probability $1/2$, a pair (j', k') such that $j' \neq k$ and $k' \neq j'$ is selected with equal probability and the state changes to $\mathit{Exam}(v, (j', k'), b_{j'}^1, b_{j'}^{obs})$.[12] With the remaining probability, the state changes to $\mathit{Report}(v, k, b_k^{right}, b_{-k}^{obs})$.

Report States. A report state, denoted $\mathit{Report}(v, k, b_k^{right}, b_{-k}^{obs})$, is indexed by a payoff profile $v \in V^{**}$, a player $k \in N$, and an action profile $(b_k^{right}, b_{-k}^{obs}) \in \{b_k^1, b_k^2\} \times A_{-k}$. Player k is the one who was under examination in the last period (by construction, the previous period is an examination period). For each player i, choose a pair of distinct actions $\{c_i', c_i''\} \subseteq A_i$ arbitrarily in advance. In this period, player $r(k)$, i.e., the referee for k defined in (23), plays the following mixed action:

$$\begin{array}{ll} 0.9 \cdot c_{r(k)}' + 0.1 \cdot c_{r(k)}'' & \text{if } b_k^{right} = b_k^1 \text{ and } \omega_{r(k)} \in \Omega_{r(k)}^{1,k}, \\ 0.9 \cdot c_{r(k)}' + 0.1 \cdot c_{r(k)}'' & \text{if } b_k^{right} = b_k^2 \text{ and } \omega_{r(k)} \in \Omega_{r(k)}^{2,k}, \\ 0.1 \cdot c_{r(k)}' + 0.9 \cdot c_{r(k)}'' & \text{if } b_k^{right} = b_k^1 \text{ and } \omega_{r(k)} \notin \Omega_{r(k)}^{1,k}, \\ 0.1 \cdot c_{r(k)}' + 0.9 \cdot c_{r(k)}'' & \text{if } b_k^{right} = b_k^2 \text{ and } \omega_{r(k)} \notin \Omega_{r(k)}^{2,k}, \end{array}$$

where $\omega_{r(k)} \in \Omega_{r(k)}$ denotes the referee's private signal in the previous period. Any other player $i \neq r(k)$ plays a mixed action $0.5 \cdot c_i' + 0.5 \cdot c_i''$.

Regardless of the public randomization in the middle of the period, all players observe all the other players. Let $c^{obs} \in A$ denote the realized action profile (possibly $c_i^{obs} \notin \{c_i', c_i''\}$ for some players). The state transition depends on the public randomization at the beginning of the next period. With probability $1/2$, a new pair of players (j, k) such that $j \neq k$ is chosen with equal probability and the state changes to $\mathit{Exam}(v, (j, k), c_j', c_j^{obs})$. With the remaining probability, the state changes to $\mathit{Coop}(v, \rho)$ where $\rho \in \Delta(A)$ depends on $(v, k, b_k^{right}, b_{-k}^{obs}, c^{obs})$ and is determined as follows.

[12] Here the choice of a particular action $b_{j'}^1$ is arbitrary; any other action works since $\beta_{j'}^k$ assigns positive probability to all actions.

First, let $v' \in \mathbb{R}^N$ be defined by

$$\begin{aligned} v'_i = v_i &+ \frac{1-\delta+\mu\delta}{(1/4)\delta^2}\Big[v_i - u_i(b_k^{right}, b_{-k}^{obs}) + \lambda_i(N \setminus \{i,k\})\Big] \\ &+ \frac{1-\delta+\mu\delta}{(1/4)\delta^2}\Big[v_i - u_i(c^{obs}) + \lambda_i(N \setminus \{i\})\Big]\frac{1}{2}\delta. \end{aligned} \tag{37}$$

By (31),

$$|v'_i - v_i| \leq \frac{(1-\delta+\mu\delta)D(3/2)}{(1/4)\delta^2} < \varepsilon.$$

Thus $v' \in \bar{N}_\varepsilon(v) \subseteq \bar{N}_{2\varepsilon}(v^*)$. Now, we choose a distribution $\rho \in \Delta(A)$ such that

$$u_i^v(\rho) = v'_i \qquad \text{if } i \neq k, \tag{38}$$

$$u_k^v(\rho) = \begin{cases} v'_k + \varepsilon(1 - AP_k) & \text{if } c_{r(k)}^{obs} = c'_{r(k)}, \\ v'_k - \varepsilon AP_k & \text{otherwise,} \end{cases} \tag{39}$$

where

$$AP_k \equiv \begin{cases} 0.9P_{r(k)}(\Omega_{r(k)}^{1,k} \mid b_k^1, \beta_{-k}^k) + 0.1P_{r(k)}(\Omega_{r(k)}^{2,k} \mid b_k^1, \beta_{-k}^k) & \text{if } b_k^{right} = b_k^1, \\ 0.9P_{r(k)}(\Omega_{r(k)}^{2,k} \mid b_k^2, \beta_{-k}^k) + 0.1P_{r(k)}(\Omega_{r(k)}^{1,k} \mid b_k^2, \beta_{-k}^k) & \text{if } b_k^{right} = b_k^2 \end{cases} \tag{40}$$

denotes the ex ante probability that k's referee $r(k)$ plays $c'_{r(k)}$ given that players follow the strategy in the examination state. To see that ρ exists, note that by construction, $u^v(\rho)$ is within ε of v' and so within 3ε of v^*. Since u^v is within ε of u, it follows that $u(\rho)$ is within 4ε of v^*. Hence, ρ exists by (13).

Minmax States. A minmax state, denoted $Minmax(h)$, is indexed by a player $h \in N$ who is to be punished. In this state, players play the modified minmax action profile m^h (see (19)–(21)). The observation activity is determined by the public randomization in the middle of the period. With probability $1 - \hat{\mu}$, where $\hat{\mu}$ is defined in (28), players do not observe any player. In this case, the state remains the same in the next period.

With probability $\hat{\mu}$, on the other hand, each player i observes $N \setminus \{i, h\}$. Let $a_{-h}^{obs} \in A_{-h}$ denote the realized action profile of players $i \neq h$. The public randomization at the beginning of the next period chooses a pair (j, k) such that $j \neq h$ and $k \neq j$, and the state changes to

$$Exam(V^h(m_h^h, a_{-h}^{obs}), (j,k), b_j^1, a_j^{obs}),$$

where V^h is defined by (36).[13]

[13] Again, b_j^1 was chosen arbitrarily as the action that determines b_k^{right} in the examination stage. Any action will do since each player $j \neq h$ plays a completely mixed action under $Minmax(h)$.

We have specified the strategy profile on the equilibrium path. To complete the specification of the strategy profile, we first add the following rules. (i) The prescribed monitoring decision for a player does not depend on the stage-game action he chose in the period. That is, a player's own deviation in terms of stage-game action does not change the prescribed monitoring decision for the player in the period. (ii) The prescribed behavior (action and monitoring) for a player does not depend on any information he obtained by observing players whom he was not prescribed to observe. That is, if a player i observed a deviation of a player j in a period when player i was not prescribed to observe j, then i is prescribed to ignore the deviation and behave as if he did not observe it.

Let $\hat{\sigma}$ be a strategy profile that follows the state-dependent play described above and satisfies rules (i) and (ii). Consider a sequence of completely mixed strategy profiles $(\hat{\sigma}^1, \hat{\sigma}^2, \dots)$ that converge to $\hat{\sigma}$ and put far smaller weights on the trembles with respect to monitoring than those with respect to actions. This sequence generates a sequence of belief systems $(\psi^1, \psi^2, \dots)$ whose limit ψ is such that, at any history, each player believes that the other players have not deviated with respect to monitoring.

For each player i, let $\hat{H}_i$ be the set of i's (private) histories throughout which i observed all the players he was prescribed to observe under the state-dependent play (with rules (i) and (ii)). Thus $\hat{H}_i$ includes histories in which i deviated in terms of action, as well as histories in which i observed j when it was not prescribed. It should be noted that at all histories $h_i \in \hat{H}_i$, player i knows the current state and can follow the state-dependent play.

For each player i, let σ_i^* be a strategy that agrees with $\hat{\sigma}_i$ on $\hat{H}_i$ and such that, at all histories outside $\hat{H}_i$, the player plays a best response given the belief ψ and given that the other players follow $\hat{\sigma}_{-i}$. Let $\sigma^* = (\sigma_1^*, \dots, \sigma_n^*)$. We show that (σ^*, ψ) is a desired sequential equilibrium. To see that ψ is consistent with σ^*, consider a sequence of completely mixed strategy profiles $(\sigma^1, \sigma^2, \dots)$ that converges to σ^* and such that each σ_i^k agrees with $\hat{\sigma}_i^k$ on $\hat{H}_i$ and puts far smaller weights on the trembles with respect to monitoring than those with respect to actions. Then, the associated sequence of belief systems also converges to ψ. In what follows, we show that σ^* attains the target payoff profile v^* and is sequentially rational given ψ.

A.3 Values

In this section, we show that the strategy profile σ^* attains the target payoff profile v^*. To compute the continuation value for each state, we need to solve a system of equations. Since the set of states is finite in equilibrium (because the number of distributions ρ used in the cooperation states is finite in equilibrium), the solution is unique. To identify the solution, we first assume that the continuation value from any state of the form $\mathit{Exam}(v, \cdot)$ is exactly equal to v. We then show that this indeed constitutes a solution.

Given the assumption, we first compute the continuation value from minmax states. Given $h \in N$, let $M(h) \in \mathbb{R}^N$ denote the continuation payoff profile at the beginning of the state *Minmax*(h). The continuation payoff for player h is given by

$$M_h(h) = (1-\delta)\big[u_h(m^h) - \hat{\mu}\lambda_h(N \setminus \{h\})\big] + (1-\hat{\mu})\delta M_h(h) + \hat{\mu}\delta(v_h^* - \varepsilon).$$

Since $\hat{\mu}\delta = (1-\delta)(1-q)/q$ by definition, reorganizing the equation gives

$$M_h(h) = W_h^h.$$

To compute the continuation payoff for players $i \neq h$, let $M_i^a(h)$ denote the continuation payoff of player i evaluated at the beginning of the state given that $a \in \operatorname{supp}(m^h)$ is the realized action profile in this period. Then

$$M_i^a(h) = (1-\delta)\big[u_i(a) - \hat{\mu}\lambda_i(N \setminus \{i,h\})\big] + (1-\hat{\mu})\delta M_i(h) + \hat{\mu}\delta V_i^h(a).$$

Using $\hat{\mu}\delta = (1-\delta)(1-q)/q$ and substituting the definition of $V_i^h(a)$ in (36) give

$$\begin{aligned} M_i^a(h) = (1-\delta)\big[&\min_{\hat{a} \in \operatorname{supp}(m^h)} u_i(\hat{a}) - \hat{\mu}\lambda_i(N \setminus \{i,h\})\big] \\ &+ (1-\hat{\mu})\delta M_i(h) + \hat{\mu}\delta(v_i^* + \varepsilon). \end{aligned} \tag{41}$$

This implies that $M_i^a(h)$ does not depend on a and hence $M_i^a(h) = M_i(h)$ for all a. Substituting this fact into (41) yields

$$M_i(h) = W_i^h \quad \text{for all } i \neq h.$$

Thus $M(h) = W^h$ for any $h \in N$.

Abusing notation, let *Coop*$(v,\rho) \in \mathbb{R}^N$ denote the continuation payoff profile at the beginning of the state *Coop*(v,ρ). Then

$$\begin{aligned} \mathit{Coop}_i(v,\rho) = (1-\delta)\sum_{a \in A}\rho(a)\Big[u_i(\alpha_i^a, \alpha_{-i}^a) - \mu\lambda_i(D^a \setminus \{i\})\Big] + (1-\mu)\delta\,\mathit{Coop}_i(v,\rho) \\ + \mu\delta v_i - \mu\delta\sum_{a \in A}\rho(a)\sum_{j \in D^a}\frac{1}{|D^a|}\eta_j^a\xi_j^v(a)(v_i - W_i^j). \end{aligned}$$

When ρ selects $a \in A$ such that $i \in D^a$, player i is instructed to play a mixed action $\alpha_i^a = (1-\eta_i^a)\cdot a_i + \eta_i^a \cdot d_i^a$. The player is indeed indifferent between a_i and d_i^a because of the definition of $\xi_i^v(a)$ (see (32)). Therefore we can rewrite $Coop_i(v,\rho)$ as

$$\begin{aligned}Coop_i(v,\rho) = (1-\delta)\sum_{a\in A}\rho(a)\Big[u_i(a_i,\alpha_{-i}^a) - \mu\lambda_i(D^a\setminus\{i\})\Big] + (1-\mu)\delta\, Coop_i(v,\rho)\\ + \mu\delta v_i - \mu\delta\sum_{a\in A}\rho(a)\sum_{j\in D^a\setminus\{i\}}\frac{1}{|D^a|}\eta_j^a\xi_j^v(a)(v_i - W_i^j).\end{aligned}$$

Substituting the definition of $\xi_j^v(a)$ for the other players $j\neq i$ yields

$$\begin{aligned}(1-\delta+\mu\delta)\, Coop_i(v,\rho) = (1-\delta)\sum_{a\in A}\rho(a)\Bigg\{u_i(a_i,\alpha_{-i}^a) - \mu\lambda_i(D^a\setminus\{i\})\\ - \sum_{j\in D^a\setminus\{i\}}\eta_j^a\Big[u_j(d_j^a,\alpha_{-j}^a) - u_j(a_j,\alpha_{-j}^a)\Big]\frac{v_i - W_i^j}{v_j - W_j^j}\Bigg\} + \mu\delta v_i.\end{aligned}$$

Using virtual payoff functions u^v defined in (34) and writing $u^v(\rho) = \sum_{a\in A}\rho(a)u^v(a)$, we obtain

$$Coop(v,\rho) = \frac{(1-\delta)u^v(\rho) + \mu\delta v}{1-\delta+\mu\delta}. \tag{42}$$

Then by the definition of ρ^* given by (35),

$$Coop(v^*,\rho^*) = v^*.$$

Since ρ^* is used in the initial period, this implies that the target payoffs v^* are indeed achieved as the repeated-game payoffs under σ^*.

We now verify that the continuation value from an examination state of the form $Exam(v,\cdot)$ is indeed v. Consider an examination state $Exam(v,(j,k),a_j,a_j^{obs})$. We need to show that for all $i\in N$,

$$\begin{aligned}v_i = E\Bigg\{&(1-\delta)\big[u_i(b_k^{right}, b_{-k}^{obs}) - \lambda_i(N\setminus\{i,k\})\big]\\ &+ \frac{1}{2}\delta v_i + \frac{1}{2}\delta(1-\delta)\big[u_i(c^{obs}) - \lambda_i(N\setminus\{i\})\big]\\ &+ \frac{1}{4}\delta^2 v_i + \frac{1}{4}\delta^2\, Coop_i(v,\rho)\Bigg\},\end{aligned} \tag{43}$$

where the expectation is taken over (b_{-k}^{obs}, c^{obs}), and ρ is determined from (b_{-k}^{obs}, c^{obs}) by (37)–(39). For player $i\neq k$, substituting (38) into (37) and using (42) to replace

u^v with $Coop_i(v, \rho)$ yield

$$\begin{aligned} v_i &= (1-\delta)\big[u_i(b_k^{right}, b_{-k}^{obs}) - \lambda_i(N \setminus \{i,k\})\big] \\ &\quad + \frac{1}{2}\delta v_i + \frac{1}{2}\delta(1-\delta)\big[u_i(c^{obs}) - \lambda_i(N \setminus \{i\})\big] \\ &\quad + \frac{1}{4}\delta^2 v_i + \frac{1}{4}\delta^2 Coop_i(v, \rho) \end{aligned} \tag{44}$$

for all $b_{-k}^{obs} \in A_{-k}$ and all $c^{obs} \in A$. Taking the expectation of (44) implies (43). For player k, computation is the same except that $u_k^v(\rho) \neq v'_k$. Thus we obtain

$$\begin{aligned} v_k &= (1-\delta)\big[u_k(b_k^{right}, b_{-k}^{obs}) - \lambda_k(N \setminus \{k\})\big] \\ &\quad + \frac{1}{2}\delta v_k + \frac{1}{2}\delta(1-\delta)\big[u_k(c^{obs}) - \lambda_k(N \setminus \{k\})\big] \\ &\quad + \frac{1}{4}\delta^2 v_k + \frac{1}{4}\delta^2 Coop_k(v, \rho) - \frac{1}{4}\delta^2 \frac{1-\delta}{1-\delta+\mu\delta}\big[u_k^v(\rho) - v'_k\big], \end{aligned} \tag{45}$$

where the only non-trivial difference from (44) is the last term. However, this term is zero in expectation since the expected value of $u_k^v(\rho) - v'_k$ is $AP_k\varepsilon(1 - AP_k) - (1 - AP_k)\varepsilon AP_k = 0$. Thus (43) also holds for player k.

The fact that (44) holds for all $i \neq k$ and all $b_{-k}^{obs} \in A_{-k}$ implies that all players $i \neq k$ are completely indifferent over all actions in the examination period.

A.4 Incentives

We now show that σ^* is sequentially rational given ψ. We begin by showing that no player i has an incentive to deviate at any history $h_i \in \hat{H}_i$. Recall that, at histories $h_i \in \hat{H}_i$, player i knows the state and believes that the other players also know the state and follow the state-dependent play. We start with incentives in terms of stage-game actions.

Cooperation States. When the public randomization selects an action profile a as the cooperation action profile, players $i \in D^a$ are prescribed to randomize between a_i and d_i^a. As mentioned earlier, these players are indeed indifferent between these actions by the definition of $\xi_i^v(a)$. These players also do not have incentives to play any other action $a'_i \notin \{a_i, d_i^a\}$; indeed, the long-run loss is at least

$$\mu\frac{1}{2n}\delta[v_i - W_i^i] > \mu\frac{1}{2n}\delta q\varepsilon$$

by (33), and this exceeds $(1-\delta)D$ by (29). On the other hand, players $i \notin D^a$ have no incentive to deviate from a_i since by Lemma, a_i is a short-run best response to α_{-i}^a, and deviations are not observed and have no effects on the future play.

Report States. We show that in report states, all players are indifferent among all actions. Fix a report state $\mathit{Report}(v, k, b_k^{right}, b_{-k}^{obs})$. Let $R_i(c_i)$ denote player i's continuation payoff from this period if i chooses $c_i \in A_i$, where the expectation is taken with respect to c_{-i} based on all the information that player i has at the beginning of this period. For players $i \neq k$, (38) and (42) imply

$$\begin{aligned} R_i(c_i) = (1-\delta)\Big[E\big[u_i(c_i, c_{-i})\big] - \lambda_i(N \setminus \{i\})\Big] \\ + \frac{1}{2}\delta v_i + \frac{1}{2}\delta \frac{(1-\delta)E\big[v_i' \mid c_i\big] + \mu\delta v_i}{1-\delta+\mu\delta}, \end{aligned} \tag{46}$$

where v_i' depends on c_{-i} through (37). The right-hand side depends on c_i because of the two terms with expectation. But by taking the expectation of (37), we can see that

$$E\big[u_i(c_i, c_{-i})\big] + \frac{1}{2}\delta \frac{E\big[v_i' \mid c_i\big]}{1-\delta+\mu\delta}$$

as a whole does not depend on c_i. Thus $R_i(c_i)$ is actually constant in c_i. This implies that each player $i \neq k$ is completely indifferent about c_i in this period and therefore willing to randomize as instructed by the strategy.

For player k, the argument is the same except that v_i' in (46) for $i = k$ has to be replaced by $u_k^v(\rho)$ since $u_k^v(\rho) \neq v_k'$ for player k by (39). Thus

$$\begin{aligned} R_k(c_k) = (1-\delta)\Big[E\big[u_k(c_k, c_{-k})\big] - \lambda_k(N \setminus \{k\})\Big] \\ + \frac{1}{2}\delta v_k + \frac{1}{2}\delta \frac{(1-\delta)\Big[E[v_k' \mid c_k] + (AP_k' - AP_k)\varepsilon\Big] + \mu\delta v_k}{1-\delta+\mu\delta}, \end{aligned} \tag{47}$$

where AP_k' denote k's current belief about the probability that his referee $r(k)$ plays $c_{r(k)}'$ in this period.[14] The only difference between (47) and (46) is the term $(AP_k' - AP_k)\varepsilon$, which does not depend on c_k. Thus, the previous argument works for player k as well. Hence $R_k(c_k)$ does not depend on c_k and player k is also indifferent about his action.

Examination States. At the end of Section A.3, we showed that, in examination periods, players who are not under examination are indifferent over all actions. Thus we now prove that the player under examination (player k) is willing to play the pure action prescribed by the strategy (i.e., b_k^{right}). Consider an examination state $\mathit{Exam}(v, (j,k), a_j, a_j^{obs})$. Let $\ell \in \{1, 2\}$ be such that $b_k^{right} = b_k^{\ell}$. Since b_k^{right} is a short-run best response within B_k^{ℓ}, it suffices to verify that player k does not gain by playing any $b_k \notin B_k^{\ell}$. The short-run gains from playing any $b_k \notin B_k^{\ell}$ are at most $(1-\delta)D$.

[14] We may have $AP_k' \neq AP_k$ since AP_k is k's belief at the beginning of the previous period and he has since then updated his belief based on his signal and observations. Moreover, player k may have deviated in stage-game action in the previous period, which is possible since we are considering a history $h_k \in \hat{H}_k$.

On the other hand, by the definition of B_k^ℓ, playing an action $b_k \notin B_k^\ell$ necessarily lowers the probability that player $r(k)$ receives a signal $\omega_{r(k)} \in \Omega_{r(k)}^{\ell,k}$ at least by $L > 0$ (see (24) and (25)). If $\omega_{r(k)} \notin \Omega_{r(k)}^{\ell,k}$ and if the next period is a report period, then the probability that the referee plays $c'_{r(k)}$ (i.e., approves of k's answer) in the report period goes down from 0.9 to 0.1. If the referee indeed gives a disapproval and if the following period is a cooperation period, then the distribution ρ used in the cooperation period changes in such a way that $u_i^v(\rho)$ goes down by ε (see (39)), which in tern implies that the continuation value $Coop_i(v, \rho)$ goes down by $(1-\delta)\varepsilon/(1-\delta+\mu\delta)$.

Altogether, the long-run losses from playing $b_k \notin B_k^\ell$ are at least

$$L\frac{1}{2}(0.9-0.1)\frac{1}{2}\delta^2\frac{(1-\delta)\varepsilon}{1-\delta+\mu\delta}.$$

This exceeds $(1-\delta)D$ by (31).

Minmax States. Consider a state *Minmax*(h). In this state, player h has no incentive to deviate since the prescribed action m_h^h is a short-run best response against m_{-h}^h. The other players $i \neq h$ are willing to play m_i^h since $M_i^a(h)$ does not depend on a and hence they are completely indifferent.

Monitoring. We now verify that players have incentives to follow the strategy with respect to monitoring. First, no player has an incentive to observe a player who is not prescribed to be monitored, because such a player is expected to play a pure action. Suppose now that a player $k \in N$ chooses not to observe a player $j \neq k$ at the end of a period t when the strategy prescribes him to observe j. Since k is prescribed to observe j, player j was prescribed to play some mixed action $\alpha_j \in \mathcal{A}_j$ in this period. By the definition of $\underline{\eta}$, any action in the support of α_j is assigned a probability at least as large as $\underline{\eta} > 0$ (see (18), (22), (26), and (27)).

By construction, period $t+1$ is an examination period with a positive probability. The probability is at least as large as

$$\min\{1/2, 1 - \max_{a\in A}\max_{i\in N}\max_{v\in V^{**}} \xi_i^v(a)\} = 1/2 \tag{48}$$

since $\xi_i^v(a) < 1/2$. It should be noted that the lower bound (48) is valid even if period t was a cooperation period and player k played a major deviation in the period, since major deviations are ignored with probability 1/2. This point is relevant since we are considering a history $h_k \in \hat{H}_k$ and h_k may contain k's own (major) deviations in terms of actions.

If period $t+1$ is an examination period, then with at least probability $1/[n(n-1)]$, player k is chosen to be examined and is prescribed to "state" j's realized action in period t. In this case, the state in period $t+1$ is of the form *Exam*$(v, (j,k), a_j, a_j^{obs})$ where $\{a_j, a_j^{obs}\} \subseteq \text{supp}(\alpha_j)$. But player k does not know a_j^{obs} and hence is uncertain of b_k^{right}.

In this contingency, there is a positive probability bounded away from 0 with which player k plays a "wrong" action, playing $b_k \notin B_k^\ell$ when $b_k^{right} = b_k^\ell$ ($\ell \in \{1,2\}$). To see this, let $\omega_k^t \in \Omega_k$ denote the private signal that k received in the last period. Suppose, without loss of generality, that player k plays an action $b_k \in B_k^2$ (the case where $b_k \in B_k^1$ works similarly). This action is "wrong" if $b_k^{right} = b_k^1$, i.e., $a_j^{obs} = a_j$. The conditional probability that $a_j^{obs} = a_j$ holds given that k received ω_k^t is bounded below by $\underline{\eta}\underline{p}$, since $\underline{\eta}$ is the minimum probability assigned to each action in $\mathrm{supp}(\alpha_j)$ and $\underline{p}$ is the minimum probability assigned to each signal $\omega_k \in \Omega_k$.

If player k plays $b_k \notin B_k^\ell$ when $b_k^{right} = b_k^\ell$ ($\ell \in \{1,2\}$), then the probability that his referee $r(k)$ receives a signal $\omega_{r(k)} \in \Omega_{r(k)}^{\ell,k}$ goes down by at least $L > 0$. If the referee's signal falls outside of $\Omega_{r(k)}^{\ell,k}$ and if period $t+2$ is a report period, then the probability that the referee plays $c'_{r(k)}$ (i.e., approves of k's answer) goes down from 0.9 to 0.1. We can now apply the argument used for the incentives in examination periods. Then, the long-run losses from not observing player j in period t are at least

$$\frac{1}{2n(n-1)}\underline{\eta}\underline{p}L\frac{1}{2}(0.9-0.1)\frac{1}{2}\delta^3\frac{(1-\delta)\varepsilon}{1-\delta+\mu\delta}. \tag{49}$$

This exceeds the maximum short-run deviation gain, $(1-\delta)D$, by (31).

Histories $h_i \notin \hat{H}_i$. It remains to consider each player i's incentives at histories $h_i \notin \hat{H}_i$. By definition, the continuation play of σ_i^* given h_i prescribes an optimal decision for i at the history given his belief $\psi(h_i)$ and given that the other players follow $\hat{\sigma}$. By the construction of ψ, player i believes that the other players $j \neq i$ are at some histories $h_j \in \hat{H}_j$ and hence, by the definition of σ^*, their continuation strategies coincide with $\hat{\sigma}$. Therefore, following σ_i^* is sequentially rational for player i at h_i given ψ. Q.E.D.

References

Abreu, D., D. Pearce, and E. Stacchetti (1990): "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, 58, 1041–1063.

Ahn, I., and M. Suominen (2001): "Word-of-Mouth Communication and Community Enforcement," *International Economic Review*, 42, 399–415.

Aoyagi, M. (2002): "Collusion in Dynamic Bertrand Oligopoly with Correlated Private Signals and Communication," *Journal of Economic Theory*, 102, 229–248.

Ben-Porath, E., and M. Kahneman (1996): "Communication in Repeated Games with Private Monitoring," *Journal of Economic Theory*, 70, 281–297.

——— (2003): "Communication in Repeated Games with Costly Monitoring," *Games and Economic Behavior*, 44, 227–250.

Bhaskar, V., and I. Obara (2002): "Belief-Based Equilibria in the Repeated Prisoners' Dilemma with Private Monitoring," *Journal of Economic Theory*, 102, 40–69.

Compte, O. (1998): "Communication in Repeated Games with Imperfect Private Monitoring," *Econometrica*, 66, 597–626.

Ely, J., J. Hörner, and W. Olszewski (2003): "Belief-free Equilibria in Repeated Games," *Econometrica*, forthcoming.

Ely, J., and J. Välimäki (2002): "A Robust Folk Theorem for the Prisoner's Dilemma," *Journal of Economic Theory*, 102, 84–105.

Fudenberg, D., and D. Levine (1994): "Efficiency and Observability with Long-Run and Short-Run Players," *Journal of Economic Theory*, 62, 103–135.

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

Fudenberg, D., and E. Maskin (1986): "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information," *Econometrica*, 54, 533–556.

Green, E., and R. Porter (1984): "Noncooperative Collusion under Imperfect Price Information," *Econometrica*, 52, 87–100.

Hörner, J., and W. Olszewski (2004): "The Folk Theorem for All Games with Almost Perfect Monitoring," *mimeo.*

Kandori, M. (2003): "Randomization, Communication, and Efficiency in Repeated Games with Imperfect Public Monitoring," *Econometrica*, 71, 345–353.

Kandori, M., and H. Matsushima (1998): "Private Observation, Communication and Collusion," *Econometrica*, 66, 627–652.

Lehrer, E. (1991): "Internal Correlation in Repeated Games," *International Journal of Game Theory*, 19, 431–456.

Mailath, G., and S. Morris (2002): "Repeated Games with Almost-Public Monitoring," *Journal of Economic Theory*, 102, 189–228.

Matsushima, H. (1991): "On the Theory of Repeated Games with Private Information, Part II: Revelation Through Communication," *Economics Letters*, 35, 257–261.

——— (2004): "Repeated Games with Private Monitoring: Two Players," *Econometrica*, 72, 823–852.

Miyagawa, E., Y. Miyahara, and T. Sekiguchi (2003): "Repeated Games with Observation Costs," *mimeo.*

Piccione, M. (2002): "The Repeated Prisoner's Dilemma with Imperfect Private Monitoring," *Journal of Economic Theory*, 102, 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, 59–69.

Sekiguchi, T. (1997): "Efficiency in Repeated Prisoner's Dilemma with Private Monitoring," *Journal of Economic Theory*, 76, 345–361.

Yamamoto, Y. (2003): "Efficiency Results in N Players Games with Imperfect Private Monitoring," *mimeo.*