Folk Theorem in Repeated Games with Private Monitoring

Takuo Sugaya[*†]
Princeton University

November 15, 2011
The latest version and the online Supplemental Materials are available at
http://www.princeton.edu/~tsugaya/

Abstract

We show that the folk theorem generically holds for N-player repeated games with private monitoring when each player's number of signals is sufficiently large. Neither cheap talk communication nor public randomization is necessary.

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

Keywords: repeated game, folk theorem, private monitoring

[*]tsugaya@princeton.edu

[†]I thank Stephen Morris for invaluable advice, Yuliy Sannikov for helpful discussion, and Dilip Abreu, Eduardo Faingold, Drew Fudenberg, Edoardo Grillo, Johannes Hörner, Yuhta Ishii, Michihiro Kandori, Vijay Krishna, George Mailath, Mihai Manea, Larry Samuelson, Tadashi Sekiguchi, Andrzej Skrzypacz, Satoru Takahashi, Yuichi Yamamoto and seminar participants at 10th SAET conference, 10th World Congress of the Econometric Society, 22nd Stony Brook Game Theory Festival and Summer Workshop on Economic Theory for useful comments. The remaining errors are my own responsibility.

1 Introduction

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

Recent papers by Hörner and Olszewski (2006 and 2009) show that the folk theorem holds in *private monitoring*, where players can observe only private noisy signals about the action profile, if the monitoring is almost perfect and almost public, respectively.

The driving force of the folk theorem in perfect or public monitoring is the coordination of future play based on common knowledge of relevant histories. Specifically, the public component of histories, such as action profiles in perfect monitoring or public signals in public monitoring, reveals past action profiles (at least statistically). Since this public information is common knowledge, players can coordinate a punishment contingent on the public information, and thereby provide dynamic incentives to choose actions that are not static best responses.

Hörner and Olszewski (2006 and 2009) show the robustness of this coordination to the limited classes of private monitoring. If monitoring is almost perfect, then players can believe that every player observes the same signal corresponding to the action profile with high probability. If monitoring is almost public, then players can believe that every player observes the same signal with high probability.[1] Hence, almost common knowledge about relevant histories still exists.

However, with *general private monitoring*, almost common knowledge may not exist and

[1] See also Mailath and Morris (2002 and 2006).

coordination is difficult (we call this problem "*coordination failure*").[2] Hence, the robustness of the folk theorem to general private monitoring has been an open question. For example, Kandori (2002) states that "[t]his is probably one of the best known long-standing open questions in economic theory."[3]

This paper is, to the best of our knowledge, the first to show that the folk theorem holds in repeated games with discounting and generic private monitoring: In any N-player repeated game with private monitoring, if each player's number of signals is sufficiently large, then any feasible and individually rational payoff is sustainable in a sequential equilibrium for a sufficiently large discount factor.[4]

Repeated games with private monitoring are relevant for many traditional economic problems. For example, Stigler (1964) proposes a repeated price-setting oligopoly, where firms set their own price in a face-to-face negotiation and cannot directly observe their opponents' prices. Instead, a firm obtains some information about opponents' prices through its own sales. Since the level of sales depends on both opponents' prices and unobservable shocks due to business cycles, the sales level is an imperfect signal. Moreover, each firm's sales level is often private information. Thus, the monitoring is imperfect and private. In principal-agent problems, if the principal evaluates the agent subjectively, then the monitoring by the principal about the agent becomes private. Despite the importance of these problems, only a limited number of papers successfully analyze the repeated games with private monitoring.[5] Our result offers a benchmark to analyze these problems in a general private-monitoring setting.

To show the folk theorem under general private monitoring, we unify and improve on three approaches in the literature on private monitoring that have been used to show the

[2] Mailath and Morris (2002 and 2006) and Sugaya and Takahashi (2011) offer the formal models of this argument.

[3] See Mailath and Samuelson (2006) for a survey.

[4] See Lehrer (1990) for the case of no discounting.

[5] Harrington and Skrzypacz (2011) show evidence of cooperative behavior (cartels) among firms in lysine and vitamin industries. After arguing that these industries fit Stigler's setup, they write a repeated-game model with private monitoring and solve a special case. See also Harrington and Skrzypacz (2007).

Fuchs (2007) applies a repeated game with private monitoring to a contract between a principal and an agent with subjective evaluation.

partial results so far: Belief-free, belief-based and communication approaches.

The belief-free approach (and its generalizations) has been successful in showing the folk theorem in the prisoners' dilemma.[6] A strategy profile is *belief-free* if, for any history profile, the continuation strategy of each player is optimal conditional on the history of the opponents. Hence, coordination failure never happens. With almost perfect monitoring, Piccione (2002) and Ely and Välimäki (2002) show the folk theorem for the two-player prisoners' dilemma.[7] Without any assumption on the precision of monitoring but with conditionally independent monitoring, Matsushima (2004) obtains the folk theorem in the two-player prisoners' dilemma, which is extended by Yamamoto (2011) to the N-player prisoners' dilemma with conditionally independent monitoring.[8]

Previously, attempts to generalize Matsushima (2004) have shown only limited results without almost perfect or conditionally independent monitoring: For some restricted classes of the distributions of private signals, Fong, Gossner, Hörner and Sannikov (2010) show that the payoff of the mutual cooperation is approximately attainable and Sugaya (2010a) shows the folk theorem in the two-player prisoners' dilemma. Sugaya (2010b) shows that the folk theorem holds with a general monitoring structure in the prisoners' dilemma if the number of players is no less than four.

Several papers construct belief-based equilibria, where players' strategies involve statistical inference about the opponents' past histories. That is, since common knowledge about relevant histories no longer exists, each player calculates the beliefs about the opponents' histories to calculate best responses. With almost perfect monitoring, Sekiguchi (1997)

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

[7]See Yamamoto (2007) for the N-player prisoners' dilemma. Ely, Hörner and Olszewski (2004 and 2005) and Yamamoto (2009) characterize the set of belief-free equilibrium payoffs for a general game. Except for the prisoners' dilemma, this set is not so large as that of feasible and individually rational payoffs.

[8]The strategy used in Matsushima (2004) is called a "belief-free review strategy." See Yamamoto (2011) for the characterization of the set of belief-free review-strategy equilibrium payoffs for a general game with conditional independence. Again, except for the prisoners' dilemma, this set is not so large as that of feasible and individually rational payoffs.

shows that the payoff of the mutual cooperation is approximately attainable and Bhaskar and Obara (2002) show the folk theorem in the two-player prisoners' dilemma.[9] Phelan and Skrzypacz (2011) characterize the set of possible beliefs about opponents' states in a finite-state automaton strategy and Kandori and Obara (2010) offer a way to verify if a finite-state automaton strategy is an equilibrium.

Another approach to analyze repeated games with private monitoring introduces public communication. Folk theorems have been proven by Compte (1998), Kandori and Matsushima (1998), Aoyagi (2002), Fudenberg and Levine (2002) and Obara (2009). Introducing a public element (the result of communication) and letting a strategy depend only on the public element allow these papers to sidestep the difficulty of coordination through private signals. However, the analyses are not applicable to settings where communication is not allowed: For example, in Stigler (1964)'s oligopoly example, anti-trust laws prohibit communication. Hörner and Olszewski (2006) argue that "communication reintroduces an element of public information that is somewhat at odds with the motivation of private monitoring as a robustness test" to the lack of common knowledge.

This paper incorporates all three approaches. First, the equilibrium strategy to show the folk theorem is *phase-belief-free*. That is, we see the repeated game as the repetition of long review phases. Each player has two strategies for the review phase; one that is generous to the opponent and another that is harsh to the opponent.[10] At the beginning of each review phase, for each player, both generous and harsh strategies are optimal conditional on any realization of the opponents' history. Within each review phase, each player can change the opponent's continuation payoff from the next review phase by changing the transition probability between the two strategies, without considering the other players' history. This equilibrium is immune to coordination failure at the beginning of each phase and gives us freedom to control the continuation payoffs.

Second, however, the belief-free property does not hold except at the beginning of the

[9] Bhaskar and Obara (2002) also derive a sufficient condition for the N-player prisoners' dilemma.

[10] As will be seen in Section 5, for a game with more than two players, one of player i's strategies is generous to player $i+1$ and another of player i's strategies is harsh to player $i+1$. In addition, players $-(i,i+1)$'s payoffs are constant regardless of which strategy player i picks from the two.

phases. Hence, we consider each player's statistical inference about the opponents' past histories as in the belief-based approach within each phase.

Finally, in our equilibrium, to coordinate on the play in the middle of the phase, the players do communicate but the message exchange is done with their actions. The difficulty to replace cheap talk with messages via actions is that, since the players need to infer the opponents' messages from their private signals, common knowledge about the past messages no longer exists. One of our methodological contributions is to offer a systematic way to replace the cheap talk with message exchange via actions in private monitoring by overcoming the lack of common knowledge.

The rest of the paper is organized as follows: Section 2 introduces the model and Section 3 states the assumptions and main result. Section 4 offers the overview of the structure of the proof. Section 5 relates the infinitely repeated game to a finitely repeated game with an auxiliary scenario (reward function) and derives a sufficient condition on the finitely repeated game to show the folk theorem in the infinitely repeated game. The remaining parts of the paper are devoted to the proof of the sufficient condition. Section 6 explains the basic structure of the finitely repeated game. As will be seen in Section 7, we concentrate on the approximate equilibrium until Section 13. Since the complete proof is long and complicated, for the rest of the main text (that is, from Section 8 to Section 15), we concentrate on a special case explained in Section 8 to illustrate the key structure. Namely, we focus on the two-player prisoners' dilemma with cheap talk and public randomization, and interested readers may refer to the Supplemental Materials for the complete proof for a *general* game *without* cheap talk or public randomization. Section 9 specifies what assumptions are sufficient in this special case. After we formally define the structure of the finitely repeated game for the two-player prisoners' dilemma in Section 10, we define the strategy in Section 11. While defining the strategy, we define many variables. Section 12 verifies that we take all the variables coherently. Section 13 shows that the strategy approximately satisfies the sufficient condition derived in Section 5. Finally, Section 14 adjusts the strategy further so that it exactly satisfies the sufficient condition (therefore, we are *not* considering an approximate equilibrium. The final strategy is an *exact* sequential equilibrium). All proofs are given in the Appendix (Section 15). Sections from 16 to 53 are in the Supplemental Materials.

2 Model

2.1 Stage Game

The stage game is given by $\left\{I, \{A_i, Y_i, U_i\}_{i \in I}, q\right\}$. $I = \{1, \ldots, N\}$ is the set of players, A_i with $|A_i| \geq 2$ is the finite set of player i's pure actions, Y_i is the finite set of player i's private signals, and U_i is the finite set of player i's ex-post utilities. Let $A \equiv \prod_{i \in I} A_i$, $Y \equiv \prod_{i \in I} Y_i$ and $U \equiv \prod_{i \in I} U_i$ be the set of action profiles, signal profiles and ex post utility profiles, respectively.

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

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

Player i's expected payoff from $a \in A$ is the ex ante value of $\tilde{u}_i$ given a and is denoted $u_i(a)$. For each $a \in A$, let $u(a)$ represent the payoff vector $(u_i(a))_{i \in I}$.

2.2 Repeated Game

Consider the infinitely repeated game of the above stage game in which the (common) discount factor is $\delta \in (0, 1)$. Let $a_{i,\tau}$ and $y_{i,\tau}$, respectively, denote the action played and the private signal observed in period τ by player i. Player i's private history up to period $t \geq 1$ is given by $h_i^t \equiv \{a_{i,\tau}, y_{i,\tau}\}_{\tau=1}^{t-1}$. With $h_i^1 = \{\emptyset\}$, for each $t \geq 1$, let H_i^t be the set of all h_i^t. A strategy for player i is defined to be a mapping $\sigma_i : \bigcup_{t=1}^{\infty} H_i^t \to \triangle(A_i)$. Let Σ_i be the set of all strategies for player i. Finally, let $E(\delta)$ be the set of sequential equilibrium payoffs with a common discount factor δ.

3 Assumptions and Result

In this section, we state two assumptions and the main result (folk theorem).

First, we state an assumption on the payoff structure. Let $F \equiv \text{co}(\{u(a)\}_{a\in A})$ be the set of *feasible payoffs*. The *minimax payoff* for player i is

$$v_i^* \equiv \min_{\alpha_{-i}\in\Pi_{j\neq i}\Delta(A_j)} \max_{a_i\in A_i} u_i(a_i, \alpha_{-i}).$$

Then, the set of *feasible and individually rational payoffs* is given by $F^* \equiv \{v \in F : v_i \geq v_i^*$ for all $i\}$. We assume the full dimensionality of F^*.

Assumption 1 *The stage game payoff structure satisfies the full dimensionality condition:* $\dim(F^*) = N$*.*

Second, we state an assumption on the signal structure.

Assumption 2 *Each player's number of signals is sufficiently large: For any $i \in I$, we have*

$$|Y_i| \geq 2\sum_{j\in I} |A_j|.$$

Under these assumptions, we can generically construct an equilibrium to attain any point in $\text{int}(F^*)$.

Theorem 1 *If Assumptions 1 and 2 are satisfied, then the folk theorem generically holds: For generic $q(\cdot \mid \cdot)$, for any $v \in \text{int}(F^*)$, there exists $\bar{\delta} < 1$ such that, for all $\delta > \bar{\delta}$, $v \in E(\delta)$.*

See Section 9 and the Supplemental Material 1 for exactly what genericity conditions we need in the proof. As will be seen, Assumption 2 is more than necessary. What we need for the proof is

$$|Y_i| \geq \begin{cases} |A_i| + 2\,|A_{i-1}| \\ \text{if } N = 2, \\ \max \left\{ \begin{array}{c} |A_i| + |A_{i+1}| - 1 + 2\sum_{j \neq i,i+1} (|A_j| - 1)\,, |A_{i-1}| + \sum_{j \neq i-1,i} (|A_j| - 1)\,, \\ 2\,|A_{i-1}|\,, \max_{j \in I} \left\{ \frac{1}{2}\,|A_j| + \sum_{n \neq j} |A_n| \right\}, \max_{j \in I} \left\{ \frac{1}{2}\,|A_j| + 2\sum_{n \neq i,j} |A_n| \right\} \end{array} \right\} \\ \text{if } N \geq 3. \end{cases}$$

From now on, we arbitrarily fix $v \in \text{int}(F^*)$ and construct an equilibrium to support v in a sequential equilibrium.

4 An Overview of the Argument

This section provides some intuition for our construction. Following Hörner and Olszewski (2006), we see a repeated game as repetition of T_P-period review phases. T_P will be formally defined later. In Section 4.1, we explain that our equilibrium is "phase-belief-free" and how it makes our equilibrium immune to coordination failure at the beginning of each phase. Section 4.2 offers the basic structure of the review phase.

To explain the details of the review phase, it is useful to consider a special case where additional communication devices are available. Section 4.3 introduces these devices. With these communication devices, in Sections 4.4, 4.5 and 4.6, we offer the detailed explanation of the review phase.

Finally, we explain how to dispense with the communication devices in Section 4.7.

4.1 Phase-Belief-Free

As Hörner and Olszewski (2006), the equilibrium is *phase-belief-free*. Each player i has two T_P-period-finitely-repeated-game strategies, denoted $\sigma_i(G)$ and $\sigma_i(B)$. At the beginning of each review phase, for each player i, independently of her history, any continuation strategy that adheres to one of the two strategies $\sigma_i(G)$ and $\sigma_i(B)$ in the review phase is optimal. We say that player i taking $\sigma_i(x_i)$ with $x_i \in \{G, B\}$ in the review phase is "in state $x_i \in \{G, B\}$."

Intuitively speaking, $\sigma_i(G)$ is a "generous" strategy that gives a high payoff to player $i+1 \pmod N$ who takes either $\sigma_{i+1}(G)$ or $\sigma_{i+1}(B)$, regardless of the other players' state profile $x_{-(i,i+1)} \in \{G, B\}^{N-2}$. On the other hand, $\sigma_i(B)$ is a "harsh" strategy that gives a low payoff to player $i+1$ regardless of player $(i+1)$'s strategy (including those different from $\sigma_{i+1}(G)$ and $\sigma_{i+1}(B)$) and $x_{-(i,i+1)}$. Hence, player $(i-1)$'s strategy controls player i's value regardless of $x_{-(i-1)}$, replacing i with $i-1$ in the previous two sentences. Since these two strategies are optimal at the beginning of the next phase, it is up to player $i-1$ whether player $i-1$ will take $\sigma_{i-1}(G)$ or $\sigma_{i-1}(B)$ in the next phase. Therefore, player $i-1$ with $\sigma_{i-1}(G)$ in the current phase can freely reduce player i's continuation payoff from the next review phase by transiting to $\sigma_{i-1}(B)$ with higher probability while player $i-1$ with $\sigma_{i-1}(B)$ can freely increase player i's continuation payoff by transiting to $\sigma_{i-1}(G)$ with higher probability.[11] In summary, we do not need to consider player $(i-1)$'s incentive to punish player i after a "bad history" in state G or to reward player i after a "good history" in state B.

4.2 Structure of the Review Phase

The basic structure of the review phase is summarized as follows. At the beginning of the review phase, the players communicate a state profile $x \in \{G, B\}^N$. This communication stage is named the "coordination block" since the players try to coordinate on x. The details will be explained in Section 4.4.

Based on the result of the coordination block, the players play the finitely repeated

[11] Here, the changes in the continuation payoffs are measured by the differences between player i's ex ante value given x_{i-1} at the beginning of the review phase and the ex post value at the end of the review phase after player $i-1$ observes the history in the phase. See Section 5 for the formal definition.

For example, if player $i-1$ with $x_{i-1} = G$ does not reduce player i's continuation value, then it means that the state of player $i-1$ in the next review phase is G with probability one, so that the ex post value is the same as the ex ante value.

game for many periods. This step consists of multiple "review rounds." The details will be explained in Section 4.6.

Finally, at the end of the phase, the players communicate the history in the coordination block and review rounds. This stage is named the "report block" since the players report the history in the review rounds. The role of this communication will be explained in Section 4.5.

4.3 Special Communication Devices

Before explaining the details of the coordination block, review rounds and report block, we introduce three special communication devices. We will dispense with all three in Section 4.7.

Perfect Cheap Talk Until Section 4.6, we assume that the players could directly communicate in the coordination block and report block. We assume that the communication were (i) cheap (not directly payoff-relevant), (ii) instantaneous and (iii) public and perfect (it generates the same signal as the message to each player).

Noisy Cheap Talk In the review rounds, we assume that the players could directly communicate via *noisy* cheap talk. We will later explain why we use noisy cheap talk rather than the perfect cheap talk in the review rounds.

"Noisy cheap talk with precision $p \in (0,1)$" is the communication device that is (i) cheap and (ii) instantaneous, but (iii) private and imprecise with probability $\exp(-O(T^p))$.[12] Specifically, when the sender (say player j) sends a binary message $m \in \{G, B\}$ via noisy cheap talk, the receiver (say player i) will observe a binary private signal $f[i](m) \in \{G, B\}$. With high probability, the message transmits correctly: $f[i](m) = m$ with probability $1 - \exp(-O(T^p))$. Given the true message m and the receiver's private signal $f[i](m)$, the controller of the receiver's payoff (player $i-1$) stochastically receives a binary private signal

[12] In general, when we say $y = O(x)$, it means that there exists $k > 0$ such that $y = kx$.

$g[i-1](m) \in \{m, E\}$. If $f[i](m) \neq m$ (if the receiver receives a wrong signal), then $g[i-1](m) = E$ with probability $1 - \exp(-O(T^p))$. That is, $g[i-1](m) = E$ implies that player $i-1$ (the controller of player i's payoff) suspects that the communication may have an error. Further, we assume that any signal pair can occur with probability at least $\exp(-O(T^p))$. Hence, the communication is noisy.

We assume that the signals are private. Therefore, $f[i](m)$ is observable only to the receiver (player i) and $g[i-1](m)$ is observable only to the controller of the receiver's payoff (player $i-1$).

There are two important features of this noisy cheap talk: First, whenever the receiver realizes that her signal was wrong: $f[i](m) \neq m$, then she puts a belief no less than $1 - \exp(-O(T^p))$ on the event that the controller of her payoff should have received the signal $g[i-1](m) = E$ and "realized" there was an error.[13] Second, any error occurs with positive probability $\exp(-O(T^p))$. It will be clear in Section 4.6 that these two features are important to construct an equilibrium in the review rounds.

Public Randomization In the report block, we assume that public randomization were available in addition to the perfect cheap talk.

With these special communication devices, Sections 4.4, 4.5 and 4.6 explain the coordination block, the report block and the review rounds, respectively.

4.4 Coordination Block

The role of the coordination block is to coordinate on x as in Hörner and Olszewski (2006). With the perfect cheap talk, each player tells the truth about her own state x_i and the state profile $x \in \{G, B\}^N$ becomes common knowledge. In the review rounds, based on x, the players play $a(x)$ with high probability on the equilibrium path. Intuitively, $a(x)$ is the action profile taken in the "regular" histories when the state profile is x. See Section 5 for the formal definition of $a(x)$.

[13] As we will see, player $(i-1)$'s continuation play is independent of $g[i-1](m)$ and so player i cannot learn $g[i-1](m)$.

4.5 Report Block

We introduce the report block where the players communicate the history in the coordination block and review rounds. This communication enables us to concentrate on ε-equilibrium until the end of the last review round. Suppose that we have constructed a strategy profile which is ε-equilibrium at the end of the last review round if we neglect the report block. We explain how to attain the exact equilibrium by using the report block.

As seen in Section 4.3, suppose that the perfect cheap talk and public randomization are available. Each player i is picked by the public randomization with probability $\frac{1}{N}$.[14] The picked player i sends the whole history in the coordination block and review rounds (denoted h_i^{main}) to player $i-1$. That is, h_i^{main} is player i's history from the beginning of the current review phase to the end of the last review round.

Assume that player i always tells the truth about h_i^{main}. Player $i-1$ changes the continuation payoff of player i such that, after any period t in the coordination block and review rounds, after any history h_i^t, it is exactly optimal for player i to follow the prescribed action by $\sigma_i(x_i)$. Since the original strategy profile was ε-equilibrium with arbitrarily small ε, this can be done by slightly changing the continuation strategy based on h_{i-1}^{main} and h_i^{main}.[15]

The remaining task with the perfect cheap talk and public randomization is to show the incentive to tell the truth about h_i^{main}. Intuitively, with defining a linear space and norm properly for the histories, player $i-1$ punishes player i proportionally to $\left\| h_{i-1}^{\text{main}} - \mathbb{E}\left[h_{i-1}^{\text{main}} \mid \hat{h}_i^{\text{main}} \right] \right\|^2$ with $\hat{h}_i^{\text{main}}$ being the reported history. The optimal report $\hat{h}_i^{\text{main}}$ to minimize the expected punishment $\mathbb{E}\left[\left\| h_{i-1}^{\text{main}} - \mathbb{E}\left[h_{i-1}^{\text{main}} \mid \hat{h}_i^{\text{main}} \right] \right\|^2 \mid h_i^{\text{main}} \right]$ is to tell the truth: $\hat{h}_i^{\text{main}} = h_i^{\text{main}}$.[16] Since the adjustment for exact optimality is small, the small punishment is enough to incentivize player i to tell the truth. Therefore, the total changes in the continuation payoff based on the report block do not affect the equilibrium payoff.

[14] For $N \geq 3$, the precise procedure is slightly different. See Section 36 in the Supplemental Material 3.

[15] With more than two players, player $i-1$ also needs to know the histories of players $-(i-1, i)$. So that players $-(i-1, i)$ can send their histories to player $i-1$, we introduce another communication stage after the report block, named the "re-report block." Since this information sent by players $-(i-1, i)$ in the re-report block is used only to control player i's continuation payoff, the truthtelling incentive for players $-(i-1, i)$ is trivially satisfied. See Section 37 in the Supplemental Material 3.

[16] Note that this logic is the same as we show the consistency of generalized-method-of-moments estimators.

4.6 Review Rounds

Between the coordination block and the report block, the players play a T-period "review round" for L times. Here, $L \in \mathbb{N}$ is a fixed integer that will be determined in Section 12, and

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

so that

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

Throughout the paper, we neglect the integer problem since it is handled by replacing each variable s that should be an integer with $\min_{\substack{n\in\mathbb{N} \\ n\geq s}} n$.

The reason why we have T periods in each review round is to aggregate private signals for many periods to get precise information as in Matsushima (2004).[17] There are two reasons why we have L review rounds. The first reason is new: As we will explain, the signals of the players can be correlated while Matsushima (2004) assumes that the signals are conditionally independent. To deal with correlation, we need multiple review rounds.

The second reason is the same as Hörner and Olszewski (2006). If we replace each period of Hörner and Olszewski (2006) with a T-period review round, then we need a sufficiently large number of review rounds so that a deviator should be punished sufficiently long to cancel out the gains in the instantaneous utility from deviation.

Below, we offer a more detailed explanation of the review rounds. In Section 4.6.1, we concentrate on the first role of the L rounds. That is, we consider the case where the block of Hörner and Olszewski (2006) has one period, that is, the stage game is the two-player prisoners' dilemma. We will explain a general two-player game and a general more-than-two-player game in Sections 4.6.2 and 4.6.3, respectively, where the second role of the L rounds is important.

Whenever we consider the two-player case and we say players i and j, we assume that player j is player i's (unique) opponent unless otherwise specified.

[17] See also Radner (1985) and Abreu, Milgrom and Pearce (1991).

4.6.1 The Two-Player Prisoners' Dilemma

In the two-player prisoners' dilemma, we consider player i's incentive to take $\sigma_i(G)$ when player j takes $\sigma_j(G)$. The other combinations of (x_i, x_j) are symmetric. Remember that since x is communicated via perfect cheap talk, x is common knowledge.

So that $\sigma_i(G)$ is generous to player j, player i needs to take cooperation with ex ante high probability. On the other hand, player j can reduce player i's continuation payoff from the next review phase based on her history within the current review phase (see the explanation of phase-belief-free in Section 4.1).

Suppose that player j has a "good" random variable (signal) which occurs with probability q_2 when player i takes cooperation and with probability $q_1 < q_2$ when player i takes defection. $q_2 > 0$ can be very small since the monitoring is imperfect. Assume that the instantaneous utility gain of taking defection instead of cooperation is $g > 0$.

If player j needs to incentivize player i to take cooperation every period independently, then player j needs to reduce player i's continuation payoff by at least $\frac{g}{q_2 - q_1}$ (for simplicity, forget about discounting) after not observing the good signal. Then, the ex ante per-period reduction of the continuation payoff is $\frac{g}{q_2 - q_1}(1 - q_2)$, which is too large to attain efficiency (if q_2 is bounded away from one). That is, player j switches to the harsh strategy (which takes defection in the prisoners' dilemma) from the next review phase too often. Hence, we need to come up with a procedure to prevent the inefficient punishment (reduction of the continuation payoff).

Conditional Independence Following Matsushima (2004), assume that player i's signals were independent of player j's signals. In this case, we could see a collection of L review rounds as one "long review round." That is, player j monitors player i for LT periods. Player j will take the generous strategy with probability one in the next review phase if the good signal is observed $(q_2 + 2\varepsilon) LT$ times or more.[18] If it is observed less, then player j reduces the continuation payoff by the shortage multiplied by $\frac{g}{q_2 - q_1}$. That is, with X_j being how

[18] We will explain why we use 2ε instead of ε later. In addition, this ε is different from ε for ε-equilibrium.

many times player j has observed the good signal in the LT periods, the reduction of the continuation payoff will be $\frac{g}{q_2-q_1}\left\{\left(q_2+2\varepsilon\right)LT-X_j\right\}_+$.[19] We call X_j "player j's score about player i."

Since player i's signals were independent of player j's signals, player i could not update any information about player j's score about player i from player i's private signals. Hence, by the law of large numbers, for sufficiently large T, player i believes that $\left(q_2+2\varepsilon\right)LT-X_j>0$ with ex post high probability after any history. Hence, it is optimal for player i to constantly take cooperation. At the same time, since the expected value of X_j is q_2LT, the ex ante per-period reduction of the continuation payoff is $\frac{g}{q_2-q_1}2\varepsilon$, which can be arbitrarily small by taking ε small. Therefore, we are done.

Conditional Dependence Now, we dispense with conditional independence. That is, player i's signals and player j's signals can be correlated arbitrarily. Intuitively, see one period as a day and a long review round as a year: $LT=365$. Since the expected score is q_2LT, to prevent an inefficient punishment, player j cannot punish player i after the score slightly exceeds q_2LT (in the above example, $\left(q_2+2\varepsilon\right)LT$). On the other hand, if the signals are correlated, then later in a year (say, November), it happens with a positive probability that player i believes that, judging from her own signals and correlation, player j's score about player i has been much more than q_2LT already (in the above example, more than $\left(q_2+2\varepsilon\right)LT$). Then, player i wants to start to defect.

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

Intuitively, player i switches to defection if player i's expectation of player j's score about player i is much higher than the ex ante mean. We want to specify exactly when each player i takes defection based on player i's expectation of player j's score about player i.

[19] $\{X\}_+$ is equal to X if $X\geq 0$ and 0 otherwise.

Chain of Learning However, this creates the following problem: Since player i switches her action based on player i's expectation of player j's score about player i, player i's action reveals player i's expectation of player j's score about player i. Since both "player i's expectation of player j's score about player i" and "player i's score about player j" are calculated from player i's history, player j may want to learn "player i's expectation of player j's score about player i" from "player j's signals about player i's action." If so, player j's decision of actions depends also on player j's expectation of player i's expectation of player j's score about player i. Proceeding one step further, player i's decision of actions depends on player i's expectation of player j's expectation of player i's expectation of player j's score about player i. This chain continues infinitely.

Noisy Cheap Talk Cuts off the Chain of Learning We want to construct an equilibrium that is not destroyed by the chain of high order expectations. From the discussion of the report block, we can focus on ε-equilibrium. This means that, to verify an equilibrium, it is enough to show that each player believes that her action is strictly optimal or any action is optimal with high probability (not probability one). To prevent the chain of learning, we take advantage of this "ε slack" in ε-equilibrium and the noise in the noisy cheap talk explained in Section 4.3.

The basic structure is as follows. We divide an LT-period long review round into L T-period review rounds. We make sure that each player takes a constant action within a review round. If player j observes a lot of good signals in a review round, then player i should take defection from the next review round. At the end of each review round, player j sends a noisy cheap talk message with precision $p = \frac{1}{2}$ to inform player i of the optimal action in the next review round. Based on player i's own history and player i's signal of player j's message via noisy cheap talk, player i may switch to a constant defection from the next review round. That is, the breakdown of incentives and switches of actions occur only at the beginning of each review round. The remaining questions are (i) how we can incentivize player j to tell the truth and (ii) how we can make sure that the chain of learning does not destroy an equilibrium.

The intuitive answer to these questions are as follows. In equilibrium, by the law of large numbers, with ex ante high probability, player i at the end of the review round puts a high belief on the event that player j has not observed a lot of good signals. In such a case, player i believes that player i's optimal action in the next review round is cooperation and disregards the signal of player j's message. That is, the precision of player i's inference about player i's optimal action from the review rounds is usually $1 - \exp(-O(T))$ because the length of the review round is T. Since this is higher than the precision of the signal of the noisy cheap talk, $1 - \exp(-O(T^{\frac{1}{2}}))$, player i disregards the signal. Player i incentivizes player j to tell the truth by changing player j's continuation payoff from the next review phase only if player i does not disregard the message. Since player i does not disregard the message only after rare histories, incentivizing player j does not affect efficiency. This answers question (i).

The answer to question (ii) is as follows: Consider the case where player i obeys the signal of player j's message and player i learns from player j's continuation play that player i's signal of player j's message was wrong. Even after realizing an error, player i keeps obeying the signal by the following reasons: By the definition of the noisy cheap talk in Section 4.3, player i believes that player j should have received E and should have realized that player i's signal was wrong. Since player j's continuation play never reveals whether player j received E or not, player i keeps this belief. As will be seen, player j after observing E makes player i indifferent between any action profile. Therefore, it is almost optimal for player i to keep obeying the signal.

Next, consider the case where player i disregards the signal of player j's message and player i learns from player j's continuation play that player j's action is different from what is consistent with player i's expectation of player j's score about player i and player i's message. For example, player i sent the message that player j should switch to defection but realizes that player j is still cooperating. This means that, if player i's message had transmitted correctly, then in order for player j to keep cooperating, player j's history should have told player j that player i has not observed a lot of good signals about player j yet (that is, player j's expectation of player i's score about player j is low). What if player j's expectation of player i's good signal about player j and player j's good signal about player i are negatively correlated and this implies that player j should have observed a lot of good signals about player i? Does player i want to switch to defection? The answer is no. Since player i's message did not transmit correctly with probability $\exp(-O(T^{\frac{1}{2}}))$, player i always attributes the inconsistency between player j's action and player i's expectation of player j's action to the error in player j's signal of player i's noisy message, rather than the mistake in player i's inference.[20]

We will define an equilibrium strategy more fully to answer the questions (i) and (ii) formally.

Full Explanation of the Strategy For each lth review round, let $X_j(l)$ be player j's score about player i in the lth review round, which denotes how many times player j observes the good signal in the lth review round.

In each lth review round, if $X_j(\tilde{l}) \leq (q_2 + 2\varepsilon) T$ for all $\tilde{l} \leq l - 1$, that is, if player j's score about player i has not been "erroneously high" in the previous review rounds, then player j monitors player i by player j's score about player i. That is, the reduction of the continuation payoff from the next review phase[21] caused by the lth review round is $\frac{g}{q_2 - q_1}((q_2 + 2\varepsilon) T - X_j(l))$. Note that this is proportional to $\frac{g}{q_2-q_1}\{(q_2 + 2\varepsilon) LT - X_j\}_+$ except that this increases *without* an upper bound within a review round.

On the other hand, if $X_j(\tilde{l}) > (q_2 + 2\varepsilon) T$ happens for some $\tilde{l} \leq l - 1$, that is, if player j's score about player i has been "erroneously high" in one of the previous review rounds, then player j stops monitoring. That is, the reduction of player i's continuation payoff from the next review phase caused by the lth review round is fixed at $gT + \frac{g}{q_2-q_1} 2\varepsilon T$. See below for how we determine this number.

[20] See 1-(b) and 2 below to make sure that after any history, there is a positive probability that player j obeys the signal of player i's message.

[21] Note that this is not a next review round.

For notational convenience, let $\lambda_j(l) = G$ denote the situation that player j monitors player i in the lth review round and let $\lambda_j(l) = B$ denote the situation that player j stops monitoring in the lth review round. That is, $\lambda_j(1) = G$ and $\lambda_j(l) = G$ if and only if $X_j(\tilde{l}) \leq (q_2 + 2\varepsilon) T$ for all $\tilde{l} \leq l - 1$.

The total reduction of the continuation payoff is

$$\frac{g}{q_2 - q_1}T + \sum_{l=1}^{L} \left(\mathbf{1}\{\lambda_j(l) = G\} \frac{g}{q_2 - q_1} \left((q_2 + 2\varepsilon) T - X_j(l)\right) + \mathbf{1}\{\lambda_j(l) = B\} \left(\frac{g}{q_2 - q_1}T + \frac{g}{q_2 - q_1} 2\varepsilon T \right) \right).$$

In general, $\mathbf{1}\{X\}$ is an index function such that

$$\mathbf{1}\{X\} = \begin{cases} 1 & \text{if } X \text{ is true,} \\ 0 & \text{if } X \text{ is not true.} \end{cases}$$

Three remarks: First, we have a constant term $\frac{g}{q_2 - q_1}T$. Note that the maximum score $X_j(l)$ for one round is T. Since the increment of the decrease in the reduction of the continuation payoff is $\frac{g}{q_2 - q_1}$, this constant term is sufficient to cover the maximum decrease of the reduction of the continuation payoff for one review round. Second, after $(q_2 + 2\varepsilon) T - X_j(l) < 0$, that is, after player j's score about player i becomes erroneously high, in the following review rounds, we have a constant positive reduction $\left(\frac{g}{q_2 - q_1}T + \frac{g}{q_2 - q_1} 2\varepsilon T \right)$. Third, from the first and second remarks, the total reduction in the continuation payoff at the beginning of the next review phase is always positive. This implies that we can find a transition probability for player j's state in the next review phase to achieve this reduction of the continuation payoff. If it were negative, then player j would need to transit to a bad strategy with a negative probability, which is infeasible.

Consider player i's incentive. If player i could know $\lambda_j(l)$, then player i wants to take cooperation (defection, respectively) constantly in the lth review round if $\lambda_j(l) = G$ ($\lambda_j(l) = B$, respectively). Verify this by backward induction: In the last Lth review round, this is true since the decrease in the reduction of the continuation payoff is always $\frac{g}{q_2 - q_1}$ (0, respectively) for an additional observation of the good signal if $\lambda_j(L) = G$ ($\lambda_j(L) = B$, respectively).

Note that with this optimal strategy, player i's payoff from the Lth review round (the instantaneous utilities in the Lth review round and the reduction of the continuation payoff caused by the Lth review round) is equal to $u_i(C,C) - \frac{g}{q_2 - q_1} 2\varepsilon T$ regardless of $\lambda_j(L)$ if player j plays cooperation.[22] That is, the reduction of the continuation payoff after $\lambda_j(L) = B$ is determined so that player i's payoff is the same between $\lambda_j(L) = G$ and $\lambda_j(L) = B$. Therefore, when we consider the $(L-1)$th review round, player i can neglect the effect of the strategy in the $(L-1)$th review round on the payoff in the Lth review round. Hence, the same argument establishes the result for the $(L-1)$th review round. We can proceed until the first review round by backward induction.

Since player i cannot observe $\lambda_j(l+1)$ directly, after the lth review round, player i wants to know whether $\lambda_j(l+1)$ is G or B. To inform player i of $\lambda_j(l+1)$, player j sends a noisy cheap talk message $m = \lambda_j(l+1)$ with precision $p = \frac{1}{2}$ at the end of each lth review round. With two players, player $i-1$ is equal to player j. If player j receives the signal $g[j](m) = E$ which implies that the communication may have an error, then player j makes player i indifferent between any action profile sequence in the following review rounds.

Intuitively, player i takes cooperation in the next review round if $f[i](\lambda_j(l+1)) = G$ and defection if $f[i](\lambda_j(l+1)) = B$. However, to incentivize player j to tell the truth without destroying efficiency of the equilibrium and to deal with the chain of learning, we need a more complicated strategy.

Specifically, after each lth review round, player i calculates the conditional belief (distribution) of $X_j(l)$ given player i's history. By the central limit theorem, given player i's history, the standard deviation of this conditional distribution is $O(T^{\frac{1}{2}})$. If the conditional expectation of $X_j(l)$ is no more than $(q_2 + \varepsilon) T$, then since $(q_2 + 2\varepsilon) T$ is far from the conditional expectation by at least εT, player i believes that player j has not observed an erroneously high score with probability at least $1 - \exp(-O(T))$. That is, player i believes that $\lambda_j(l+1) = G$ with probability at least $1 - \exp(-O(T))$.[23] Therefore, if player i's condi-

[22] As player i switches to defection after some history, player j does not always take cooperation. We will take this into account in the formal proof.

[23] Precisely speaking, $\lambda_j(l+1) = B$ if and only if player j has observed an erroneously high score in the

tional expectation of player j's score about player i is no more than $(q_2 + \varepsilon) T$, then player i will think that it is an error with probability at least $1 - \exp(-O(T))$ if player i receives $f[i](\lambda_j(l+1)) = B$.

Given the discussion above, player i will take the following strategy:

1. If player i's conditional expectation of player j's score about player i is no more than $(q_2 + \varepsilon) T$, then player i will mix the following two:

   (a) With probability $1-\eta$, player i disregards the message and believes that $\lambda_j(l+1) = G$, thinking that it is an error if player i receives $f[i](\lambda_j(l+1)) = B$.

   (b) With probability η, player i obeys player i's signal of player j's message: Player i takes cooperation in the $(l+1)$th review round if $f[i](\lambda_j(l+1)) = G$ and defection if $f[i](\lambda_j(l+1)) = B$.

2. If player i's conditional expectation of player j's score about player i is more than $(q_2 + \varepsilon) T$, then player i always obeys player i's signal of player j's message: Player i takes cooperation in the $(l+1)$th review round if $f[i](\lambda_j(l+1)) = G$ and defection if $f[i](\lambda_j(l+1)) = B$.

In addition, if 1-(b) or 2 happens, then player i makes player j indifferent between any action profile sequence.

Verify that this is an ε-equilibrium: From player i's perspective at the beginning of the $(l+1)$th review round, 1-(a) is ε-optimal by the reason explained above. For 1-(b) and 2, it is always ε-equilibrium to obey the message since whenever player i's signal is wrong: $f[i](\lambda_j(l+1)) \neq \lambda_j(l+1)$, player j receives $g[j](\lambda_j(l+1)) = E$ and makes player i indifferent between any action profile sequence with probability $1 - \exp(-O(T^{\frac{1}{2}}))$.

Does player i want to learn from player j's continuation strategy? The answer is no in ε-equilibrium.

$\tilde{l}$th review round for some $\tilde{l} \leq l$. Hence, even if player j's score in the lth review round is not erroneously high, it is possible to have $\lambda_j(l+1) = B$ when player j has observed an erroneous score before the lth review round. We will take this into account in the formal proof in Section 13.

If 1-(b) or 2 is the case for player i, then since player j's strategy is independent of $g[j](\lambda_j(l+1))$,[24] player i always believes that if $f[i](\lambda_j(l+1)) \neq \lambda_j(l+1)$, then $g[j](\lambda_j(l+1)) = E$.

If 1-(a) is the case for player i, then player i's belief on $\lambda_j(l+1) = G$ at the beginning of the $(l+1)$th review round is no less than $1 - \exp(-O(T))$. On the other hand, 1-(b) or 2 is the case for player j with probability at least η regardless of player j's history. Hence, player j obeys player j's signal of player i's message with probability at least η. Since player j's signal of player i's message is noisy, regardless of player i's true message and $g[i](\lambda_i(l+1))$, any realization of player j's signal is possible with probability at least $\exp(-O(T^{\frac{1}{2}}))$. Thus, player i believes that any action of player j happens with probability at least $\exp(-O(T^{\frac{1}{2}}))$. Since the initial belief on $\lambda_j(l+1) = G$ is $1 - \exp(-O(T))$, which is very high compared to $\exp(-O(T^{\frac{1}{2}}))$, player i will not learn from player j's continuation play in ε-equilibrium.

In other words, when player i obeys the signal, player i believes that if player i's signal is wrong, then player j should have known that. When player i disregards the message based on her inference from the review round, then whenever player i observes player j's action different from player i's expectation, player i attributes the inconsistency to an error in player j's signals, rather than to player i's inference about player j's score about player i. This is possible since the inference from the review round is precise with probability $1-\exp(-O(T))$ while the signals of the noisy cheap talk are imprecise with probability $\exp(-O(T^{\frac{1}{2}}))$.

Finally, consider player j's incentive. The incentive to tell the truth about $\lambda_j(l+1)$ is satisfied since whenever player i's signal of player j's message affects player i's continuation play, that is, if 1-(b) or 2 is the case for player i, then player i makes player j indifferent between any action profile sequence.

We also need to consider player j's incentive in the lth review round. If 1-(a) is the case, then player i cooperates and player i does not make player j indifferent between any action profile sequence. This is better than 1-(b) or 2, where player i makes player j indifferent be-

[24] As player i's continuation play is independent of $g[i](\lambda_i(l+1))$, player j's continuation play is independent of $g[j](\lambda_j(l+1))$.

tween any action profile sequence.[25] Therefore, if player j can decrease player i's conditional expectation of player j's score about player i, then player j wants to do so. We construct the good signal so that player j cannot manipulate player i's conditional expectation of player j's score about player i. That is, player j's expectation of player i's conditional expectation of player j's score about player i is constant with respect to player j's action. See (19) and (27) for the formal definition of the good signal.

Therefore, this is an ε-equilibrium.

We are left to check efficiency. An erroneously high realization of player j's score about player i or player i's conditional expectation of player j's score about player i does not occur with high probability. In addition, $g[j](m) = E$ does not happen with high probability either. Hence, if we take η (the probability that 1-(b) is the case) sufficiently small, then with high probability, player i takes cooperation for all the review rounds and player j monitors player i by

$$\frac{g}{q_2 - q_1}T + \sum_{l=1}^{L} \frac{g}{q_2 - q_1} \left(\left(q_2 + 2\varepsilon \right) T - X_j(l) \right).$$

Since the ex ante mean of $X_j(l)$ is $q_2 T$, the per-period expected reduction of the continuation payoff is $\frac{g}{q_2-q_1}\left(\frac{1}{L} + 2\varepsilon\right)$, which can be arbitrarily small for large L and small ε.

Summary Let us summarize the equilibrium construction. Although the breakdown of cooperation after erroneous histories is inevitable, we need to verify that the chain of learning does not destroy the incentives. First, we divide the long review round into L review rounds. We make sure that, in each review round, the constant action is optimal. To do so, we have a constant term $\frac{g}{q_2-q_1}T$ for the reduction of the continuation payoff. This is enough to cover the maximum decrease in the reduction of the continuation payoff in one review round. At the same time, since the length of one review round is only $\frac{1}{L}$ of the total length of the review phase, the per-period reduction of the continuation payoff from this constant term is

[25] Since player i is in the good state, when player i makes player j indifferent between any action profile sequence, player i will do so by reducing player j's continuation payoff from the next review phase so that player j's payoff (the summation of the instantaneous utilities and the reduction of the continuation payoff) is flatten at the lowest level with respect to action profiles.

sufficiently small for large L. So, this does not affect efficiency.

To inform player i of the optimal action in the next review round, player j sends a noisy message. The noise plays two roles: First, player i (the receiver) disregards the message with ex ante high probability (this is 1-(a) in the above explanation). To incentivize player j to tell the truth, player i makes player j indifferent between any action profile sequence in the following review rounds whenever player i's signal of player j's message affects player i's continuation play. Since player i disregards the message with high probability, this does not destroy efficiency. Second, since each player obeys her signal of the opponent's message with a positive probability, whenever a player observes the opponent's action different from what she expected, she thinks that this is due to an error in the noisy communication. This cut down the chain of learning.

Finally, we construct the good signal from player j's private signals such that player j's expectation of player i's conditional expectation of player j's score about player i is constant with respect to player j's action.

4.6.2 A General Two-Player Game

Now, we consider the second role of L, that is, we consider a game where the block of Hörner and Olszewski (2006) has more than one period. We still concentrate on the two-player case.

Imagine that we replace each period in Hörner and Olszewski (2006) with a T-period review round. We need L review rounds so that, when player i uses the harsh strategy, regardless of player j's deviation, we can keep player j's value low enough. If player j deviates for a non-negligible part of a review round, then by the law of large numbers, player i can detect player j's deviation with high probability. If player i minimaxes player j from the next review round after such an event, then player j can get a payoff higher than the targeted payoff only for one review round. With sufficiently long L, therefore, player j's average payoff from a review phase can be arbitrarily close to the minimax payoff.

A known problem to replace one period in Hörner and Olszewski (2006) with a review round is summarized in Remark 5 in their Section 5. Player i's optimal action in a round depends on player j's signals in the past rounds. Player i calculates the belief about player j's past signals at the beginning of the round and starts to take an action that is optimal from her belief. While player i observing signals in that round, since player j's actions depend on player j's signals in the past rounds, player i may realize that player j's actions are different from what player i expected from her belief about player j's signals. Then, player i needs to correct her belief about player j's past signals.

Realize that this is the same "chain of learning" problem as we have dealt with for $\lambda_j(l)$. Here, we will proceed as follows: Player j has a "signal to check her own deviation" which occurs less often if player j does not follow the equilibrium path. Let $G_j(l)$ be how many times player j observes this signal in the lth review round. We call $G_j(l)$ "player j's score about player j's own deviation." If the realization of $G_j(l)$ is sufficiently low, then player j allows player i to minimax player j from the next review round. Specifically, player j makes player i indifferent between any action profile sequence from the $(l+1)$th review round.[26] At the end of the lth review round, player j sends the noisy cheap talk message about whether player j will allow player i to minimax player j from the next review round.

On the other hand, at the end of the lth review round, player i calculates the conditional expectation of player j's score about player j's own deviation. With probability $1-\eta$, player i decides the action in the $(l+1)$th review round as follows:

1. If player i's conditional expectation of player j's score about player j's own deviation is very low, then player i disregards player i's signal of player j's message and will minimax player j. Player i believes that it is an error if the signal says that player j will not allow player i to minimax player j.

2. Otherwise, player i will not minimax player j. Since player j makes player i indifferent if player j allows player i to minimax, not minimaxing is always optimal.

[26] To prevent player i from manipulating whether player j makes player i indifferent, we construct player j's score about player j's own deviation so that player i cannot change player i's expectation of player j's score about player j's own deviation. This is parallel to making sure that player j cannot change player j's expectation of player i's conditional expectation of player j's score about player i.

With a small probability η, player i obeys the signal of player j's message. In this case, player i makes player j indifferent between any action profile sequence.

Then, from 1, player j will be minimaxed if player j deviates with high probability regardless of player j's message. For this procedure to trigger the punishment properly, we construct player j's score about player j's own deviation so that player i's conditional expectation of player j's score about player j's own deviation will become low if player j deviates. Hence, we can keep player j's payoff low regardless of player j's deviation both in actions and messages. Since the players obey the signals of the messages with positive probability, if player i realizes that player j's action is different from what player i expected, then player i thinks that it is due to an error in the noisy communication. Hence, the chain of learning will not be a problem in ε-equilibrium.

4.6.3 A General Game with More Than Two Players

Finally, we consider a general game with more than two players. There are two problems unique to a game with more than two players: First, if player i's state x_i is B, then player $(i+1)$'s value should be low. Since player i is in the bad state, player i can only increase the continuation payoff. That is, we cannot punish player $i+1$ by reducing the continuation payoff. Hence, players $-(i+1)$ need to minimax player $i+1$ if player $i+1$ seems to have deviated. With two players, player i is the only opponent of player $i+1$ and so it suffices for player i to unilaterally punish player $i+1$. Hence, the punishment explained in Section 4.6.2 works (note that player $i+1$ is player j in the two-player case). On the other hand, with more than two players, we need to make sure that players $-(i+1)$ can coordinate on the punishment. This coordination can be done by communication among all the players about who will be punished at the end of each review round. See the Supplemental Material 3 for the details.

Second, there will be a new problem when we dispense with the perfect cheap talk in the coordination block. We will address this issue in Section 4.7.2.

4.7 Dispensing with Special Communication Devices

We are left to dispense with the special communication devices introduced in Section 4.3. We first explain the dispensability in the two-player game and then proceed to the dispensability in the more-than-two-player game.

4.7.1 Two Players

Dispensing with the Perfect Cheap Talk for x We explain how to replace the perfect cheap talk for the coordination on x in the coordination block with messages via actions. We proceed in steps.

First, we replace the perfect cheap talk with the noisy cheap talk sending a binary message. The property of the noisy cheap talk here is the same as the one in Section 4.3. By exchanging the noisy cheap talk messages several times, each player i can construct the inference of x, denoted $x(i)$. The important properties to establish are (i) $x(i) = x$ for all i with high probability, (ii) the communication is incentive compatible, and (iii) after realizing that $x(i) \neq x(j)$, that is, after player i realizes that player i's inference is different from player j's inference, player i believes that player j should have realized there was an error in the communication and that player j has made player i indifferent between any action profile sequence in all the review rounds with high probability. See the Supplemental Material 4 for the details.

Dispensing with the Noisy Cheap Talk Second, we replace all the noisy cheap talk with messages via actions. Given the discussion above, by doing so, we can dispense with the perfect cheap talk in the coordination block and the noisy cheap talk in the review rounds. Consider the situation where player j sends a binary noisy cheap talk message $m \in \{G, B\}$ to player i with precision $p \in (0, 1)$. Again, with two players, player $i-1$ is equal to player j. Remember that the noisy cheap talk with precision p is (i) cheap, (ii) instantaneous, and (iii) precise with probability $1 - \exp(-O(T^p))$: (iii-a) $f[i](m) = m$ with probability $1 - \exp(-O(T^p))$; (iii-b) If $f[i](m) \neq m$, then $g[j](m) = E$ with probability

$1 - \exp(-O(T^p))$; (iii-c) any signal pair can occur with probability $\exp(-O(T^p))$.

Instead of sending a noisy cheap talk message, suppose that player j sends the message via actions: Player j (sender) picks two actions a_j^G and a_j^B and takes a_j^m for T^p period. Player i (receiver) takes some fixed action, say a_i^G. Player i needs to infer the message from her private signals.

There are three difficulties: The message exchange is now (i) payoff-relevant, (ii) takes time and (iii) imprecise.

Since $T^p < T$ with $p \in (0, 1)$, the length of the communication is much shorter than that of the review rounds. Therefore, we can deal with the first difficulty by changing the continuation payoffs to cancel out the differences in the instantaneous utilities. With $T^p < T$, this does not affect the equilibrium payoff, that is, the equilibrium payoff is mainly determined by the instantaneous utilities and the changes in the continuation payoffs from the T-period review rounds. (ii) In addition, $T^p < T$ implies that the second difficulty does not affect the equilibrium payoff either.

(iii) We are left to consider the third difficulty. We want to create a mapping from player j's history to $g[j](m) \in \{m, E\}$ and a mapping from player i's history to $f[i](m) \in \{G, B\}$ to preserve (iii-a), (iii-b) and (iii-c). The latter cannot depend on the true message.

The basic intuition is as follows. Suppose that player i's signals and player j's signals are correlated. Player i infers that the message is m if the empirical distribution of player i's signals is close to the true distribution under player j sending m. If player i makes a mistake, then it means that player i observes the empirical distribution of her signals that is far away from the true distribution. Since the signals are correlated, with high probability, player j also observes the empirical distribution of her signals that is far away from the ex ante distribution under m.[27] Since player j knows her own message m, player j should realize that there may be an error. That is, if player j infers $g[j](m) = E$ if the empirical distribution of player j's signals is far from the true distribution under m, then (iii-b) is

[27] This is not generically true if player i's number of signals is much larger than player j's number of signals. This corresponds to the case where $f[i](m)$ below is not well defined. See the Supplemental Material 4 for how to deal with this case.

satisfied. (iii-a) follows from the law of iterated expectation and the low of large numbers and (iii-c) follows from the full support assumption of the signal distribution.

Formally, by the law of large numbers, with very high ex ante probability, the empirical distribution of player j's signals with a message m is very close to the affine hull of the true distribution of player j's signals with respect to player i's deviation: $\text{aff}\left(\left\{\left(q_j(y_j \mid a_j^m, a_i)\right)_{y_j}\right\}_{a_i \in A_i}\right)$. Hence, if not, then player j thinks that there is an error. That is,

1. $g[j](m) = m$ if the empirical distribution of player j's signals is very close to $\text{aff}\left(\left\{\left(q_j(y_j \mid a_j^m, a_i)\right)_{y_j}\right\}_{a_i \in A_i}\right)$, and

2. $g[j](m) = E$ if it is not close to $\text{aff}\left(\left\{\left(q_j(y_j \mid a_j^m, a_i)\right)_{y_j}\right\}_{a_i \in A_i}\right)$.

On the other hand, we define $f[i](m)$ as follows:

1. Player i calculates the conditional expectation of the empirical distribution of player j's signals as if player i knew $m = G$. If this conditional expectation is not far away from $\text{aff}\left(\left\{\left(q_j(y_j \mid a_j^G, a_i)\right)_{y_j}\right\}_{a_i \in A_i}\right)$, then $f[i](m) = G$.

2. Player i calculates the conditional expectation of the empirical distribution of player j's signals as if player i knew $m = B$. If this conditional expectation is not far away from $\text{aff}\left(\left\{\left(q_j(y_j \mid a_j^B, a_i)\right)_{y_j}\right\}_{a_i \in A_i}\right)$, then $f[i](m) = B$.

Suppose that this is well defined. That is, there is no player i's history such that

- If player i calculates the conditional expectation of the empirical distribution of player j's signals as if player i knew $m = G$, then this conditional expectation is not far away from $\text{aff}\left(\left\{\left(q_j(y_j \mid a_j^G, a_i)\right)_{y_j}\right\}_{a_i \in A_i}\right)$.

- At the same time, if player i calculates the conditional expectation of the empirical distribution of player j's signals as if player i knew $m = B$, then this conditional expectation is not far away from $\text{aff}\left(\left\{\left(q_j(y_j \mid a_j^B, a_i)\right)_{y_j}\right\}_{a_i \in A_i}\right)$ either.

Then, we are done: First, (iii-a) is satisfied by the law of iterated expectations and the law of large numbers. Second, (iii-b) is satisfied. To see why, suppose that the true message is $m = G$ and that player i has $f[i](m) = B$. Then, since 1 is not the case, player i's conditional expectation of the empirical distribution of player j's signals as if player i knew $m = G$ (this is the true message) is far away from aff $\left(\left\{\left(q_j(y_j \mid a_j^G, a_i)\right)_{y_j}\right\}_{a_i \in A_i}\right)$. This implies that, by the central limit theorem, player i puts a belief no less than $1-\exp(-O(T^p))$ on the event that the empirical distribution of player j's signals is not close to aff $\left(\left\{\left(q_j(y_j \mid a_j^G, a_i)\right)_{y_j}\right\}_{a_i \in A_i}\right)$ and that $g[j](m) = E$, as desired. The case with $m = B$ is symmetric. Finally, if we make the full support assumption on the signal distribution, then (iii-c) is automatically satisfied.

See the Supplemental Material 4 for how we deal with the case where the above definition of $f[i](m)$ is not well defined.

Dispensing with the Perfect Cheap Talk and Public Randomization in the Report Block We are left to dispense with the perfect cheap talk and public randomization in the report block about h_i^{main}.

First, we replace the perfect rich cheap talk to send h_i^{main} with perfect cheap talk that can send only a binary message. We attach a sequence of binary messages to h_i^{main}. To send h_i^{main}, player i sends the sequence of binary messages corresponding to h_i^{main}. Expecting that we will replace the perfect cheap talk with messages via actions, we make sure that the number of binary messages sent is sufficiently smaller than T. Otherwise, it would be impossible to replace the cheap and instantaneous talk with payoff-relevant and taking-time messages via actions. Since each period in each review round is *i.i.d.*, it suffices that player i reports how many times player i observes an action-signal pair (a_i, y_i) for each $(a_i, y_i) \in A_i \times Y_i$ for each review round. Hence, the cardinality of the relevant history is approximately $T^{L|A_i||Y_i|}$. Since each message is binary, the number of binary messages necessary to send the relevant history is $\log_2 T^{L|A_i||Y_i|}$, which is much smaller than T.

Second, we dispense with the public randomization. Recall that we use the public randomization to determine who will report the history such that (i) ex ante (before the report block), every player has a positive probability to report the history, and that (ii) ex post (after the realization of the public randomization), there is only one player who reports the history.

To see why both (i) and (ii) are important, remember that the equilibrium strategy would be only ε-optimal without the adjustment based on the report block. Thus, to attain the exact optimality, it is important for each player in the review rounds to believe that the reward will be adjusted with positive probability. Therefore, (i) is essential.

(ii) is important because, the logic to incentivize player i to tell the truth uses the fact that player i does not know h_j^{main} (again, with two players, player $i-1$ is player j). If player i could observe a part of player j's sequential messages which partially reveal h_j^{main}, then player i may want to tell a lie.

We show that the players use their actions and private signals to establish the properties (i) and (ii), without the public randomization.

Third, we replace the perfect binary cheap talk with noisy binary cheap talk. Before doing so, we explain what property of the communication is important in the report block. The role of the report block is for player j to adjust player i's continuation payoff so that $\sigma_i(G)$ and $\sigma_i(B)$ are both exactly optimal. Since this adjustment does not affect player j's payoff, while player i sends h_i^{main}, player j (the receiver) does not care about the precision of the message. On the other hand, if player i realizes that her past messages may not have transmitted correctly in the middle of sending a sequence of messages, then we cannot pin down player i's optimal strategy after that.

Therefore, we consider *conditionally independent* noisy cheap talk such that, when player i sends $m \in \{G, B\}$, player j receives a signal $f^{\text{ci}}[j](m) \in \{G, B\}$. The message transmits correctly, that is, $f^{\text{ci}}[j](m) = m$, with high probability. Player i receives no information about $f^{\text{ci}}[j](m)$, so that player i can always believe that the message transmits correctly with high probability. Then, the truthtelling is still optimal after any history.

Finally, we replace the conditionally independent noisy cheap talk with messages via repetition of actions. Although we do not assume conditional independence of signals a

priori or do not assume that $2|Y_i| \leq |A_j||Y_j|$,[28] as long as the adjustment of the continuation payoff based on the messages is sufficiently small, we can construct a message exchange protocol such that the sender always believes that the message transmits correctly with high probability. We defer the detailed explanation to Section 44.4.1 in the Supplemental Material 4.

4.7.2 More Than Two Players

With more than two players, we follow the same step as in the two-player case to dispense with the communication devices. Each step is the same as in the two-player case with player j replaced with player $i-1$ except for how to replace the perfect cheap talk in the coordination block with the noisy cheap talk.

Recall that player i informs the other players $-i$ of x_i in the coordination block. With two players, there is only one receiver of the message. On the other hand, with more than two players, there are more than one receivers of the message. If some players infer x_i is G while the others infer x_i is B, then the action that will be taken in the review rounds may not be included in $\{a(x)\}_x$. Since we do not have any bound on player i's payoff in such a situation, it might be of player i's interest to induce this. Since we assume that the signals from the noisy cheap talk when player i sends the message to player j are private, if we let player i inform each player j of x_i separately, then player i may want to tell a lie to a subset of players. In the Supplemental Material 5, we create a message protocol so that, while the players exchange messages and infer the other players' messages from private signals in order to coordinate on x_i, there is no player who can induce a situation where some players infer x_i is G while the others infer x_i is B in order to increase her own equilibrium payoff. Yamamoto (2011) offers a procedure to achieve this goal with conditionally independent monitoring. Our contribution is a non-trivial extension of his procedure so that it is applicable to a general monitoring structure.

[28] The latter implies that we cannot use the method that Fong, Gossner, Hörner and Sannikov (2010) create $\lambda^j(y^j)$ in their Lemma 1, which preserves the conditional independence property.

5 Finitely Repeated Game

In this section, we consider a T_P-period *finitely* repeated game with auxiliary scenarios. Intuitively, a finitely repeated game corresponds to a review phase in the infinitely repeated game and auxiliary scenarios correspond to changes in continuation payoffs.

We derive sufficient conditions on strategies and auxiliary scenarios in the finitely repeated game such that we can construct a strategy in the infinitely repeated game to support v. The sufficient conditions are summarized in Lemma 1.

Let $\sigma_i^{T_P} : H_i^{T_P} \to \Delta(A_i)$ be player i's strategy in the finitely repeated game. Let $\Sigma_i^{T_P}$ be the set of all strategies in the finitely repeated game. Each player i has a state $x_i \in \{G, B\}$. In state x_i, player i plays $\sigma_i(x_i) \in \Sigma_i^{T_P}$.

In addition, locate all the players on a circle clockwise. Each player i with x_i gives an "auxiliary scenario" (or "reward function") $\pi_{i+1}(x_i, \cdot : \delta) : H_i^{T_P+1} \to \mathbb{R}$ to the left-neighbor $i+1$ (identify player $N+1$ as player 1).[29] The auxiliary scenarios are functions from player i's histories in the finitely repeated game to the real numbers.

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

$$\underline{v}_i < v_i < \bar{v}_i \tag{2}$$

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

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

$$\sigma_i(G), \sigma_i(B) \in \arg\max_{\sigma_i^{T_P} \in \Sigma_i^{T_P}} \mathbb{E}\left[\sum_{t=1}^{T_P} \delta^{t-1} u_i(a_t) + \pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta) \mid \sigma_i^{T_P}, \sigma_{-i}(x_{-i})\right]. \tag{3}$$

[29] The players are inward-looking.

2. Regardless of $x_{-(i-1)}$, the discounted average of player i's instantaneous utilities and player $(i-1)$'s auxiliary scenario on player i is equal to $\bar{v}_i$ if player $(i-1)$'s state is good ($x_{i-1} = G$) and equal to $\underline{v}_i$ if player $(i-1)$'s state is bad ($x_{i-1} = B$):

$$\frac{1-\delta}{1-\delta^{T_P}} \mathbb{E}\left[\sum_{t=1}^{T_P} \delta^{t-1} u_i(a_t) + \pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta) \mid \sigma(x)\right] = \begin{cases} \bar{v}_i \text{ if } x_{i-1} = G, \\ \underline{v}_i \text{ if } x_{i-1} = B \end{cases} \quad (4)$$

for all $x_{-(i-1)} \in \{G, B\}^{N-1}$.

Intuitively, since $\lim_{\delta \to 1} \frac{1-\delta}{1-\delta^{T_P}} = \frac{1}{T_P}$, this requires that the time average of the expected sum of the instantaneous utilities and the auxiliary scenario is close to the targeted payoffs $\underline{v}_i$ and $\bar{v}_i$.

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

$$\begin{aligned} \lim_{\delta \to 1} \frac{1-\delta}{\delta^{T_P}} \sup_{x_{i-1}, h_{i-1}^{T_P+1}} \left|\pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta)\right| &= 0, \\ \pi_i(G, h_{i-1}^{T_P+1} &: \delta) \leq 0, \\ \pi_i(B, h_{i-1}^{T_P+1} &: \delta) \geq 0. \end{aligned} \quad (5)$$

We call (5) the "feasibility constraint."

As seen in Section 4, (5) implies that player $i-1$ with $x_{i-1} = G$ can reduce player i's continuation payoff by transiting to $x_{i-1} = B$ with higher probability while player $i-1$ with $x_{i-1} = B$ can increase player i's continuation payoff by transiting to $x_{i-1} = G$ with higher probability.

We explain why these conditions are sufficient. As explained in Section 4, we see the infinitely repeated game as the repetition of T_P-period review phases.

In each review phase, each player i has two possible states $\{G, B\} \ni x_i$ and player i with state x_i takes $\sigma_i(x_i)$ in the phase. (3) implies that both $\sigma_i(G)$ and $\sigma_i(B)$ are optimal regardless of the other players' strategy. (4) implies that player i's ex ante value at the beginning of the phase is solely determined by player $(i-1)$'s state: $\sigma_{i-1}(G)$ gives a high value while $\sigma_{i-1}(B)$ gives a low value.

Here, $\pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta)$ represents the differences between player i's ex ante value given x_{i-1} at the beginning of the phase and the ex post value at the end of the phase after player $i-1$ observes $h_{i-1}^{T_P+1}$. $\pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta) = 0$ implies that the ex post value is the same as the ex ante value since player $i-1$ transits to the same state in the next phase with probability one. With $x_{i-1} = G$ (B, respectively), the smaller $\pi_i(G, h_{i-1}^{T_P+1} : \delta)$ (the larger $\pi_i(B, h_{i-1}^{T_P+1} : \delta)$, respectively), the more likely it is for player $i-1$ to transit to the opposite state B (G, respectively) in the next phase.[30] The feasibility of this transition is guaranteed by (5).

The following lemma summarizes the discussion:

Lemma 1 *For Theorem 1, it suffices to show that, for sufficiently large δ, there exist $\{\underline{v}_i, \bar{v}_i\}_{i\in I}$ with (2), T_P with $\lim_{\delta\to 1} \delta^{T_P} = 1$, $\{\{\sigma_i(x_i)\}_{x_i\in\{G,B\}}\}_{i\in I}$ and $\{\{\pi_i(x_{i-1}, \cdot : \delta)\}_{x_{i-1}\in\{G,B\}}\}_{i\in I}$ such that (3), (4) and (5) are satisfied in the T_P-period finitely repeated game.*

Proof. See the Appendix. ■

Let us specify $\underline{v}_i$ and $\bar{v}_i$. This step is the same as Hörner and Olszewski (2006). Given $x \in \{G,B\}^N$, pick 2^N action profiles $\{a(x)\}_{x\in\{G,B\}^N}$ and corresponding payoff vectors $\{w(x)\}_{x\in\{G,B\}^N}$:

$$w(x) = u(a(x)) \text{ with } x \in \{G,B\}^N. \tag{6}$$

As we have mentioned, player $(i-1)$'s state x_{i-1} refers to player i's payoff and indicates whether this payoff is strictly above or below v_i no matter what the other players' states are. That is, *player $(i-1)$'s state controls player i's payoff*. Formally,

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

[30] Here, we define $\pi_i(G, h_{i-1}^{T_P+1} : \delta)$ as the movement of the continuation payoff. On the other hand, in Section 4, we consider the reduction of the continuation payoff. Therefore, $-\pi_i(G, h_{i-1}^{T_P+1} : \delta)$ corresponds to the reduction of the continuation payoff in Section 4.

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

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

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

Below, we construct $\{\sigma_i(x_i)\}_{x_i,i}$ and $\{\pi_i(x_{i-1}, \cdot : \delta)\}_{x_{i-1},i}$ satisfying (3), (4) and (5) with $\bar{v}_i$ and $\underline{v}_i$ defined above in the finitely repeated game.

6 Coordination, Main and Report Blocks

In this section, we explain the basic structure of the T_P-period finitely repeated game. At the beginning of the finitely repeated game, there is the "coordination block." In the finitely repeated game, the players play the action profile $a(x)$ depending on the state profile $x = (x_n)_{n\in I} \in \{G, B\}^N$. Since x_i is player i's private state, player i informs the other players $-i$ of x_i by sending messages about x_i.

As seen in Section 4, we first assume that the players can communicate x via perfect cheap talk. The players take turns: Player 1 tells x_1 first, player 2 tells x_2 second, and so on until player N tells x_N. With the perfect cheap talk, this block is instantaneous and x becomes common knowledge. Second, we replace the perfect cheap talk with the noisy cheap talk. As we will see, with two players, this block is still instantaneous while with more than two players, this block now consists of many periods. More importantly, x is no longer common knowledge. Finally, we replace the noisy cheap talk with messages via actions. Since the players repeat the messages to increase the precision, this block takes time.

After the coordination block, we have "main blocks." One main block consists of a review round and a few supplemental rounds. The review round lasts T periods with

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

as seen in Section 4. After that, for each player i, each player $j \in -i$ sends messages about what is player i's optimal action in the next round. As explained in Section 4, we first assume that player j sends the messages via noisy cheap talk. With the noisy cheap talk, this message is sent instantaneously. Then, we replace the noisy cheap talk with messages via actions. Since the players repeat the messages to increase the precision, sending the messages takes time.

Let h_i^{main} be a generic element of player i's history at the end of the last main block, that is, player i's history in the coordination block and all the main blocks.

After the last main block, we have the "report block" where each player reports h_i^{main}. We first assume that the players decide who will report the history by the public randomization device and that the picked player reports h_i^{main} by the perfect cheap talk. Then, this block is instantaneous. Second, we dispense with the public randomization. Third, we replace the perfect cheap talk with conditionally independent (noisy) cheap talk. Fourth, we dispense with the conditionally independent cheap talk.

When we say $h_i^{T_P+1}$, this denotes player i's history at the end of the report block, that is, $h_i^{T_P+1}$ contains both h_i^{main} and what information player i receives about $\left(h_n^{\text{main}}\right)_{n \in I}$ in the report block.

7 Almost Optimality

As seen in Section 4, we first show that player i's strategy is "almost optimal," or that the strategy profile is "ε-equilibrium" with $\varepsilon = \exp(-O(T^{\frac{1}{2}}))$ until the end of the last main block if we neglect the report block. After that, based on the communication in the report block, player $i-1$ adjusts the reward function so that player i's strategy is exactly optimal after any history in any period of the review phase if we take the report block into account.

We divide the reward function into two parts:

$$\pi_i(x_{i-1}, h_{i-1}^{T_P+1} : \delta) = \pi_i^{\text{main}}(x_{i-1}, h_{i-1}^{\text{main}} : \delta) + \pi_i^{\text{report}}(x_{i-1}, h_{i-1}^{T_P+1} : \delta).$$

Note that $\pi_i^{\text{main}}(x_{i-1}, h_{i-1}^{\text{main}} : \delta)$ is the reward based on player $(i-1)$'s history except for the report block and that $\pi_i^{\text{report}}(x_{i-1}, h_{i-1}^{T_P+1} : \delta)$ is the reward based on player $(i-1)$'s whole history including the report block.

As a preparation to prove the existence of π_i with (3), (4) and (5), we first construct π_i^{main} such that

1. $\sigma_i(x_i)$ is "almost optimal with $\exp(-O(T^{\frac{1}{2}})) > 0$ if we ignore the report block": For all $i \in I$ and $x \in \{G, B\}^N$, for any τ and h_i^τ in the coordination and main blocks,

$$\begin{aligned} & \max_{\sigma_i \in \Sigma_i^{\text{main}}} \mathbb{E}\left[\sum_{t=1}^{T_P} \delta^{t-1} u_i(a_t) + \pi_i^{\text{main}}(x_{i-1}, h_{i-1}^{\text{main}} : \delta) \mid h_i^\tau, \sigma_i, \sigma_{-i}(x_{-i})\right] \\ & \quad -\mathbb{E}\left[\sum_{t=1}^{T_P} \delta^{t-1} u_i(a_t) + \pi_i^{\text{main}}(x_{i-1}, h_{i-1}^{\text{main}} : \delta) \mid h_i^\tau, \sigma(x)\right] \\ \leq & \exp(-O(T^{\frac{1}{2}})). \end{aligned} \tag{8}$$

   Here, Σ_i^{main} is the set of all possible strategies in the coordination and main blocks.

2. (4) and (5) are satisfied with π_i replaced with π_i^{main} (neglecting π_i^{report}).

That is, our temporary objective is to construct $\sigma_i(x_i)$ and $\pi_i^{\text{main}}(x_{i-1}, h_{i-1}^{\text{main}} : \delta)$ satisfying (8), (4) and (5). We concentrate on this problem until Section 13. After constructing such π_i^{main}, we construct the strategy in the report block and π_i^{report} such that $\sigma_i(x_i)$ and $\pi_i = \pi_i^{\text{main}} + \pi_i^{\text{report}}$ satisfy (3), (4) and (5) in Section 14.

8 A Special Case

Since the complete proof is long and complicated, in the proof of the main text, we illustrate the main structure by focusing on a special case where (i) there are two players ($N = 2$), (ii) the payoff structure is prisoners'-dilemma and for all i,

$$u_i(D,D) < v_i < u_i(C,C), \tag{9}$$

(iii) public randomization is available, (iv) perfect cheap talk is available, and (v) noisy cheap talk with precision $p \in (0,1)$ is available.

We comment on each of these five simplifications.

Two Players As we have explained in Section 4.6.3, the two-player case is special in the following two aspects: First, if player i's state x_i is B, with more than two players, players $-(i+1)$ need to coordinate on minimaxing player $i+1$ if player $i+1$ seems to have deviated. Second, when the players coordinate on x_i in the coordination block, we need to make sure that no player can induce the situation where some players infer x_i is G while the others infer x_i is B.

In the Supplemental Material 3, we concentrate on the first aspect and explain how players $-(i+1)$ coordinate on the punishment. In the Supplemental Material 3, player i sends the message about x_i via perfect cheap talk that is also public, that is, if player i sends x_i to a player, then all the players can observe the message precisely. Since all the players coordinate on the same x_i, we are free from the second problem.

In the Supplemental Material 5, we consider the second problem. We replace the perfect and public cheap talk with the noisy cheap talk and then replace the noisy cheap talk with messages via actions. We need to show that, while the players exchange messages and infer the other players' messages from private signals in order to coordinate on x_i, there is no player who can induce a situation where some players infer x_i is G while the others infer x_i is B in order to increase her own equilibrium payoff.

Below, we concentrate on the two-player case. See the Supplemental Materials 3 and 5 for the case with more than two players. Since we assume two players, let player j be player i's unique opponent.

The Prisoners' Dilemma Remember that we take $a(x)$ such that (7) holds. If (9) is the case, then we can take

$$a_i(x) \equiv \begin{cases} C_i & \text{if } x_i = G, \\ D_i & \text{if } x_i = B. \end{cases}$$

Then, it happens to be the case that $a_i(x)$ with $x_i = B$ minimaxes player j at the same time of satisfying (7).

In a general game, $a_i(x)$ with $x_i = B$ is not a minimaxing strategy. Since player i with $\sigma_i(B)$ needs to keep player j's payoff low with a non-negative reward for any strategy of player j, player i needs to switch to a minimaxing action if player i believes that player j has deviated with high probability.

For this reason, Hörner and Olszewski (2006) have a block consisting of more than one period and in each period, if player i observes a signal indicating player j's deviation, then player i switches to a minimaxing action. As explained in Section 4, in our equilibrium, if player i observes signals indicating player j's deviation in a review round, then player i minimaxes player j from the next review round. See the Supplemental Material 2 for the formal treatment of a general game with two players.

Public Randomization As mentioned in Section 4, the players use public randomization in the report block to determine who will report the history h_i^{main} such that (i) ex ante (during the main blocks), every player has a positive probability to report the history, and that (ii) ex post (after the realization of the public randomization), there is only one player who reports the history.

Specifically, we assume that the players can draw a public random variable y^p from the uniform distribution on $[0, 1]$ whenever they want.

In the Supplemental Materials 4 and 5, we show that the public randomization is dispensable and that the players use their actions and private signals to establish the properties (i) and (ii).

Perfect Cheap Talk Perfect cheap talk will be used in the coordination block to coordinate on x and in the report block to report the whole history h_i^{main}. When the sender sends a message m via perfect cheap talk, the sender does not receive any private signal while the receiver receives a perfect signal, that is, the receiver's private signal is m. Hence, we can say that the receiver observes m directly and that m becomes common knowledge.

In the Supplemental Materials 4 and 5, we show that the perfect cheap talk is dispensable. As explained in Section 4, for the coordination block, we first replace the perfect cheap talk with the noisy cheap talk and then replace the noisy cheap talk with messages via actions. For the report block, we first replace the perfect cheap talk with the conditional independence (noisy) cheap talk and then replace the conditional independence cheap talk with messages via actions.

Noisy Cheap Talk with Precision $p \in (0,1)$ We assume that each player j has an access to a noisy cheap talk device with precision $p \in (0,1)$ to send a binary message $m \in \{G, B\}$.[31] When player j sends m to player i via noisy cheap talk with precision p, it generates player i's private signal $f[i](m) \in \{G, B\}$ with the following probability:

$$\Pr(\{f[i](m) = f\} \mid m) = \begin{cases} 1 - \exp(-O(T^p)) & \text{for } f = m, \\ \exp(-O(T^p)) & \text{for } f = \{G, B\} \setminus \{m\}. \end{cases}$$

That is, $f[i](m)$ is the correct signal with high probability.

Given player i's signal $f[i](m)$, it generates player j's private signal $g[j](m) \in \{m, E\}$ with the following probability:

$$\Pr(\{g[j](m) = E\} \mid m, f[i](m)) = 1 - \exp(-O(T^p))$$

for all $(m, f[i](m))$ with $f[i](m) \neq m$. That is, if player i observes a wrong signal, then player j observes the signal E ("error") with high probability. This also implies that player j with $g[j](m) = m$ believes that $f[i](m) = m$ with probability at least $1 - \exp(-O(T^p))$

[31] Except for the Supplemental Material 4, p is always equal to $\frac{1}{2}$.

since otherwise, player j should have received $g[j](m) = E$.

We do not specify the probability for the other cases except that

- anything happens with probability at least $\exp(-O(T^p))$:

$$\Pr\left(\{(f[i](m), g[j](m)) = (f, g)\} \mid m\right) \geq \exp(-O(T^p))$$

  for all m and (f, g), and

- unconditionally on $f[i](m)$, $g[j](m) = m$ with high probability:

$$\Pr\left(\{g[j](m) = m\} \mid m\right) \geq 1 - \exp(-O(T^p))$$

  for all m.

Finally, player i observes her second private signal $f_2[i](m) \in \{G, B\}$ and player j observes her second private signal $g_2[j](m) \in \{G, B\}$. We assume that there exists $\eta > 0$ such that

- $f_2[i](m)$ and $g_2[j](m)$ are very imprecise signals compared to $f[i](m)$ and $g[j](m)$:
  - For all $f_2, m, f[i](m), g[j](m) \in \{G, B\}$,

$$\Pr\left(\{f_2[i](m) = f_2\} \mid m, f[i](m), g[j](m)\right) \geq \eta. \tag{10}$$

    By (10), after observing any $f_2[i](m)$, player i still believes that if $f[i](m) \neq m$, then $g[j](m) = E$ with probability $1 - \exp(-O(T^p))$.
  - For all $g_2, m, f[i](m), g[j](m) \in \{G, B\}$,

$$\Pr\left(\{g_2[j](m) = g_2\} \mid m, f[i](m), g[j](m)\right) \geq \eta. \tag{11}$$

    By (11), after observing any $g_2[j](m)$, player j with $g[j](m) = m$ still believes that $f[i](m) = m$ with probability $1 - \exp(-O(T^p))$.

This implies that, while we are considering almost optimality, we can ignore the second signals $f_2[i](m)$ and $g_2[j](m)$.

- $f_2[i](m)$ and $g_2[j](m)$ have some information about the other player's information:

  - For any $m \in \{G, B\}$, $g[j](m) \in \{G, B\}$, $f[i](m), f[i](m)' \in \{G, B\}$ and $f_2[i](m), f_2[i](m)' \in \{G, B\}$, if $(f[i](m), f_2[i](m)) \neq (f[i](m)', f_2[i](m)')$, then

$$\left\| \begin{array}{c} \mathbb{E}\left[\mathbf{1}_{g_2[j](m)} \mid m, g[j](m), f[i](m), f_2[i](m)\right] \\ -\mathbb{E}\left[\mathbf{1}_{g_2[j](m)} \mid m, g[j](m), f[i](m)', f_2[i](m)'\right] \end{array} \right\| > \eta. \tag{12}$$

    In this paper, we use the Euclidean norm. Here, $\mathbf{1}_{g_2[j](m)}$ is 2×1 vector such that

$$\mathbf{1}_{g_2[j](m)} = \begin{cases} \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \text{if } g_2[j](m) = G, \\ \begin{bmatrix} 0 \\ 1 \end{bmatrix} & \text{if } g_2[j](m) = B. \end{cases}$$

    This implies that, in the report block, even after knowing m and $g[j](m)$, player i who does not know $g_2[j](m)$ has the incentive to tell the truth about $(f[i](m), f_2[i](m))$. See Lemma 9 for the formal argument.

  - For any $m, m' \in \{G, B\}$, $f[i](m) \in \{G, B\}$, $g[j](m), g[j](m)' \in \{G, B\}$ and $g_2[j](m), g_2[j](m)' \in \{G, B\}$, if $(m, g[j](m), g_2[j](m)) \neq (m', g[j](m)', g_2[j](m)')$, then

$$\left\| \begin{array}{c} \mathbb{E}\left[\mathbf{1}_{f_2[i](m)} \mid m, g[j](m), g_2[j](m), f[i](m)\right] \\ -\mathbb{E}\left[\mathbf{1}_{f_2[i](m)} \mid m', g[j](m)', g_2[j](m)', f[i](m)\right] \end{array} \right\| > \eta. \tag{13}$$

Here, $\mathbf{1}_{f_2[i](m)}$ is 2×1 vector such that

$$\mathbf{1}_{f_2[i](m)} = \begin{cases} \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \text{if } f_2[i](m) = G, \\ \begin{bmatrix} 0 \\ 1 \end{bmatrix} & \text{if } f_2[i](m) = B. \end{cases}$$

This implies that, in the report block, even after knowing $f[i](m)$, player j who does not know $f_2[i](m)$ has the incentive to tell the truth about $(m, g[j](m), g_2[j](m))$. Again, see Lemma 9 for the formal argument.

The following lemma summarizes the important features of the noisy cheap talk:

Lemma 2 *The signals by the noisy cheap talk with precision $p \in (0,1)$ for player j to send $m \in \{G, B\}$ satisfy the following conditions:*

1. *For any $m \in \{G, B\}$, player i's signal $f[i](m)$ is correct with high probability:*

$$\Pr(\{f[i](m) = m\} \mid m) \geq 1 - \exp(-O(T^p)).$$

2. *For any $m \in \{G, B\}$, $f[i](m) \in \{G, B\}$ and $f_2[i](m) \in \{G, B\}$, after knowing m, $f[i](m)$ and $f_2[i](m)$, player i puts a high belief on the events that either $f[i](m)$ is correct or $g[j](m) = E$:*

$$\Pr(\{f[i](m) = m \text{ or } g[j](m) = E\} \mid m, f[i](m), f_2[i](m)) \geq 1 - \exp(-O(T^p)).$$

3. *For any $m \in \{G, B\}$ and $g_2[j](m) \in \{G, B\}$, if $g[j](m) = m$, then player j puts a high belief on the event that player i's first signal is correct:*

$$\Pr(\{f[i](m) = m\} \mid m, \{g[j](m) = m\}, g_2[j](m)) \geq 1 - \exp(-O(T^p)).$$

4. For any $m \in \{G, B\}$, any signal profile can happen with positive probability:

$$\begin{aligned} & \Pr\left(\{(f\,[i]\,(m), g[j](m), f_2[i](m), g_2[j](m)) = (f, g, f_2, g_2)\} \mid m\right) \\ \geq \; & \exp(-O(T^p)) \end{aligned}$$

for all $(f, g, f_2, g_2) \in \{G, B\}^4$.

Proof. See the discussion above. ■

Condition 1 implies that the signal is correct with high probability. Condition 2 implies that, even after player i realizes that her signal is not correct ($f[i](m) \neq m$), player i believes that player j realizes the mistake (that is, $g[j](m) = E$) with high probability, as required in Section 4. On the other hand, Condition 3 implies that, after observing $g[j](m) = m$, player j believes that player i received the correct signal (since otherwise player i should have received $g[j](m) = E$) with high probability.[32] Further, Condition 4 implies that all the players believe that any mistake happens with probability $\exp(-O(T^p)$. As seen in Section 4, this is important to incentivize the players to follow the equilibrium path after observing the opponent's action different from her expectation.

In the Supplemental Materials 4 and 5, we show that we can replace the noisy cheap talk with messages via actions, so that we can keep the important features summarized in Lemma 2.

History with Cheap Talk and Public Randomization Since the players communicate via cheap talk, the players store the signals from the cheap talk in the history. When a sender sends a message m, then the sender observes the true message and her own private signals while the receiver observes only her own private signals. With abuse of notation, when the communication is done before the players take actions in period t, we include the true message and the private signals of the sender (only the private signals of the receiver, respectively) to the history in period t of player i, h_i^t, if player i is the sender (the receiver, respectively).

[32] We use this property only in Section 39 in the Supplemental Material 4.

In addition, since the players coordinate the future play via public randomization, the players store the realization of the public randomization in the history. With abuse of notation, when a public randomization device is drawn before the players take actions in period t, we include the realization of the public randomization device to the history in period t of each player i, h_i^t.

Summary In summary, for the proof in the main text, we focus on the two-player prisoners' dilemma: $I = 2$, $A_i = \{C_i, D_i\}$ and

$$u_i(D_i, C_j) > u_i(C_i, C_j) > u_i(D_i, D_j) > u_i(C_i, D_j). \tag{14}$$

Further, we focus on v with

$$v \in \text{int}([u_1(D_1, D_2), u_1(C_1, C_2)] \times [u_2(D_2, D_1), u_2(C_2, C_1)]). \tag{15}$$

For v with (15), we can take $a(x)$, $\underline{v}_i$ and $\bar{v}_i$ such that

$$a_i(x) \equiv \begin{cases} C_i & \text{if } x_i = G, \\ D_i & \text{if } x_i = B. \end{cases} \tag{16}$$

and

$$u_i(D_1, D_2) < \underline{v}_i < v_i < \bar{v}_i < u_i(C_1, C_2). \tag{17}$$

In addition, for notational convenience, whenever we say players i and j, unless otherwise specified, i and j are different.

Finally, we assume that the perfect cheap talk, noisy cheap talk with precision $p \in (0, 1)$ and public randomization are available, all of which are shown to be dispensable in the Supplemental Material 4.

For the rest of the main paper, we prove the folk theorem in this special case: We arbitrarily fix v with (15) and then construct a strategy profile (action plans and rewards) in the finitely repeated game with (3), (4) and (5).

9 Assumption

Before proceeding to the proof, we explain what are generic conditions that we need in the special case defined above. See the Supplemental Material 1 for what are additional assumptions that we need for more general cases.

Here, we make the following three assumptions: Assumption 3 implies that the monitoring has full support. Assumption 4 guarantees that we can construct the good signal introduced in Section 4.6.1. Assumption 5 is used to give the incentive to tell the truth in the report block. As will be seen, all of them are generic under Assumption 2.

9.1 Full Support

We assume that the monitoring has full support:

Assumption 3 *$q(y \mid a) > 0$ for all $a \in A$ and $y \in Y$.*

This assumption has two implications: First, Condition 4 of Lemma 2 is satisfied when we replace the noisy cheap talk with precision p with repeatedly taking actions for T^p periods.[33]

Second, a Nash equilibrium is realization equivalent to a sequential equilibrium.[34] Therefore, for the rest of the paper, we focus on Nash equilibria.

9.2 Identifiability

To incentivize the players to follow the equilibrium path, it is important that, for each player $i \in I$ and action profile $a \in A$, her opponent j statistically identifies player i's deviation. That is, we want to create a statistics $\psi_j^a(y_j)$ whose expectation is higher when player i

[33] This property is used in the Supplemental Materials 4 and 5.
[34] See Sekiguchi (1997) and Kandori and Matsushima (1998).

follows the prescribed action a_i than $\tilde{a}_i \neq a_i$: With some $q_2 > q_1$,

$$\mathbb{E}\left[\psi_j^a(y_j) \mid \tilde{a}_i, a_j\right] \equiv \sum_{y_j} q(y_j \mid \tilde{a}_i, a_j)\psi_j^a(y_j) = \begin{cases} q_2 & \text{if } \tilde{a}_i = a_i, \\ q_1 & \text{if } \tilde{a}_i \neq a_i. \end{cases} \tag{18}$$

Further, in our equilibrium, player i calculates the conditional expectation of $\psi_j^a(y_j)$ after observing y_i, believing that a is taken:

$$\sum_{y_j} \psi_j^a(y_j) q(y_j \mid a, y_i).$$

We want to make sure that player j cannot change player j's expectation of player i's conditional expectation of player j's statistics $\psi_j^a(y_j)$ by player j's deviation: For each $\tilde{a}_j \neq a_j$,

$$\sum_{y_i} \left(\sum_{y_j} \psi_j^a(y_j) q(y_j \mid a, y_i) \right) q\left(y_i \mid a_i, \tilde{a}_j\right) = q_2. \tag{19}$$

Note that this is an equilibrium calculation, that is, player i believes that the equilibrium action a is taken. Note also that (19) is equivalent to

$$\sum_{y_j} \left(\sum_{y_i} q(y_j \mid a, y_i) q(y_i \mid a_i, \tilde{a}_j) \right) \psi_j^a(y_j) = q_2. \tag{20}$$

A sufficient condition for the existence of such ψ_j^a is as follows: Let $Q_1(\tilde{a}_i, a_j) \equiv \left(q(y_j \mid \tilde{a}_i, a_j)\right)_{y_j}$ be the vector expression of the conditional distribution of player j's signals given $\tilde{a}_i, a_j$. In addition, let $Q_2(a_i, \tilde{a}_j) \equiv \left(\sum_{y_i} q(y_j \mid a, y_i) q(y_i \mid a_i, \tilde{a}_j)\right)_{y_j}$ be the ex ante distribution of player j's signals when y_i is first generated according to $q(y_i \mid a_i, \tilde{a}_j)$ and then y_j is generated according to $q(y_j \mid a, y_i)$. We assume that all the vectors $Q_1(\tilde{a}_i, a_j)$ with $\tilde{a}_i \in A_i$ and $Q_2(a_i, \tilde{a}_j)$ with $\tilde{a}_j \neq a_j$ are linearly independent:

Assumption 4 *For any $i \in I$ and $a \in A$, $Q_1(\tilde{a}_i, a_j)$ with $\tilde{a}_i \in A_i$ and $Q_2(a_i, \tilde{a}_j)$ with $\tilde{a}_j \neq a_j$ are linearly independent.*

This assumption is generic if $|Y_j| \geq |A_i| + |A_j| - 1$. We can show that Assumption 4 is sufficient for (18) and (19).

Lemma 3 *If Assumption 4 is satisfied, then there exist $q_2 > q_1$ such that, for each $i \in I$ and $a \in A$, there exists a function $\psi_j^a : Y_j \to (0,1)$ such that (18) and (19) are satisfied.*

Proof. See the Appendix. ■

In addition, since the linear independence of $Q_1(\tilde{a}_i, a_j)$ with respect to $\tilde{a}_i$ implies that player j can statistically identify player i's action, player j can give a reward that cancels out the effect of discounting:

Lemma 4 *If Assumption 4 is satisfied, then for each $i \in I$, there exists $\pi_i^\delta : \mathbb{N} \times A_j \times Y_j \to \mathbb{R}$ such that*

$$\delta^{t-1} u_i(a_t) + \mathbb{E}\left[\pi_i^\delta(t, a_{j,t}, y_{j,t}) \mid a_t\right] = u_i(a_t) \text{ for all } a_t \in A \text{ and } t \in \{1, ..., T_P\} \tag{21}$$

and

$$\lim_{\delta \to 1} \frac{1-\delta}{1-\delta^{T_P}} \sum_{t=1}^{T_P} \sup_{a_{j,t}, y_{j,t}} \left|\pi_i^\delta(t, a_{j,t}, y_{j,t})\right| = 0 \tag{22}$$

for $T_P = O(T)$ with $T = (1-\delta)^{-\frac{1}{2}}$.

Proof. See the Appendix. ■

The intuition is straightforward. Since player j can identify player i's action, player j rewards player i if player i takes an action with a lower instantaneous utility in earlier periods rather than postponing it. Since the discount factor converges to unity, this adjustment is small. As we will see in Section 11.3, we add

$$\sum_{t=1}^{T_P} \pi_i^\delta(t, a_{j,t}, y_{j,t}) \tag{23}$$

to π_i^{main} so that we can neglect discounting within the review phase. With abuse of notation, we do not consider (5) for π_i^δ since (22) guarantees that we can always subtract (add, respectively) a small number depending on x_j to (from, respectively) π_i^{main} to make π_i^{main} negative (positive, respectively) without affecting the incentives and equilibrium payoff.

Further, since player j can statistically identify player i's action, player j can give a reward that cancels out the difference in the instantaneous utilities:

Lemma 5 *If Assumption 4 is satisfied, then, there exists $\bar{u} > 0$ such that, for each $i \in I$, there exist $\pi_i^G : A_j \times Y_j \to [-\bar{u}, 0]$ and $\pi_i^B : A_j \times Y_j \to [0, \bar{u}]$ such that*

$$\begin{aligned} u_i(a) + \mathbb{E}\left[\pi_i^G(a_j, y_j) \mid a\right] &= \textit{constant} \in [-\bar{u}, \bar{u}] \textit{ for all } a \in A, \\ u_i(a) + \mathbb{E}\left[\pi_i^B(a_j, y_j) \mid a\right] &= \textit{constant} \in [-\bar{u}, \bar{u}] \textit{ for all } a \in A. \end{aligned}$$

Proof. See the Appendix. ■

Suppose that player i's reward function is time-separable and that, in period t, the reward from period t is $\sum_{\tau \geq t} \pi_i^{x_j}(a_{j,\tau}, y_{j,\tau})$. Then, player i is indifferent between any action profile sequence from period t.

9.3 Slight Correlation

As briefly mentioned in Section 7, the reward π_i^{report} on player i is adjusted based on player i's messages about h_i^{main} in the report block. At the same time, we need to establish player i's truthtelling incentive about h_i^{main}.

When player i reports her history $(a_{i,t}, y_{i,t})$ for some period t in the coordination or main blocks, intuitively, player j punishes player i proportionally to

$$\left\| \mathbf{1}_{y_{j,t}} - \mathbb{E}\left[\mathbf{1}_{y_{j,t}} \mid \hat{a}_{i,t}, \hat{y}_{i,t}, a_{j,t}\right] \right\|^2.$$

Here, $\mathbf{1}_{y_{j,t}}$ is a $|Y_j| \times 1$ vector whose element corresponding to $y_{j,t}$ is one and other elements are zero. $(\hat{a}_{i,t}, \hat{y}_{i,t})$ is player i's message. Intuitively,[35] player i wants to minimize

$$\mathbb{E}\left[\left\|\mathbf{1}_{y_{j,t}} - \mathbb{E}\left[\mathbf{1}_{y_{j,t}} \mid \hat{a}_{i,t}, \hat{y}_{i,t}, a_{j,t}\right]\right\|^2 \mid a_{i,t}, y_{i,t}, a_{j,t}\right]. \tag{24}$$

We assume that player i knew player j's action $a_{j,t}$.[36]

We assume that a different $(a_{i,t}, y_{i,t})$ has different information about $y_{j,t}$ conditional on $a_{j,t}$:

Assumption 5 *For any $i \in I$, $a_j \in A_j$, $a_i, a_i' \in A_i$ and $y_i, y_i' \in Y_i$, if $(a_i, y_i) \neq (a_i', y_i')$, then*

$$\mathbb{E}\left[\mathbf{1}_{y_j} \mid a_i, y_i, a_j\right] \neq \mathbb{E}\left[\mathbf{1}_{y_j} \mid a_i', y_i', a_j\right].$$

Note that this excludes the conditional independence. Given Assumption 5, the truthtelling is uniquely optimal.

Lemma 6 *If Assumption 5 is satisfied, then for any $a_t \in A$ and $y_{i,t} \in Y_i$, $(\hat{a}_{i,t}, \hat{y}_{i,t}) = (a_{i,t}, y_{i,t})$ is a unique minimizer of (24).*

Proof. By algebra. ■

Take ex ante value of (24) before observing $y_{i,t}$ assuming the truthtelling $(\hat{a}_{i,t}, \hat{y}_{i,t}) = (a_{i,t}, y_{i,t})$:

$$\mathbb{E}\left[\left\|\mathbf{1}_{y_{j,t}} - \mathbb{E}\left[\mathbf{1}_{y_{j,t}} \mid a_{i,t}, y_{i,t}, a_{j,t}\right]\right\|^2 \mid a_{i,t}, a_{j,t}\right]. \tag{25}$$

From Lemma 3, we can show the existence of player j's reward on player i which cancels out the difference in (25) for different $a_{i,t}$'s:

Lemma 7 *If Assumptions 4 and 5 are satisfied, then for any $j \in I$ and $a_j \in A_j$, there exists $\Pi_i : A_j \times Y_j \to \mathbb{R}$ such that*

$$\mathbb{E}\left[\Pi_i(a_j, y_j) \mid a_i, a_j\right] = \mathbb{E}\left[\left\|\mathbf{1}_{y_j} - \mathbb{E}\left[\mathbf{1}_{y_j} \mid a_i, y_i, a_j\right]\right\|^2 \mid a_i, a_j\right]$$

[35] That is, except that player i can learn about $y_{j,t}$ from the continuation play between period t and the report block.

[36] If the incentive to tell the truth is provided assuming that player i knew player j's action, then the incentive automatically holds if player i does not know player j's action.

for all $a_i \in A_i$.

Proof. The same as Lemma 3. ■

10 Structure of the Phase

In this section, we formally define the structure of the T_P-period finitely repeated game (review phase), which is summarized in Figure 1 below. T_P depends on L and T. $L \in \mathbb{N}$ will be pinned down in Section 12 and $T = (1-\delta)^{-\frac{1}{2}}$.

As seen in Section 6, at the beginning of the phase, there is the coordination block. The players take turns to communicate x. First, player 1 sends x_1 via perfect cheap talk. Second, player 2 sends x_2 via perfect cheap talk. For notational convenience, let the round for x_i denote the moment that player i sends x_i.

After the coordination blocks, there are L "main blocks." Each of the first $(L-1)$ main blocks is further divided into three rounds. That is, for $l \in \{1, ..., L-1\}$, the lth main block consists of the following three rounds: First, the players play a T-period review round. Second, there is a supplemental round for $\lambda_1(l+1)$. Third, there is a supplemental round $\lambda_2(l+1)$. As seen in Section4, $\lambda_i(l+1) \in \{G, B\}$ is an index of whether player i has observed an "erroneous score" in the review rounds $1, \ldots, l$. In the supplemental round for $\lambda_i(l+1)$, player i sends $\lambda_i(l+1)$ via noisy cheap talk with precision $p = \frac{1}{2}$.

The last Lth main block has only the T-period review round.

Let $T(l)$ be the set of T periods in the lth review round.

After the last main block, there is the report block, where player i who is picked by the public randomization reports the whole history h_i^{main}.

Given this structure, we show that, for sufficiently large δ, with $T_P = L(1-\delta)^{-\frac{1}{2}}$, there exist $\sigma_i(x_i)$ and $\pi_i(x_j, h_j^{T_P+1} : \delta)$ satisfying (3), (4) and (5).

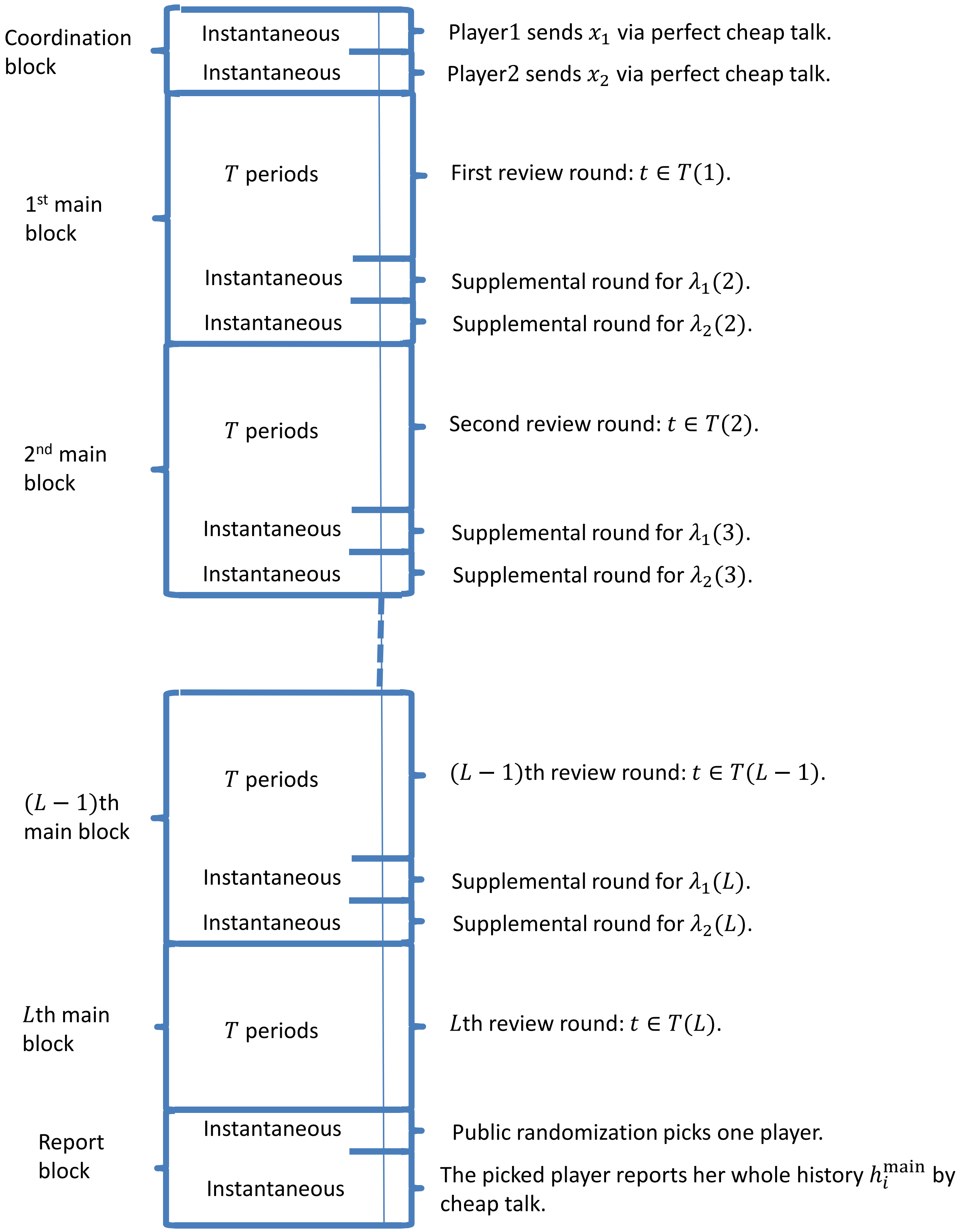

Figure 1: Structure of the Phase

11 Equilibrium Strategies

In this section, we define $\sigma_i(x_i)$ in the coordination and main blocks and $\pi_i^{\text{main}}(x_j, h_j^{\text{main}} : \delta)$. See Section 14 for the definition of $\sigma_i(x_i)$ in the report block and $\pi_i^{\text{report}}(x_j, h_j^{T_P+1} : \delta)$.

In Section 11.1, we define the state variables that will be used to define the action plans and rewards. Given the states, Section 11.2 defines the action plan $\sigma_i(x_i)$ and Section 11.3 defines the reward function $\pi_i^{\text{main}}(x_j, h_j^{\text{main}} : \delta)$. Finally, Section 11.4 determines the transition of the states defined in Section 11.1.

11.1 States x_i, $\lambda_i(l+1)$, $\hat{\lambda}_j(l+1)$, $\theta_i(l)$ and $\theta_i(\lambda_i(l+1))$

The state $x_i \in \{G, B\}$ is determined at the beginning of the review phase and fixed. With the perfect cheap talk, after player 2 sends x_2 in the coordination block, x becomes common knowledge.

As seen in Section 4, $\lambda_i(l+1) \in \{G, B\}$ is player i's state. Intuitively, $\lambda_i(l+1) = B$ implies that player i has observed an erroneous score about player j in the lth round or before. As will be formally defined in Section 11.4, $\lambda_i(l+1)$ is determined at the end of the lth review round.

On the other hand, since player j's reward on player i in the $(l+1)$th review round depends on $\lambda_j(l+1)$ as seen in Section 4, it is natural to consider player i's belief about $\lambda_j(l+1) = G$. The space for player i's possible beliefs about $\lambda_j(l+1) = G$ in each period t in the $(l+1)$th review round is $[0, 1]$ and it depends on the details of a history h_i^t. However, we classify the set of player i's histories into two partitions: The set of histories labeled as $\hat{\lambda}_j(l+1) = G$ and that labeled as $\hat{\lambda}_j(l+1) = B$. Intuitively, $\hat{\lambda}_j(l+1) = G$ ($\hat{\lambda}_j(l+1) = B$, respectively) implies that player i believes that $\lambda_j(l+1) = G$ ($\lambda_j(l+1) = B$, respectively) is likely.

To make the equilibrium tractable, $\hat{\lambda}_j(l+1)$ depends only on player i's history at the beginning of the $(l+1)$th review round and is fixed during the $(l+1)$th review block, as will be defined in Section 11.4. Further, in the $(l+1)$th review round, player i takes a constant action that depends only on $x \in \{G, B\}^2$ and $\hat{\lambda}_j(l+1) \in \{G, B\}$.

Further, as we have briefly mentioned in Sections 4, player i makes player j indifferent between any action profile after some history. If she does in the lth review round, then π_j^{main} will be $\sum_\tau \pi_j^{x_i}(a_{i,\tau}, y_{i,\tau})$ for period τ in the lth review round and after. $\theta_i(l) \in \{G, B\}$ and $\theta_i(\lambda_i(l+1)) \in \{G, B\}$ are indices of whether player i uses such a reward. See Section 11.3 for how the reward function depends on these two states and see Section 11.4 for the transition of the states.

11.2 Player i's Action

In the coordination block, player i tells the truth about x_i.

In the each lth review round, player i with $\sigma_i(x_i)$ takes $a_i(x)$ with

$$a_i(x) \equiv \begin{cases} C_i & \text{if } x_i = G, \\ D_i & \text{if } x_i = B \end{cases} \tag{26}$$

if $\hat{\lambda}_j(l) = G$ and D_i if $\hat{\lambda}_j(l) = B$. That is, if player i believes that player j has observed an erroneous score before, then player i takes D_i, a static best response to player j's action $a_j(x)$. This is the breakdown of the incentives explained in Section 4.

In the supplemental round for $\lambda_i(l+1)$, player i sends the message $\lambda_i(l+1)$ truthfully via noisy cheap talk with precision $p = \frac{1}{2}$. We assume that the players cannot manipulate p.[37] That is, in the supplemental round for $\lambda_i(l+1)$, only the noisy cheap talk with $p = \frac{1}{2}$ is available.

11.3 Reward Function

In this subsection, we explain player j's reward function on player i, $\pi_i^{\text{main}}(x_j, h_j^{\text{main}} : \delta)$.

[37] The same constraint is applicable whenever a player sends a message via noisy cheap talk with precision $p \in (0, 1)$.

Score We formally define player j's score about player i in the lth review round denoted $X_j(l)$ by which player j monitors player i for all $x \in \{G,B\}^2$.

Since the expected value of $\psi_j^{a(x)}(y_{j,t})$ in Lemma 3 increases if the players take $a(x)$, intuitively, $X_j(l)$ should be proportional to $\sum_{t\in T(l)} \psi_j^{a(x)}(y_{j,t})$.

Instead of using $\psi_j^{a(x)}(y_{j,t})$ as the actual score, we use $\psi_j^{a(x)}(y_{j,t})$ as the probability with which the score increases by 1, rather than 0.[38] For that purpose, player j constructs $\Psi_{j,t}^{a(x)} \in \{0,1\}$ from $\psi_j^{a(x)}(y_{j,t})$ as follows: After taking $a_{j,t}$ and observing $y_{j,t}$,[39] player j calculates $\psi_j^{a(x)}(y_{j,t})$. After that, player j draws a random variable from the uniform distribution on $[0,1]$. If the realization of this random variable is less than $\psi_j^{a(x)}(y_{j,t})$, then $\Psi_{j,t}^{a(x)} = 1$, and otherwise, $\Psi_{j,t}^{a(x)} = 0$. That is, $\Psi_{j,t}^{a(x)}$ is a time-independent Bernoulli random variable with mean $\psi_j^{a(x)}(y_{j,t})$. $\Psi_{j,t}^{a(x)}$ corresponds to the "good signal" in Section 4.

Player j's score about player i in the lth review round, $X_j(l)$, is equal to $\sum_{t\in T(l)} \Psi_{j,t}^{a(x)}$ except that player j does not use one random period in $T(l)$ for monitoring. That is, player j randomly picks one period $t_j(l)$ from $T(l)$: $\Pr(\{t_j(l) = t\}) = \frac{1}{T}$ for all $t \in T(l)$. Let $T_j(l) \equiv T(l) \setminus \{t_j(l)\}$ be the other periods than $t_j(l)$ in the lth review round. Player j monitors player i during $T(l)$ by the score

$$X_j(l) \equiv \sum_{t\in T_j(l)} \Psi_{j,t}^{a(x)} + \mathbf{1}_{t_j(l)}. \tag{27}$$

Here, $\mathbf{1}_{t_j(l)} \in \{0,1\}$ is a random variable with $\Pr\left(\left\{\mathbf{1}_{t_j(l)} = 1\right\}\right) = q_2$ conditional on $t_j(l)$. Hence, instead of monitoring by $\sum_{t\in T(l)} \Psi_{j,t}^{a(x)}$, player j randomly picks $t_j(l)$ and replaces $\Psi_{j,t_j(l)}^{a(x)}$ with the random variable $\mathbf{1}_{t_j(l)}$ that is independent of the players' action. Note that, since $\mathbb{E}\left[\Psi_{j,t}^{a(x)} \mid a(x)\right] = \mathbb{E}\left[\mathbf{1}_{t_j(l)}\right] = q_2$, the expected increment of $X_j(l)$ is the same for each period as long as $a(x)$ is played.

Since player j excludes $t_j(l)$ randomly, player i cannot learn player j's signal in period $t_j(l)$ by observing player j's continuation play. This plays an important role to incentivize

[38] This convert from the probability to the Bernouli random variable is the same as Fong, Gossner, Hörner and Sannikov (2010).

[39] As we have seen in Section 11.2, $a_{j,t} = a_j(x)$ as long as $\hat{\lambda}_i(l) = G$.

player i to tell the truth in the report block (see also Section 15.7). The same remark is applicable whenever we define $t_j(l)$.

Slope of the Reward Take $\bar{L}$ sufficiently large:

$$\bar{L}(q_2 - q_1) > \max_{a,i} 2|u_i(a)|. \tag{28}$$

Intuitively, $\bar{L}$ corresponds to $\frac{g}{q_2-q_1}$ in Section 4.

If the reward for the lth review round is $\bar{L}X(l)$ except for a constant and player j plays $a_j(x)$, then player i wants to take $a_i(x)$: Intuitively, for each $t \in T(l)$, the marginal increase of the probability that $\Psi_j^{a(x)} = 1$ is

$$\underbrace{\left(1 - \frac{1}{T}\right)}_{\Pr(\{t_j(l) \neq t\})} (q_2 - q_1).$$

From (28), for sufficiently large T, that is, for sufficiently large δ from (1), the expected gain from the reward dominates the gain from the instantaneous utilities.

The Reward Function The reward $\pi_i^{\text{main}}(x_j, h_j^{\text{main}} : \delta)$ is written as

$$\pi_i^{\text{main}}(x_j, h_j^{\text{main}} : \delta) = \sum_{l=1}^{L} \sum_{t \in T(l)} \pi_i^{\delta}(a_{j,t}, y_{j,t}) + \begin{cases} -\bar{L}T + \sum_{l=1}^{L} \pi_i^{\text{main}}(x, h_j^{\text{main}}, l) & \text{if } x_j = G, \\ \bar{L}T + \sum_{l=1}^{L} \pi_i^{\text{main}}(x, h_j^{\text{main}}, l) & \text{if } x_j = B. \end{cases}$$

Remember that $T(l)$ is the set of periods in the lth review round.

Five remarks. First, we add (23) so that we can neglect discounting within the review phase.

Second, since x becomes common knowledge, with abuse of notation, we let π_i^{main} depend directly on x.

Third, intuitively, $\pi_i^{\text{main}}(x, h_j^{\text{main}}, l)$ is the reward for the lth review round.

Fourth, for $x_j = G$, there exists a term $-\bar{L}T$. With $\frac{g}{q_2-q_1}$ replaced with $\bar{L}$, this corresponds to the constant term $\frac{g}{q_2-q_1}T$ in Section 4. As we will define below formally, if $X_j(\tilde{l}) \leq (q_2+2\varepsilon)T$ for all $\tilde{l} \leq l-1$, then player j's reward on player i in the lth review round is

$$\pi_i^{\text{main}}(x, h_j^{\text{main}}, l) = \text{a non-positive constant} + \bar{L}\left(X_j\left(l\right) - \left(q_2+2\varepsilon\right)T\right), \tag{29}$$

which is non-positive for $X_j(l) \leq (q_2+2\varepsilon)T$.[40] After $X_j(l) > (q_2+2\varepsilon)T$ in the lth review round, that is, after $X_j(l)$ is "erroneously high," $\pi_i^{\text{main}}(x, h_j^{\text{main}}, \tilde{l})$ will be a non-positive constant for $\tilde{l} \geq l+1$.

In summary, $\pi_i^{\text{main}}(x, h_j^{\text{main}}, l)$ is positive at most for one review round when $X_j(l)$ becomes erroneously high for the first time. Therefore, for (5), it suffices to have

$$\max_{X_j(l)} \bar{L}\left(X_j\left(l\right) - \left(q_2+2\varepsilon\right)T\right) - \bar{L}T \leq 0.$$

Since $X_j(l) \leq T$, this is satisfied.

Fifth, for $x_j = B$, symmetrically to $x_j = G$, there exists a term $\bar{L}T$. If $X_j(\tilde{l}) \geq (q_2-2\varepsilon)T$ for all $\tilde{l} \leq l-1$, then the reward in the lth review round is

$$\pi_i^{\text{main}}(x, h_j^{\text{main}}, l) = \text{a } \textit{non-negative} \text{ constant} + \bar{L}\left(X_j\left(l\right) - \left(q_2-2\varepsilon\right)T\right), \tag{30}$$

which is non-negative if $X_j(l) \geq (q_2-2\varepsilon)T$.[41] After $X_j(l) < (q_2-2\varepsilon)T$, that is, after $X_j(l)$ becomes "erroneously *low*," $\pi_i^{\text{main}}(x, h_j^{\text{main}}, \tilde{l})$ will be a *non-negative* constant for $\tilde{l} \geq l+1$.

In summary, $\pi_i^{\text{main}}(x, h_j^{\text{main}}, l)$ is *negative* at most for one review round and for (5), it suffices to have

$$\min_{X_j(l)} \bar{L}\left(X_j\left(l\right) - \left(q_2-2\varepsilon\right)T\right) + \bar{L}T \geq 0.$$

Since $X_j(l) \geq 0$, this is satisfied.

[40] "A non-positive constant" in the above reward is zero in the example in Section 4. Generally, this term is negative so that the equilibrium payoff is equal to the targeted value $\bar{v}_i$. See Section 13.2.

[41] "A non-negative constant" is introduced so that the equilibrium payoff is equal to the targeted value $\underline{v}_i$.

Reward Function for the lth Review Round Now, we can formally define $\pi_i^{\text{main}}(x, h_j^{\text{main}}, l)$ for each $l = 1, ..., L$.

In the lth review round, if $\theta_j(\tilde{l}) = B$ or $\theta_j(\lambda_j(\tilde{l}+1)) = B$ happens for some $\tilde{l} \leq l-1$, then player j makes player i indifferent between any action profile by

$$\pi_i^{\text{main}}(x, h_j^{\text{main}}, l) = \sum_{t \in T(l)} \pi_i^{x_j}(a_{j,t}, y_{j,t}). \tag{31}$$

Otherwise, that is, if $\theta_j(\tilde{l}) = \theta_j(\lambda_j(\tilde{l}+1)) = G$ for all $\tilde{l} \leq l-1$, then player j's reward on player i depends on the state profile, x, the index of the past erroneous history, $\lambda_j(l)$, and player j's score about player i in the lth review round, $X(l)$. As seen in Section 4, the reward is linearly increasing in $X_j(l)$ if $\lambda_j(l) = G$ and is flat if $\lambda_j(l) = B$:

$$\pi_i^{\text{main}}(x, h_j^{\text{main}}, l) = \begin{cases} \bar{\pi}_i(x, G, l) + \bar{L}\left(X_j(l) - (q_2 T + 2\varepsilon T)\right) & \text{if } x_j = G \text{ and } \lambda_j(l) = G, \\ \bar{\pi}_i(x, G, l) + \bar{L}\left(X_j(l) - (q_2 T - 2\varepsilon T)\right) & \text{if } x_j = B \text{ and } \lambda_j(l) = G, \\ \bar{\pi}_i(x, B, l) & \text{if } \lambda_j(l) = B. \end{cases} \tag{32}$$

Here, $\bar{\pi}_i(x, \lambda_j(l), l)$ is a constant that will be determined in Section 13.2 so that (8), (4) and (5) are satisfied. $\bar{\pi}_i(x, G, l)$ with $x_j = G$ corresponds to the non-positive constant in (29) and that with $x_j = B$ corresponds to the non-negative constant in (30).

11.4 Transition of the States

In this subsection, we explain the transition of the players' states. Since x_i is fixed in the phase, we consider the following four states:

11.4.1 Transition of $\lambda_j(l+1) \in \{G, B\}$

As mentioned in Section 4, $\lambda_j(l+1) \in \{G, B\}$ is player j's index of the past erroneous history. Here, we consider player j's state rather than player i's state since later, we will consider how player j's reward on player i incentivizes player i to take $\sigma_i(x_i)$. Since player j's reward is affected by $\lambda_j(l+1)$, it is more convenient to summarize the transition of $\lambda_j(l+1)$ rather than $\lambda_i(l+1)$.

The initial condition is $\lambda_j(1) = G$. Inductively, given $\lambda_j(l) \in \{G, B\}$, $\lambda_j(l+1)$ is determined as follows: If $\lambda_j(l) = B$, then $\lambda_j(l+1) = B$. That is, once $\lambda_j(l) = B$ happens, it lasts until the end of the phase. If $\lambda_j(l) = G$, then $\lambda_j(l+1) = B$ if and only if the score in the lth review round is "erroneous." That is,

1. If
$$X_j(l) \in [q_2 T - 2\varepsilon T, q_2 T + 2\varepsilon T],$$
then $\lambda_i(l+1) = G$.

2. If
$$X_j(l) \notin [q_2 T - 2\varepsilon T, q_2 T + 2\varepsilon T],$$
then $\lambda_i(l+1) = B$.

Compared to the above explanation, both erroneously low and erroneously high scores induce $\lambda_i(l+1) = B$ regardless of x_j. This lightens the notation since we do not need to analyze different transitions for different x_j's.

11.4.2 Transition of $\hat{\lambda}_j(l+1) \in \{G, B\}$

As we have mentioned in Section 11.1, $\hat{\lambda}_j(l+1) \in \{G, B\}$ is the partition of player i's histories. Intuitively, player i believes that $\lambda_j(l+1) = \hat{\lambda}_j(l+1)$ with high probability.

Since $\lambda_j(1) = G$ is common knowledge, define $\hat{\lambda}_j(1) = G$. We define $\hat{\lambda}_j(l)$ inductively. If $\hat{\lambda}_j(l) = B$, then $\hat{\lambda}_j(l+1) = B$. Hence, once $\hat{\lambda}_j(l) = B$ happens, it lasts until the end of the phase. If $\hat{\lambda}_j(l) = G$, then $\hat{\lambda}_j(l+1) \in \{G, B\}$ is defined as follows.

Suppose that $\hat{\lambda}_j(l) = G$ would be a correct inference (not always in private monitoring). Then, $\lambda_j(l+1)$ is determined as

$$\lambda_j(l+1) = \begin{cases} G & \text{if } X_j(l) \in [q_2 T - 2\varepsilon T, q_2 T + 2\varepsilon T] \\ B & \text{if } X_j(l) \notin [q_2 T - 2\varepsilon T, q_2 T + 2\varepsilon T] \end{cases} \tag{33}$$

with $X_j(l) = \sum_{t\in T_j(l)} \Psi_{j,t}^{a(x)} + \mathbf{1}_{t_j(l)}$. Therefore, it is natural to consider the conditional expectation of $X_j(l)$:

$$\mathbb{E}\left[X_j(l) \mid a(x), \{y_{i,t}\}_{t\in T(l)}\right].$$

Here, we assume that player j would play $a_j(x)$ in the lth review round (not always if $\hat{\lambda}_i(l) = B$).[42]

Intuitively, player i calculates $\mathbb{E}\left[X_j(l) \mid a(x), \{y_{i,t}\}_{t\in T(l)}\right]$. Then, in the supplemental round for $\lambda_j(l+1)$, player j sends $\lambda_j(l+1)$ via noisy cheap talk with precision $p = \frac{1}{2}$ and player i receives a signal $f[i](\lambda_j(l+1))$.[43] Based on $\mathbb{E}\left[X_j(l) \mid a(x), \{y_{i,t}\}_{t\in T(l)}\right]$ and $f[i](\lambda_j(l+1))$, player i constructs $\hat{\lambda}_j(l+1)$.

Formally, instead of using the conditional expectation $\mathbb{E}\left[X_j(l) \mid a(x), \{y_{i,t}\}_{t\in T(l)}\right]$, we consider

$$\sum_{t\in T_i(l)} \mathbb{E}\left[\Psi_{j,t}^{a(x)} \mid a(x), y_{i,t}\right] + q_2. \tag{34}$$

Notice that player i calculates the summation of the conditional expectation of $\Psi_{j,t}^{a(x)}$ over $T_i(l)$ (the set of periods that player i uses to monitor player j), not $T_j(l)$ (the set of periods that player j uses to monitor player i).[44] Since $T_i(l)$ and $T_j(l)$ are different at most for two periods, this difference is negligible for almost optimality (8).

Further, instead of using $\mathbb{E}\left[\Psi_{j,t}^{a(x)} \mid a(x), y_{i,t}\right]$, player i constructs $(E_i\Psi_j^{a(x)})_t \in \{0,1\}$ as follows: After taking $a_i(x)$ and observing $y_{i,t}$, player i calculates $\mathbb{E}\left[\Psi_{j,t}^{a(x)} \mid a(x), y_{i,t}\right]$. After that, player i draws a random variable from the uniform distribution on $[0,1]$. If the realization of this random variable is less than $\mathbb{E}\left[\Psi_{j,t}^{a(x)} \mid a(x), y_{i,t}\right]$, then $(E_i\Psi_j^{a(x)})_t = 1$ and otherwise, $(E_i\Psi_j^{a(x)})_t = 0$. Let

$$E_iX_j(l) = \sum_{t\in T_i(l)} (E_i\Psi_j^{a(x)})_t + q_2.$$

[42] See Section 11.4.4 for the reason why player i can always believe that player j takes $a_j(x)$.

[43] By (10), we can neglect $f_2[i](\lambda_j(l+1))$ for almost optimality.

[44] The term q_2 reflects the fact that the expected value of $\mathbf{1}_{t_j(l)}$ is q_2.

Since

$$\Pr\left(\left\{(E_i\Psi_j^{a(x)})_t = 1\right\} \mid a_t, y_t\right) = \mathbb{E}\left[\Psi_{j,t}^{a(x)} \mid a(x), y_{i,t}\right]$$

for all a_t and y_t given $\mathbb{E}\left[\Psi_{j,t}^{a(x)} \mid a(x), y_{i,t}\right]$, the *ex post* probability *given* $\{a_t, y_t\}_{t\in T(l)}$ of the event that

$$\left|\sum_{t\in T_i(l)} \mathbb{E}\left[\Psi_{j,t}^{a(x)} \mid a(x), y_{i,t}\right] + q_2 - E_i X_j(l)\right| \leq \frac{1}{4}\varepsilon T \tag{35}$$

is $1-\exp(-O(T))$ by the law of large numbers.

Consider player i's belief about $\lambda_j(l+1)$ in the case with (35) and

$$E_i X_j(l) \in [q_2 T - \frac{1}{2}\varepsilon T, q_2 T + \frac{1}{2}\varepsilon T]. \tag{36}$$

Since $T_i(l)$ and $T_j(l)$ are different only for two periods, player i has

$$\mathbb{E}\left[X_j(l) \mid a(x), \{y_{i,t}\}_{t\in T(l)}, T_i(l), T_j(l)\right] \in [q_2 T - \varepsilon T, q_2 T + \varepsilon T]. \tag{37}$$

Given $a(x)$, $\{y_{i,t}\}_{t\in T(l)}$, $T_i(l)$ and $T_j(l)$, the conditional distribution of $X_j(l)$ is approximately a normal distribution with the mean (37) and a standard deviation $O(T^{\frac{1}{2}})$ by the central limit theorem. Since (37) implies that the conditional mean is inside of $[q_2 T - 2\varepsilon T, q_2 T + 2\varepsilon T]$ at least by

$$\varepsilon T = T^{\frac{1}{2}} \times \underbrace{O(T^{\frac{1}{2}})}_{\text{the order of the standard deviation}},$$

player i believes that $X_j(l) \in [q_2 T - 2\varepsilon T, q_2 T + 2\varepsilon T]$ with probability $1-\exp(-O(T))$. Hence, at the end of the lth review round (before receiving signals from the noisy cheap talk message in the supplemental rounds), player i believes that $\lambda_j(l+1) = G$ with probability $1-\exp(-O(T))$.

Consider the posterior after receiving $f[i](\lambda_j(l+1))$. Since player i knows that errors occur with probability $\exp(-O(T^{\frac{1}{2}})) \gg \exp(-O(T))$ in the communication via noisy cheap talk, player i keeps a belief

$$1 - \exp(-O(T)) \tag{38}$$

on $\lambda_j(l+1) = G$ regardless of $f[i](\lambda_j(l+1))$. Especially, if the signal is $f[i](\lambda_j(l+1)) = B$, then player i believes that it is an error with high probability.

Further, we need to consider the posterior after observing player j's continuation play in the following review rounds. Section 13.1 offers the proof that learning from player j's continuation play does not change the belief so much.

On the other hand, forget about the belief about $\lambda_j(l+1)$ and suppose that player i could know $\lambda_j(l+1)$ (she cannot in private monitoring). Consider the two possible realizations of the signals in the supplemental round for $\lambda_j(l+1)$. If $f[i](\lambda_j(l+1)) = \lambda_j(l+1)$, then player i receives a correct message. If $f[i](\lambda_j(l+1)) \neq \lambda_j(l+1)$, then with probability $1 - \exp(-O(T^{\frac{1}{2}}))$, player j should receive the signal telling that player i did not receive the correct signal, that is, $g[j](\lambda_j(l+1)) = E$. If $g[j](\lambda_j(l+1)) = E$, then player j will use the reward (31) for $\tilde{l} \geq l+1$ and any action will be optimal for player i as we will see in Section 11.4.3.

Therefore, regardless of $\lambda_j(l+1)$ and $f[i](\lambda_j(l+1))$, if player i uses

$$\hat{\lambda}_j(l+1) = f[i](\lambda_j(l+1)), \tag{39}$$

then $\sigma_i(x_i)$ defined in Section 11.2 is almost optimal. Note that player j's action plan in the main blocks is independent of $g[j](\lambda_j(l+1))$ and that player i cannot learn $g[j](\lambda_j(l+1))$ from player i's history in the main blocks.

Given the discussion above, we consider the following transition of $\hat{\lambda}_j(l+1)$:

1. If (35) and (36) are satisfied, then player i randomly picks the following two procedures:

   (a) With large probability $1-\eta$, player i disregards $f[i](\lambda_j(l+1))$ and has $\hat{\lambda}_j(l+1) = G$. This is almost optimal from (38).

   (b) With small probability $\eta > 0$, player i will obey $f[i](\lambda_j(l+1))$: $\hat{\lambda}_j(l+1)$ is determined by (39). This is almost optimal since if $f[i](\lambda_j(l+1)) \neq \lambda_j(l+1)$, then $g[j](\lambda_j(l+1)) = E$ with probability $1 - \exp(-O(T^{\frac{1}{2}}))$.

2. If (35) is not satisfied or (36) is not satisfied, then $\hat{\lambda}_j(l+1)$ is determined by (39). As 1-(b) above, this is almost optimal.

For concreteness, we define that player i who has deviated before the beginning of the $(l+1)$th review round determines $\hat{\lambda}_j(l+1)$ by (39).

11.4.3 Transition of $\theta_i(l) \in \{G, B\}$ and $\theta_i(\lambda_i(l+1)) \in \{G, B\}$

As we have seen in Section 11.3, $\theta_i(\tilde{l}) = B$ or $\theta_i(\lambda_i(\tilde{l}+1)) = B$ with $\tilde{l} \leq l-1$ implies that player j is indifferent between any action profile (except for the incentives from π_j^{report}).

$\theta_i(l) = G$ if 1-(a) is the case when player i creates $\hat{\lambda}_j(l+1)$ in Section 11.4.2. On the other hand, $\theta_i(l) = B$ if 1-(b) or 2 is the case.

After sending $\lambda_i(l+1)$ via noisy cheap talk message in the supplemental round for $\lambda_i(l+1)$, if player i receives the signal that player j may receive a wrong signal, that is, if $g[i](\lambda_i(l+1)) = E$, then $\theta_i(\lambda_i(l+1)) = B$. Otherwise, that is, if $g[i](\lambda_i(l+1)) = \lambda_i(l+1)$, then $\theta_i(\lambda_i(l+1)) = G$.

11.4.4 Summary of the Transitions of θ_j

We summarize the implications of the transitions of θ_j. Since we want to consider player i's incentive, we consider θ_j, not θ_i.

First, player j makes player i almost indifferent between any action profile after receiving $g[j](\lambda_j(l+1)) = E$. Since player i believes that, whenever her signal is wrong: $f[i](\lambda_j(l+1)) \neq \lambda_j(l+1)$, player j receives $g[j](\lambda_j(l+1)) = E$ and so $\theta_i(\lambda_i(l+1)) = B$ with high probability. Therefore, (39) is an almost optimal inference.

Second, consider how player j constructs $\hat{\lambda}_i(l+1)$. Reversing the indices i and j in Sections 11.4.2, whenever player j uses the signal of the noisy cheap talk $f[j](\lambda_i(l+1))$, $\theta_j(l) = B$ or $\theta_j(\lambda_j(l+1)) = B$ happens. This implies that player j will use the reward (31) and player i is indifferent between any action of player j from the $(l+1)$th review round, whenever player i's message has an impact on player j's continuation action plan. Hence, player i is almost indifferent between any message.

Third, from Section 11.2, player j does not take $a_j(x)$ in the lth review round only if $\hat{\lambda}_i(l)$ is not equal to G. Section 11.4.2 implies that player j has $\hat{\lambda}_i(l) \neq G$ only if 1-(b) or 2 is the case in Section 11.4.2 for some review round $\tilde{l} \leq l-1$. Hence, $\hat{\lambda}_i(l) \neq G$ implies that $\theta_j(\tilde{l}) = B$ or $\theta_j(\lambda_j(\tilde{l}+1)) = B$ with $\tilde{l} \leq l-1$. Therefore, when player i calculates the conditional expectation of $X_j(l)$, for almost optimality, player i can assume that player j takes $a_j(x)$. See the proof of Lemma 8 for the formal argument.

Fourth, suppose that $\theta_j(\tilde{l}) = \theta_j(\lambda_j(\tilde{l}+1)) = G$ for all $\tilde{l} \leq l-1$ (otherwise, player i is indifferent between any action profile except for π_i^{report}). From the third observation, player j takes $a_j(x)$. We will show that the joint distribution of $\theta_j(l)$ and $\theta_j(\lambda_j(l+1))$ is independent of player i's action in the lth review round with probability no less than $1 - \exp(-O(T^{\frac{1}{2}}))$. To see why, consider the following three reasons:

1. The event that (35) is not satisfied happens with the *ex post* probability $\exp(-O(T))$ *given* $\{a_t, y_t\}_{t \in T(l)}$, as we have verified in Section 11.4.2.

2. Suppose that (35) is the case. $\theta_j(l) = B$ if (i) (36) is not satisfied or (ii) (36) is satisfied and 1-(b) happens in Section 11.4.2 (with the roles of players i and j reversed).

   Since player j takes $a_j(x)$, from Lemma 3 (with the roles of players i and j reversed), the distribution of $\left(E_j \Psi_i^{a(x)}\right)_t$ is independent of player i's action. Therefore, whether (36) is satisfied or not is independent of player i's action.

   Conditional on that (36) is satisfied, whether 1-(a) or 1-(b) is the case depends on player j's mixture and is independent of player i's action.

3. $\theta_j(\lambda_j(l+1)) = B$ if and only if $g[j](\lambda_j(l+1)) = E$, which happens with probability $\exp(-O(T^{\frac{1}{2}}))$ for all $\lambda_j(l+1)$.[45]

[45] The coefficient of $T^{\frac{1}{2}}$ in the explicit expression of $O(T^{\frac{1}{2}})$ can depend on $\lambda_j(l+1)$ and $\lambda_j(l+1)$ may depend on player i's strategy. However, the probability of $g[j](\lambda_j(l+1)) = E$ is $\exp(-O(T^{\frac{1}{2}}))$ regardless of $\lambda_j(l+1)$ and so almost independent of $\lambda_j(l+1)$.

12 Variables

In this section, we show that all the variables can be taken consistently satisfying all the requirements that we have imposed: q_2, q_1, $\bar{u}$, $\bar{L}$, L, η and ε.

First, q_1 and q_2 are determined in Lemma 3 and $\bar{u}$ is determined in Lemma 5, independently of the other variables. These are determined by the precision of the monitoring.

Given q_1 and q_2, we define $\bar{L}$ to satisfy (28):

$$\bar{L}\left(q_2 - q_1\right) > \max_{a,i} 2\left|u_i\left(a\right)\right|. \tag{40}$$

This implies that the marginal expected increase of the continuation payoff by taking $a_i(x)$ is sufficiently higher than the gain of the deviation in the instantaneous utilities.

Given $\bar{L}$, we take $L \in \mathbb{N}$ sufficiently large and $\varepsilon > 0$ sufficiently small such that

$$u_i\left(D_1, D_2\right) + \frac{\bar{L}}{L} + 2\varepsilon\bar{L} < \underline{v}_i < \overline{v}_i < u_i\left(C_1, C_2\right) - \frac{\bar{L}}{L} - 2\varepsilon\bar{L}.$$

From (17), this is feasible. As seen in Section 4, a large L makes it possible to incentivize the players to take a constant action within a review round without destroying efficiency. A small ε together with large T implies that we can make sure that erroneous histories do not occur frequently without affecting efficiency.

Given $\bar{u}$, $\bar{L}$, L and ε, take η sufficiently small such that

$$\begin{aligned} & u_i\left(D_1, D_2\right) + \frac{\bar{L}}{L} + 2\varepsilon\bar{L} + L\eta\left(\bar{u} - \min_{i,x} w_i\left(x\right)\right) \\ < \; & \underline{v}_i < \overline{v}_i < u_i\left(C_1, C_2\right) - \frac{\bar{L}}{L} - 2\varepsilon\bar{L} - L\eta\left(\bar{u} + \max_{i,x} w_i\left(x\right)\right). \end{aligned} \tag{41}$$

As explained in Section 4, a small η implies that the event that player i needs to incentivize player j to tell the truth in the supplemental rounds does not occur too often.

Since $T_P = LT$ and $T = (1-\delta)^{-\frac{1}{2}}$, we have

$$\lim_{\delta \to 1} \delta^{T_P} = 1.$$

Therefore, discounting for the payoffs in the next review phase goes to zero.

13 Almost Optimality of $\sigma_i(x_i)$

Since we have defined $\sigma_i(x_i)$ and π_i^{main} except for $\bar{\pi}_i(x, \lambda_j(l), l)$, we now show that if we properly define $\bar{\pi}_i(x, \lambda_j(l), l)$, then $\sigma_i(x_i)$ and π_i^{main} satisfy (8), (4) and (5).

13.1 Almost Optimality of $\hat{\lambda}_j(l)$

First, we show the "almost optimality of $\hat{\lambda}_j(l)$." Let $\alpha_j(l)$ be player j's action plan in the lth review round[46] and $\boldsymbol{\alpha}_j(l) = (\alpha_j(1), ..., \alpha_j(l))$ be the sequence of player j's action plans from the first review round to the lth review round (excluding what messages player j sent via noisy cheap talk in the supplemental rounds). We want to show that, for any lth review round, for any h_i^t with period t in the lth review round, conditional on $\boldsymbol{\alpha}_j(l)$, player i puts a belief no less than $1 - \exp(-O(T^{\frac{1}{2}}))$ on the events that $\lambda_j(l) = \hat{\lambda}_j(l)$ or there exists $\tilde{l} \leq l-1$ with $\theta_j(\tilde{l}) = B$ or $\theta_j(\lambda_j(\tilde{l}+1)) = B$ (and so player i is indifferent between any action profile except for π_i^{report}).

Lemma 8 *For any lth review round, for any h_i^t with period t in the lth review round, conditional on $\boldsymbol{\alpha}_j(l)$, player i puts a belief no less than*

$$1 - \exp(-O(T^{\frac{1}{2}})) \tag{42}$$

on the events that $\lambda_j(l) = \hat{\lambda}_j(l)$ or there exists $\tilde{l} \leq l-1$ with $\theta_j(\tilde{l}) = B$ or $\theta_j(\lambda_j(\tilde{l}+1)) = B$.

[46] Note that player j takes an *i.i.d.* action plan within a review round. We use α_j instead of a_j since player j may take a mixed strategy to minimax player i in a general game. See the Supplemental Materials 2 and 3.

Proof. See the Appendix. ■

Since the statement is correct after conditioning on player j's action plan in the lth review round, learning about player j's action from player i's signals in the lth review round is irrelevant.

Let us illustrate the main logic by concentrating on $l = 2$. It is common knowledge that the players took $a(x)$ in the first review round.

From the discussion in Section 11.4.2, we know the following:

1. If 1-(a) is the case in Section 11.4.2, then at the end of the first review round, player i puts a belief no less than $1 - \exp(-O(T))$ on the event that $\lambda_j(2) = G$.

2. If 1-(b) or 2 is the case, then at the beginning of the second round, given $\lambda_j(2)$, player i puts a conditional belief no less than $1-\exp(-O(T^{\frac{1}{2}}))$ on the events that $\lambda_j(2) = \hat{\lambda}_j(2)$ or "$g[j]\left(\lambda_j(2)\right) = E$ and $\theta_j(\lambda_j(2)) = B$."

Conditional on $\lambda_j(2)$, player i will not learn anything about $\theta_j(\lambda_j(2))$ from player j's strategy in the main blocks. Hence, if Case 2 is the case, then Lemma 8 is true.

Suppose that Case 1 happens. We consider how the belief changes after receiving the signals from the noisy cheap talk messages in the supplemental rounds for $\lambda_1(2)$ and $\lambda_2(2)$ and after learning player j's action in the second review round. Since player j's action in the second review round is determined solely by x and $\hat{\lambda}_i(2)$, we want to show that, given $f[i](\lambda_j(2))$, $f_2[i](\lambda_j(2))$, x and $\hat{\lambda}_i(2)$, player i puts a belief no less than $1 - \exp(-O(T^{\frac{1}{2}}))$ on the events that $\lambda_j(2) = G$ or "$\theta_j(1) = B$ or $\theta_j(\lambda_j(2)) = B$."

Consider learning from the signals $f[i](\lambda_j(2))$ and $f_2[i](\lambda_j(2))$. Since any pair happens with probability at least $\exp(-O(T^{\frac{1}{2}}))$ by Lemma 2, the upper bound of changes in the likelihood by observing the signals is $\exp(O(T^{\frac{1}{2}}))$.

Next, consider learning from $\hat{\lambda}_j(2)$. Remember that 1-(b) or 2 is the case with probability at least η when player j constructs $\hat{\lambda}_i(2)$ in Section 11.4.2 (the roles of players i and j are reversed). If 1-(b) or 2 is the case, then given player i's message $\lambda_i(2)$ and player i's signals about player j's signals $g[i](\lambda_i(2))$ and $g_2[i](\lambda_i(2))$, player i believes that any $f[j](\lambda_i(2))$

happens with probability at least $\exp(-O(T^{\frac{1}{2}}))$ by Lemma 2. Therefore, the upper bound of changes in the likelihood by observing $\hat{\lambda}_i(2)$ is $\frac{1-\eta}{\eta}\exp(O(T^{\frac{1}{2}}))$.

In total, learning from $f[i](\lambda_j(2))$, $f_2[i](\lambda_j(2))$ and $\hat{\lambda}_i(2)$ changes the likelihood by

$$\exp(O(T^{\frac{1}{2}}))\frac{1-\eta}{\eta}\exp(O(T^{\frac{1}{2}})) = \exp(O(T^{\frac{1}{2}})).$$

Since the prior on $\lambda_j(2) = G$ is $1-\exp(-O(T))$, the posterior on $\lambda_j(2) = G$ after learning $f[i](\lambda_j(2))$, $f_2[i](\lambda_j(2))$ and $\hat{\lambda}_i(2)$ (and so after learning the signals from the noisy cheap talk messages in the supplemental rounds for $\lambda_1(2)$ and $\lambda_2(2)$ and player j's action in the second review round) is at least

$$1-\exp(-O(T))\exp(O(T^{\frac{1}{2}})) = 1-\exp(-O(T)).$$

Therefore, Lemma 8 holds for Case 1.

See the Appendix for the proof with $l \geq 3$. The additional difficulty is that the players do not always take $a(x)$ from the second review rounds.

13.2 Determination of $\bar{\pi}_i(x, \lambda_j(l), l)$

Second, based on Lemma 8, we determine $\bar{\pi}_i(x, \lambda_j(l), l)$ such that $\sigma_i(x_i)$ and π_i^{main} satisfy (8), (4) and (5):

Proposition 1 *For sufficiently large δ, there exists $\bar{\pi}_i(x, \lambda_j(l), l)$ such that*

1. *$\sigma_i(x_i)$ is almost optimal: For each $l \in \{1, ..., L\}$,*

   (a) *For any period t in the lth review round, (8) holds.*

   (b) *When player i sends the noisy cheap talk message about $\lambda_i(l+1)$, (8) holds.*[47]

[47] With $l = L$, this is redundant.

2. *(4) is satisfied with π_i replaced with π_i^{main}. Since each $x_i \in \{G, B\}$ gives the same value conditional on x_j, the strategy in the coordination block is optimal.*[48]

3. *π_i^{main} satisfies (5).*

Proof. See the Appendix. ■

Here, we offer the intuitive explanation. First, we construct $\bar{\pi}_i(x, \lambda_j(l), l)$, assuming that the players follow $\sigma_i(x_i)$. We want to make sure that $\bar{\pi}_i(x, \lambda_j(l), l) \leq 0$ ($\bar{\pi}_i(x, \lambda_j(l), l) \geq 0$, respectively) if $x_j = G$ ($x_j = B$, respectively) and that player i's value from the lth review round

$$\sum_{t \in T(l)} u_i(a_t) + \pi_i^{\text{main}}(x, h_j^{\text{main}}, l)$$

is close to $\bar{v}_i$ ($\underline{v}_i$, respectively) if $x_j = G$ ($x_j = B$, respectively), $\hat{\lambda}_j(l) = \lambda_j(l)$ and $\theta_j(\tilde{l}) = \theta_j(\lambda_j(\tilde{l}+1)) = G$ for all $\tilde{l} \leq l-1$. Note that the last condition implies that $\hat{\lambda}_i(l) = G$ and that player j takes $a_j(x)$. Remember that in Section 4.6.1, the reward $gT + \frac{g}{q_2 - q_1} 2\varepsilon T$ after $\lambda_j(l) = B$ is determined so that player i's value is independent of whether $\lambda_j(l)$ is equal to G or B, as long as $\hat{\lambda}_j(l) = \lambda_j(l)$.

If $\hat{\lambda}_j(l) = \lambda_j(l) = G$, then the players take $a(x)$ and the average instantaneous utility during the lth review round is $w_i(x)$.[49] On the other hand, the ex ante mean of the score $X_j(l)$ is $q_2 T$ and the expected reward is close to 0 except for $\bar{\pi}_i(x, \lambda_j(l), l)$. Therefore, from (41), if $x_j = G$ ($x_j = B$, respectively), then there exists $\bar{\pi}_i(x, \lambda_j(l), l) \leq 0$ ($\bar{\pi}_i(x, \lambda_j(l), l) \geq 0$, respectively) such that player i's value from the lth review round is close to $\bar{v}_i$ ($\underline{v}_i$, respectively).

If $x_j = G$ and $\hat{\lambda}_j(l) = \lambda_j(l) = B$, then player i takes a best response to player j's action $a_j(x)$ and the average instantaneous utility during the lth review round is more than $w_i(x)$. The reward is 0 except for $\bar{\pi}_i(x, \lambda_j(l), l)$. Therefore, from (41), player i's value from the lth review round is close to $\bar{v}_i$ if we properly determine $\bar{\pi}_i(x, \lambda_j(l), l) \leq 0$.

[48] This is not precise since we will further adjust the reward function based on the report block. However, as we will see, even after the adjustment of the report block, any $x_i \in \{G, B\}$ still gives exactly the same value and so the strategy in the coordination block is exactly optimal.

[49] With abuse of language, we take the limit where δ goes to one.

If $x_j = B$, $\hat{\lambda}_i(l) = G$, and $\hat{\lambda}_j(l) = \lambda_j(l) = B$, then player i takes a best response to player j's action $a_j(x)$. Since player j with $x_j = B$ takes the minimaxing action D_j, the average instantaneous utility during the lth review round is $v_i^* = u_i(D, D) \leq \underline{v}_i$. The reward is 0 except for $\bar{\pi}_i(x, \lambda_j(l), l)$. Therefore, from (41), player i's value from the lth review round is close to $\underline{v}_i$ if we properly determine $\bar{\pi}_i(x, \lambda_j(l), l) \geq 0$.

Second, we verify 1-(a): In the lth review round, it is almost optimal for player i to follow $\sigma_i(x_i)$. We concentrate on the case where $\theta_j(\tilde{l}) = \theta_j(\lambda_j(\tilde{l}+1)) = G$ for all $\tilde{l} \leq l-1$ and player j uses the reward (32). Lemma 8 guarantees that, for almost optimality, player i can assume $\lambda_j(l) = \hat{\lambda}_j(l)$ and Section 11.4.4 guarantees that player j takes $a_j(x)$.

For the last Lth review round, player i maximizes

$$\sum_{t \in T(l)} u_i(a_t) + \pi_i^{\text{main}}(x, h_j^{\text{main}}, l) \tag{43}$$

with $l = L$. If $\lambda_j(L) = \hat{\lambda}_j(L) = G$, then $\pi_i^{\text{main}}(x, h_j^{\text{main}}, L)$ is linearly increasing with slope $\bar{L}$ in the score $X_j(L)$. Since (40) implies that the slope $\bar{L}$ is sufficiently large, it is almost optimal to take $a_i(x)$ in order to maximize (43). If $\lambda_j(L) = \hat{\lambda}_j(L) = B$, then $\pi_i^{\text{main}}(x, h_j^{\text{main}}, L)$ is flat and so player i wants to take a best response to player j's action. Therefore, D_i is almost optimal in order to maximize (43). Therefore, $\sigma_i(x_i)$ is almost optimal for the Lth review round.

We proceed backward. Suppose that player i follows $\sigma_i(x_i)$ from the $(l+1)$th review round and consider player i's incentive in the lth review round. Note that we define $\bar{\pi}_i$ such that player i's value is almost independent of $\lambda_j(l+1)$ as long as player i follows $\sigma_i(x_i)$ from the $(l+1)$th review round and $\lambda_j(l+1) = \hat{\lambda}_j(l+1)$.[50] In addition, Lemma 8 implies that player i in the main blocks does not put a high belief on the case that "$\lambda_j(l+1) \neq \hat{\lambda}_j(l+1)$ and player i's value depends on action profiles in the $(l+1)$th review round." Further,

[50]In the above discussion, we have verified that this claim is correct for the case with $\theta_j(\tilde{l}) = \theta_j(\lambda_j(\tilde{l}+1)) = G$ for all $\tilde{l} \leq l$ (with L replaced with $l+1$).

For the other cases, player i is indifferent between any action profile sequence, which implies that player i's value is constant for any action profile, as desired.

Section 11.4.4 guarantees that the probability that $\theta_j(l) = B$ or $\theta_j(\lambda_j(l+1)) = B$ given $\theta_j(\tilde{l}) = \theta_j(\lambda_j(\tilde{l}+1)) = G$ for all $\tilde{l} \leq l-1$ is almost independent of player i's strategy in the lth review round and supplemental round for $\lambda_i(l+1)$. Therefore, for almost optimality, we can assume that player i in the lth review round maximizes (43), assuming that $\theta_j(\tilde{l}) = \theta_j(\lambda_j(\tilde{l}+1)) = G$ for all $\tilde{l} \leq l-1$, that player j uses the reward (32), that $\lambda_j(l) = \hat{\lambda}_j(l)$, and that player j takes $a_j(x)$. Therefore, the same argument as for the Lth review round establishes that $\sigma_i(x_i)$ is almost optimal for the lth review round.

Third, 1-(b) is true since, as seen in Section 11.4.4, whenever player i's message affects player j's continuation action, player i has been almost indifferent between any action profile. In addition, although player i's message and signal observation affect player i's posterior about the optimality of $\hat{\lambda}_j(l+1)$, the effect is sufficiently small for almost optimality (see the discussion in Section 13.1).

Fourth, 2 is true since except for rare events $\hat{\lambda}_j(l) = B$ or $\hat{\lambda}_i(l) = B$, the players take $a(x)$ in the lth review round. In addition, $\theta_j(l) = B$ or $\theta_j(\lambda_j(l+1)) = B$ does not happen with high probability. Hence, in total, the players play $a(x)$ and player j uses the reward (32) for each review round with high probability. Therefore, the ex ante value is $\bar{v}_i$ ($\underline{v}_i$, respectively) if $x_j = G$ ($x_j = B$, respectively) by construction of $\bar{\pi}_i$.

Finally, from Section 11.3, π_i^{main} satisfies 3 since we take $\bar{\pi}_i(x, \lambda_j(l), l) \leq 0$ ($\bar{\pi}_i(x, \lambda_j(l), l) \geq 0$, respectively) if $x_j = G$ (B, respectively).

Therefore, we are left to construct the strategy in the report block and π_i^{report} such that $\sigma_i(x_i)$ and $\pi_i^{\text{main}} + \pi_i^{\text{report}}$ satisfy (3), (4) and (5).

14 Exact Optimality

In this section, we explain the strategy and the reward π_i^{report} in the report block. As briefly mentioned in Sections 6 and 7, player i reports h_i^{main} to player j if player i is picked by the public randomization. Player j calculates π_i^{report} based on the reported history $\hat{h}_i^{\text{main}}$ so that $\sigma_i(x_i)$ is exactly optimal against $\sigma_j(x_j)$ and $\pi_i^{\text{main}} + \pi_i^{\text{report}}$.

With the perfect cheap talk, the players could report h_i^{main} simultaneously and instantaneously. However, as seen in Section 4, for the dispensability of the cheap talk, it is important to construct the report block so that only one player sends the message and that the cardinality of the messages is sufficiently small.

For the first purpose, the players use public randomization. Player 1 reports h_1^{main} if $y^p \leq \frac{1}{2}$ and player 2 reports h_2^{main} if $y^p > \frac{1}{2}$. Below, we consider the case where player i reports the history.

From Section 10, there is a chronological order for the rounds. Hence, we can number all the rounds serially. For example, the round for x_1 is round 1, the round for x_2 is round 2, the first review round is round 3, the supplemental round for $\lambda_1(l+1)$ is round 4, the supplemental round for $\lambda_2(l+1)$ is round 5, and so on.

Let h_i^{r+1} be player i's history at the beginning of the $(r+1)$th round.

The reward from the report block is the summation of the rewards for each round:

$$\pi_i^{\text{report}}(x_j, h_j^{T_P+1} : \delta) = \sum_r \pi_i^{\text{report}} \left(h_j^{r+1}, \hat{h}_i^{r+1}, r \right).$$

Here, $\hat{h}_i^{r+1}$ is player i's report about h_i^{r+1}. Precisely, to reduce the cardinality of the messages, player i reports the summary of h_i^{r+1}. The details will be determined below. Note that the reward for round r, $\pi_i^{\text{report}} \left(h_j^{r+1}, \hat{h}_i^{r+1}, r \right)$, depends on the history until the end of round r.

We define $\pi_i^{\text{report}} \left(h_j^{r+1}, \hat{h}_i^{r+1}, r \right)$ such that

1. During the main blocks, for each period t and each h_i^t, player i believes that player i will tell the truth about h_i^{r+1}.

2. Based on the truthful report h_i^{r+1}, π_i^{report} will be adjusted so that $\sigma_i(x_i)$ is exactly optimal.

Suppose that we have shown the truthtelling incentive 1 and let us concentrate on the adjustment 2. See Section 15.7 for the proof of the truthtelling incentive.

Since we need to keep the cardinality of the messages sufficiently small, we consider the summary statistics $\#_i^{\tilde{r}}$ for the history in each round $\tilde{r}$: For round 1, let $\#_i^1$ be x_1, the message sent by the perfect cheap talk in round 1. Similarly, for round 2, let $\#_i^2$ be x_2.

For round $\tilde{r}$ corresponding to a review round, for each $(a_i, y_i) \in A_i \times Y_i$, let $\#_i^{\tilde{r}}(a_i, y_i)$ be how many times player i observed an action-signal pair (a_i, y_i) in round $\tilde{r}$. Let $\#_i^{\tilde{r}}$ be a vector $\left(\#_i^{\tilde{r}}(a_i, y_i)\right)_{a_i, y_i}$.

For round $\tilde{r}$ where player i sends a message m via noisy cheap talk, let $\#_i^{\tilde{r}}$ be player i's message and signals $(m, g[i](m), g_2[i](m))$.[51]

For round $\tilde{r}$ where player i receives a message m via noisy cheap talk, let $\#_i^{\tilde{r}}$ be player i's signals $(f[i](m), f_2[i](m))$.

By backward induction, we construct $\pi_i^{\text{report}}\left(h_j^{r+1}, \hat{h}_i^{r+1}, r\right)$. For round r corresponding to a review round, let $(T(r, a_i))_{a_i \in A_i} \in T^{|A_i|}$ be the set of strategies that take a_i for $T(r, a_i)$ times in round r. Given the summary of the history at the beginning of round r, player j calculates the ex ante continuation value of taking $(T(r, a_i))_{a_i}$ at the beginning of round r. To do so, player j assumes that player i follows the equilibrium path from the next round if $(T(r, a_i))_{a_i}$ is a strategy taken on the equilibrium path. If $(T(r, a_i))_{a_i}$ is a deviation, then player j assumes that player i takes a best response from the next round. In addition, player j takes into account the reward in the report block for the subsequent rounds $\sum_{\tilde{r} \geq r+1} \pi_i^{\text{report}}\left(h_j^{\tilde{r}+1}, \hat{h}_i^{\tilde{r}+1}, \tilde{r}\right)$. This calculation is well defined by the following three reasons: (i) Ex ante, by (23), we can ignore discounting. (ii) The set of $(T(r, a_i))_{a_i}$'s that player i should take with positive probability on the equilibrium path is determined by a summary of player i's history $\mathfrak{h}_i^r \equiv \left\{\#_i^{\tilde{r}}\right\}_{\tilde{r} \leq r-1}$. (iii) Player j determines her continuation strategy treating each period within a round identically. Hence, given the truthtelling strategy in the report block, even after a deviation, $\mathfrak{h}_i^r$ and $\#_i^r$ are enough to calculate the best responses and the continuation value for player i at the beginning of the next round. To calculate the ex ante value of taking $(T(r, a_i))_{a_i}$, we take the expectation of this continuation value using the conditional distribution of $\#_i^r$ given $(T(r, a_i))_{a_i}$. Since player j's strategy treats each

[51] Although we neglected the secondary signals $f_2[i](m)$ and $g_2[i](m)$ for the almost optimality, for the exact optimality, we need to take into account the belief updates from the secondary signals.

period in round r identically, the timing of taking a_i does not change the expectation as long as $(T(r, a_i))_{a_i}$ is fixed.

On the other hand, from $\#_i^r$, player j can know how many times player i took a_i in round r. Let $(\#_i^r(a_i))_{a_i}$ be this number. With abuse of notation, we say player i reports that she took $(T(r, a_i))_{a_i}$ if $T(r, a_i)$ is equal to $\#_i^r(a_i)$ for all a_i.

Therefore, based on the report of $\mathfrak{h}_i^r$ and $(T(r, a_i))_{a_i}$, player j can construct $\pi_i^{\text{report}}\left(h_j^{r+1}, \hat{h}_i^{r+1}, r\right)$ such that, given $\mathfrak{h}_i^r$ and player j's action plan in round r, $\alpha_j(r)$,

- for all $(T(r, a_i))_{a_i}$ that should be taken with positive probability, player i's ex ante payoff at the beginning of round r is the same, and
- if player i reports $(T(r, a_i))_{a_i}$ that should not be taken on the equilibrium path, then player j punishes player i. We make sure that this punishment is sufficiently large to discourage any deviation (not only ex ante but also) after any history.

Since the original strategy is almost optimal conditional on $\alpha_j(r)$ and a strategy which takes non-constant actions in a review round is a deviation, the second bullet point incentivizes player i to take a constant action in each review round.[52] Hence, the ex ante optimality at the beginning of each review round established by the first bullet point is sufficient for the sequential optimality.

For round r corresponding to a round where player i sends a message m, we replace $(T(r, a_i))_{a_i}$ with the set of possible messages m's in the above discussion.

By backward induction, we can verify that $\sigma_i(x_i)$ is optimal, taking all the continuation strategies into account after a deviation. See Section 15.7 for how to incentivize the players to tell the truth.

[52] See the Supplemental Material 2 for how we allow a mixed strategy minimax in a general game.

15 Appendix

15.1 Proof of Lemma 1

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

$$
\begin{aligned}
p(G, h_{i-1}^{T_P+1} &: \delta) \equiv 1 + \frac{1-\delta}{\delta^{T_P}} \frac{\pi_i(G, h_{i-1}^{T_P+1} : \delta)}{\bar{v}_i - \underline{v}_i}, \\
p(B, h_{i-1}^{T_P+1} &: \delta) \equiv \frac{1-\delta}{\delta^{T_P}} \frac{\pi_i(B, h_{i-1}^{T_P+1} : \delta)}{\bar{v}_i - \underline{v}_i}.
\end{aligned} \tag{44}
$$

If (5) is satisfied, then for sufficiently large δ, $p(G, h_{i-1}^{T_P+1} : \delta)$, $p(B, h_{i-1}^{T_P+1} : \delta) \in [0,1]$ for all $h_{i-1}^{T_P+1}$. We see the repeated game as the repetition of T_P-period "review phases." In each phase, player i has a state $x_i \in \{G, B\}$. Within the phase, player i with state x_i plays according to $\sigma_i(x_i)$ in the current phase. After observing $h_i^{T_P+1}$ in the current phase, the state in the next phase is equal to G with probability $p(x_i, h_i^{T_P+1} : \delta)$ and B with the remaining probability.

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

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

Then, since

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

and

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

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

1. Conditional on the opponents' state, the above strategy in the infinitely repeated game is optimal.

2. Regardless of $x_{-(i-1)}$, if player $i-1$ is in the state G, then player i's payoff from the infinitely repeated game is $\bar{v}_i$ and if player $i-1$ is in the state B, then player i's payoff is $\underline{v}_i$.

3. The payoff in the initial period is $p_v^{i-1}\bar{v}_i+(1-p_v^{i-1})\underline{v}_i=v_i$ as desired.

15.2 Proof of Lemma 3

If Assumption 4 is satisfied, then for any $q_2>q_1$, the system of equations (18) and (19) has a solution $\{\psi_i^a(y_j)\}_{y_j}$. If $\{\psi_i^a(y_j)\}_{y_j}$ solves the system for q_2 and q_1, then, for any $m,m'\in\mathbb{R}_{++}$, $\left\{\frac{\psi_i^a(y_j)+m}{m'}\right\}_{y_j}$ solves the system for $\frac{q_2+m}{m'}$ and $\frac{q_1+m}{m'}$. Therefore, we can make sure that $\psi_j^a:Y_j\rightarrow(0,1)$.

15.3 Proof of Lemma 5

This follows from the assumption that $Q_1(\tilde{a}_i,a_j)$ is linearly independent with respect to $\tilde{a}_i$.[53] Since $\left(1-\delta^{t-1}\right)u_i\left(a_t\right)$ converges to 0 as δ goes to unity for all $t\in\{1,...,T_P\}$ with $T_P=O((1-\delta)^{-\frac{1}{2}})$, we have

$$
\lim_{\delta\rightarrow1}\sup_{t\in\{1,...,T_P\},a_{j,t},y_{j,t}}\left|\pi_i^\delta\left(t,a_{j,t},y_{j,t}\right)\right|=0,
$$

[53] See Yamamoto (2007) for the formal proof.

which implies (22).

15.4 Proof of Lemma 5

This follows from the assumption that $Q_1(\tilde{a}_i, a_j)$ is linearly independent with respect to $\tilde{a}_i$.[54]

15.5 Proof of Lemma 8

Once $\lambda_j(\tilde{l}) = B$ is induced, then $\lambda_j(\tilde{l}') = B$ for all the following rounds. Hence, there exists a unique l^* such that $\lambda_j(\tilde{l}) = B$ is initially induced in the (l^*+1)th review round: $\lambda_j(1) = \cdots = \lambda_j(l^*) = G$ and $\lambda_j(l^*+1) = \cdots = \lambda_j(L) = B$. Similarly, there exists $\hat{l}^*$ with $\hat{\lambda}_j(1) = \cdots = \hat{\lambda}_j(\hat{l}^*) = G$ and $\hat{\lambda}_j(\hat{l}^*+1) = \cdots = \hat{\lambda}_j(L) = B$. If $\lambda_j(L) = G$ ($\hat{\lambda}_j(L) = G$, respectively), then define $l^* = L$ ($\hat{l}^* = L$, respectively).

Then, there are following three cases:

- $l^* = \hat{l}^*$: This means $\lambda_j(l) = \hat{\lambda}_j(l)$ for all l as desired.
- $l^* > \hat{l}^*$: This means that 1-(b) or 2 is the case when player i creates $\hat{\lambda}_j(\hat{l}^*+1)$ in Section 11.4.2 and that $\hat{\lambda}_j(\hat{l}^*+1)$ is determined by (39).

  By Lemma 2, player i believes that, conditional on $\lambda_j(\hat{l}^*+1)$, $f[i](\lambda_j(\hat{l}^*+1)) = \lambda_j(\hat{l}^*+1)$ or $g[j](\lambda_j(\hat{l}^*+1)) = E$ with probability no less than $1 - \exp(-O(T^{\frac{1}{2}}))$. If the latter is the case, then $\theta_j(\lambda_j(\hat{l}^*+1)) = B$. Therefore, player i believes that, conditional on $\lambda_j(\hat{l}^*+1)$, $f[i](\lambda_j(\hat{l}^*+1)) = \lambda_j(\hat{l}^*+1)$ or $\theta_j(\lambda_j(\hat{l}^*+1)) = B$ with probability no less than $1 - \exp(-O(T^{\frac{1}{2}}))$.

  Since player j's continuation play in the main blocks does not depend on $g[j](\lambda_j(\hat{l}^*+1))$, we are done.
- $l^* < \hat{l}^*$: There are following two cases:
  - If 1-(b) or 2 is the case when player i creates $\hat{\lambda}_j(l^*+1)$ in Section 11.4.2, then by the same reason as above, we are done.

[54] Again, see Yamamoto (2007).

– If 1-(a) is the case when player i creates $\hat{\lambda}_j(l^*+1)$ in Section 11.4.2, then player i at the end of the l^*th review round believes that $\lambda_j(l^*+1) = G$ with probability no less than $1 - \exp(-O(T))$.

Player j's continuation strategy reveals $\lambda_j(l^*+1)$ through (i) the strategy in the supplemental round for $\lambda_j(l^*+1)$ and (ii) $\hat{\lambda}_i(l^*+1)$. By Condition 4 of Lemma 2, any signals happen in the supplemental round for $\lambda_j(l^*+1)$ with probability at least $\exp(-O(T^{\frac{1}{2}}))$. Hence, the update of the likelihood from (i) is bounded by $\exp(O(T^{\frac{1}{2}}))$.

In addition, for $\hat{\lambda}_i(l^*+1)$, player j is in 1-(b) or 2 when player j creates $\hat{\lambda}_i(l^*+1)$ in Section 11.4.2 with probability at least η. Since Condition 4 of Lemma 2 implies that any signals happen in the supplemental round for $\lambda_i(l^*+1)$ with probability at least $\exp(-O(T^{\frac{1}{2}}))$, the update of the likelihood from (ii) is bounded by $\frac{1-\eta}{\eta}\exp(O(T^{\frac{1}{2}}))$.

Therefore, after observing player j's continuation strategy, player i believes $\lambda_j(l+1) = B$ with probability at most

$$\frac{\exp(-O(T))\frac{1-\eta}{\eta}\left(\exp(O(T^{\frac{1}{2}}))\right)^2}{1-\exp(-O(T^{\frac{1}{2}}))} = \exp(-O(T)), \tag{45}$$

as desired.

15.6 Proof of Proposition 1

For 3, it suffices to have

$$\bar{\pi}_i\left(x, \lambda_j(l), l\right) \begin{cases} \leq 0 & \text{if } x_j = G, \\ \geq 0 & \text{if } x_j = B, \end{cases} \tag{46}$$

$$\left|\bar{\pi}_i\left(x, \lambda_j(l), l\right)\right| \leq \max\nolimits_{i,a} 2\left|u_i\left(a\right)\right| T \tag{47}$$

for all $x \in \{G, B\}^2$, $\lambda_j(l) \in \{G, B\}$ and $l \in \{1, ..., L\}$.

To see why (46) and (47) are sufficient, notice the following: (47) with $T = (1-\delta)^{-\frac{1}{2}}$ implies

$$\lim_{\delta \to 1} \frac{1-\delta}{\delta^{T_P}} \sup_{x, h_j^{\text{main}}} \left| \sum_{l=1}^{L} \pi_i^{\text{main}}(x, h_j^{\text{main}}, l) \right| = 0.$$

(46) implies that $\pi_i^{\text{main}}(x, h_j^{\text{main}}, l) \geq 0$ with $x_j = G$ or $\pi_i^{\text{main}}(x, h_j^{\text{main}}, l) \leq 0$ with $x_j = B$ happens only if $\lambda_j(l) = G$ and $X_j(l) \notin [q_2 T - 2\varepsilon T, q_2 T + 2\varepsilon T]$. Since we have $\lambda_j(\tilde{l}) = B$ for $\tilde{l} > l$ after those events from Section 11.4.1, $\pi_i^{\text{main}}(x, h_j^{\text{main}}, l) \leq 0$ with $x_j = G$ or $\pi_i^{\text{main}}(x, h_j^{\text{main}}, l) \geq 0$ with $x_j = B$ except for one review round. In addition, (46) implies that $\pi_i^{\text{main}}(x, h_j^{\text{main}}, l) \leq \bar{L}T$ for $x_j = G$ and $\pi_i^{\text{main}}(x, h_j^{\text{main}}, l) \geq -\bar{L}T$ for $x_j = B$. Hence, in total, for any h_j^{main}, $-\bar{L}T + \sum_{l=1}^{L} \pi_i^{\text{main}}(x, h_j^{\text{main}}, l) \leq 0$ with $x_j = G$ and $\bar{L}T + \sum_{l=1}^{L} \pi_i^{\text{main}}(x, h_j^{\text{main}}, l) \geq 0$ with $x_j = B$. Therefore, (5) is satisfied.[55]

Now, we are left to prove 1 and 2. 1-(b) is true by the reasons that we have explained in the main text.

We will verify 1-(a) by backward induction. Section 11.4.4 guarantees that the probability that $\theta_j(l) = B$ or $\theta_j(\lambda_j(l+1)) = B$ given $\theta_j(\tilde{l}) = \theta_j(\lambda_j(\tilde{l}+1)) = G$ for all $\tilde{l} \leq l-1$ is almost independent of player i's strategy in the lth review round and supplemental round for $\lambda_i(l+1)$ and so we can neglect the effect of player i's strategy on θ_j for almost optimality. Further, for a moment, forget about the first term in π_i^{main}, $-\bar{L}T$ ($\bar{L}T$, respectively) for $x_j = G$ ($x_j = B$, respectively).

In the Lth review round, for almost optimality, we can assume that $\lambda_j(L) = \hat{\lambda}_j(L)$ and that player j uses (32) by the following reason: By Lemma 8, player i has a posterior no less than $1 - \exp(-O(T^{\frac{1}{2}}))$ on the events that $\lambda_j(L) = \hat{\lambda}_j(L)$ or any action is optimal. Since the per-period difference of the payoff from two different strategies is bounded by $\bar{U} \equiv \bar{L} + \max_{i,a} 2|u_i(a)|$, the expected loss from assuming $\lambda_j(L) = \hat{\lambda}_j(L)$ is no more than $\exp(-O(T^{\frac{1}{2}}))\bar{U}T$. Therefore, for almost optimality, we can assume that $\lambda_j(L) = \hat{\lambda}_j(L)$. Further, if (31) is used, then any action is optimal. Therefore, we can assume that player j uses (32).

[55] We neglect (23) because of (22).

In addition, if player j does not play $a_j(x)$, then it means that $\hat{\lambda}_i(L) = B$ and that, by Section 11.4.4, any action is optimal for player i (that is, (31) is used). Hence, we can concentrate on the case where player j plays $a_j(x)$.

If $\lambda_j(L) = \hat{\lambda}_j(L) = G$, then $a_i(x)$ is strictly optimal for sufficiently large T since (40) implies that the marginal expected increase in $\bar{L}X_j(L)$ is sufficiently large.[56] If $\lambda_j(L) = \hat{\lambda}_j(L) = B$, then D_i is strictly optimal since the reward (32) is constant. Therefore, $\sigma_i(x_i)$ is optimal.

Further, if player j uses (32) and $\lambda_j(L) = \hat{\lambda}_j(L)$, then player i's average continuation payoff at the beginning of the Lth review round except for $\bar{\pi}_i(x, \lambda_j(L), L)$ is

$$\begin{array}{cl} w_i(x) - 2\varepsilon\bar{L} & \text{if } x_j = G, \lambda_j(L) = \hat{\lambda}_j(L) = G, \\ w_i(x) + 2\varepsilon\bar{L} & \text{if } x_j = B, \lambda_j(L) = \hat{\lambda}_j(L) = G, \\ u_i(D_i, C_j) & \text{if } x_j = G, \lambda_j(L) = \hat{\lambda}_j(L) = B, \\ u_i(D_i, D_j) & \text{if } x_j = B, \lambda_j(L) = \hat{\lambda}_j(L) = B. \end{array} \tag{48}$$

Hence, there exists $\bar{\pi}_i(x, \lambda_j(L), L)$ with (46) and (47) such that player i's average continuation payoff is equal to $w_i(x) - 2\varepsilon\bar{L}$ if $x_j = G$ and $w_i(x) + 2\varepsilon\bar{L}$ if $x_j = B$.

Therefore, we define $\bar{\pi}_i(x, \lambda_j(L), L)$ such that player i's value from the Lth review round is independent of $\lambda_j(L)$ as long as $\lambda_j(L) = \hat{\lambda}_j(L)$.[57] In addition, Lemma 8 implies that player i in the main blocks does not put a belief more than $\exp(-O(T^{\frac{1}{2}}))$ on the events that $\lambda_j(L) \neq \hat{\lambda}_j(L)$ and player i's value depends on action profiles in the Lth review round. Further, again, we can neglect the effect of player i's strategy on θ_j for almost optimality. Therefore, for almost optimality, we can assume that player i in the $(L-1)$th review round maximizes

$$\sum_{t \in T(L-1)} u_i(a_t) + \pi_i^{\text{main}}(x, h_j^{\text{main}}, L-1), \tag{49}$$

[56] For large T, the effect of dropping one $t_j(l)$ is negligible.

[57] In the above discussion, we have verified that this claim is correct for the case with $\theta_j(\tilde{l}) = \theta_j(\lambda_j(\tilde{l}+1)) = G$ for all $\tilde{l} \leq L-1$.

For the other cases, player i is indifferent between any action profile sequence, which implies that player i's value is constant for any action profile, as desired.

assuming $\lambda_j(L-1) = \hat{\lambda}_j(L-1)$.

Therefore, the same argument as for the Lth review round establishes that $\sigma_i(x_i)$ is almost optimal in the $(L-1)$th review round.

Further, if player j uses (32) and $\lambda_j(L-1) = \hat{\lambda}_j(L-1)$, then player i's average payoff from the $(L-1)$th review round except for $\bar{\pi}_i(x, \lambda_j(L-1), L-1)$ is given by (48). The cases where (31) will be used in the Lth review round will happen with probability no more than η (player j is in 1-(b) in Section 11.4.2) plus some negligible probabilities for not having (35) or (36). When (31) is used, per period payoff is bounded by $[-\bar{u}, \bar{u}]$ by Lemma 5.

Therefore, there exists $\bar{\pi}_i(x, \lambda_j(L-1), L-1)$ with (46) and (47) such that if player j uses (32) and $\lambda_j(L-1) = \hat{\lambda}_j(L-1)$, then player i's average continuation payoff from the $(L-1)$th and Lth review rounds is $w_i(x) - 2\varepsilon\bar{L} - \eta\left(\bar{u} + \max_{i,x} w_i(x)\right)$ if $x_j = G$ and $w_i(x) + 2\varepsilon\bar{L} + \eta\left(\bar{u} - \min_{i,x} w_i(x)\right)$ if $x_j = B$.

Recursively, for $l = 1$, 1-(a) of the proposition is satisfied and the average ex ante payoff of player i is $w_i(x) - 2\varepsilon\bar{L} - L\eta\left(\bar{u} - \max_{i,x} w_i(x)\right)$ if $x_j = G$ and $w_i(x) + 2\varepsilon\bar{L} + L\eta\left(\bar{u} - \min_{i,x} w_i(x)\right)$ if $x_j = B$. Note that, in the first review round, player j uses (32) and $\hat{\lambda}_j(1) = \lambda_j(1) = G$ with probability one.

Taking the first term $-\bar{L}T$ ($\bar{L}T$, respectively) for $x_j = G$ ($x_j = B$, respectively) into account, the average ex ante payoff is $w_i(x) - \frac{\bar{L}}{L} - 2\varepsilon\bar{L} - L\eta\left(\bar{u} - \max_{i,x} w_i(x)\right)$ if $x_j = G$ and $w_i(x) + \frac{\bar{L}}{L} + 2\varepsilon\bar{L} + L\eta\left(\bar{u} - \min_{i,x} w_i(x)\right)$ if $x_j = B$.

From (41), we can further modify $\bar{\pi}_i(x, G, 1)$ with (46) and (47) such that $\sigma_i(x_i)$ gives $\bar{v}_i$ ($\underline{v}_i$, respectively) if $x_j = G$ (B, respectively). Therefore, 2 of the proposition is satisfied.

15.7 Formal Construction of the Report Block

We formally construct the strategy in the report block and π_i^{report} so that player i tells the truth about h_i^{main} and that $\sigma_i(x_i)$ is exactly optimal. Here, we do not consider the feasibility constraint (5). As we will see, π_i^{report} is bounded by $[-T^{-1}, T^{-1}]$ and we can restore (5) by adding or subtracting a small constant depending on x_j without affecting efficiency or incentive.

Let $\mathcal{A}_j(r)$ be the set of information up to and including round r consisting of

- What state x_j player j is in, and
- What action plan $\alpha_j(l)$ player j took in the lth review round if round r is the lth review round or after.

We want to show that $\sigma_i(x_i)$ is exactly optimal in round r conditional on $\mathcal{A}_j(r)$. Note that $\mathcal{A}_j(r)$ contains x_j and so the equilibrium is belief-free at the beginning of the finitely repeated game.

The following notations are useful: As we have defined in Section 14, $\#_i^r$ is the summary of player i's history within round r, $\mathfrak{h}_i^r$ is the summary of player i's history at the beginning of round r, and $(T(r,a_i))_{a_i}$ is the set of player i's strategies that take a_i for $T(r,a_i)$ times in round r ex ante (if round r corresponds to a review round). On the other hand, let $\hat{\#}_i^r$ be player i's report of $\#_i^r$ and $\left(\hat{T}(r,a_i)\right)_{a_i}$ be such that, according to player i's report $\hat{\#}_i^r$, player i takes each a_i for $\hat{T}(r,a_i)$ times in round r. In addition, let t_r be the first period of round r.

For round r corresponding to a review round, we divide a review round into $T^{\frac{3}{4}}$ review subrounds. Each kth subround is from period $t_r + (k-1)\,T^{\frac{1}{4}} + 1$ to period $t_r + kT^{\frac{1}{4}}$ with $k \in \{1, \ldots, T^{\frac{3}{4}}\}$. Let $T(r,k)$ be the set of periods in the kth subround of round r. Let $\#_i^r(k)(a_i, y_i) \in \{1, ..., T^{\frac{1}{4}}\}$ be how many times player i observed an action-signal pair (a_i, y_i) in the kth subround of round r and $\#_i^r(k)$ be $(\#_i^r(k)(a_i,y_i))_{a_i,y_i}$.

When player i is picked by the public randomization device with probability $\frac{1}{2}$, player i sends the messages via perfect cheap talk: Sequentially from round 1 to the last round, player i reports the history as follows:

- If round r corresponds to a review round, then
  - First, player i reports the summary $\#_i^r$.
  - Second, for each subround k, player i reports the summary $\#_i^r(k)$.

– Third, public randomization is drawn such that each subround k is randomly picked with probability $T^{-\frac{3}{4}}$. Let $k(r)$ be the subround picked by the public randomization.

– Fourth, for $k(r)$, player i reports the whole history $\{a_{i,t}, y_{i,t}\}_{t\in T(r,k(r))}$ in the $k(r)$th subround.

- If player i sends the noisy cheap talk message in round r, then player i reports $\#_i^r$, which is her true message m and signals $g[i](m)$ and $g_2[i](m)$.

- If player i receives the noisy cheap talk message in round r, then player i reports $\#_i^r$, which is her signals $f[i](m)$ and $f_2[i](m)$.

Remember that we want to use *binary* perfect cheap talk as mentioned in Section 4. For round r corresponding to a review round, for each $\#_i^r \in \{1, ..., T\}^{|A_i||Y_i|}$, we can attach a sequence of binary messages $\{G, B\}$ so that the sequence uniquely identifies $\#_i^r$. The length of the sequence is $|A_i|\,|Y_i| \log_2 T$. Similarly, for each $\#_i^r(k)$, we can attach a sequence of binary messages $\{G, B\}$ with length $\frac{1}{4}\,|A_i|\,|Y_i| \log_2 T$. For each (a_i, y_i), we can attach a sequence of binary messages $\{G, B\}$ with length $\log_2 |A_i|\,|Y_i|$. Then, the number of binary messages to send all the messages is $O(T^{\frac{1}{4}})$ since the fourth message $\{a_{i,t}, y_{i,t}\}_{t\in T(r,k(r))}$ is the longest. For the other rounds, the length of the necessary messages is at most 3. Therefore, in total, the number of messages we need is

$$O(T^{\frac{1}{4}}). \tag{50}$$

As a preparation to show the incentive to tell the truth, we prove the following lemma:

Lemma 9 *If Assumption 5 is satisfied, then there exists $\bar{\varepsilon} > 0$ such that*

1. *For each $l \in \{1, ..., L\}$, in the lth review round, there exists $g_i(h_j^{\text{main}}, a_i, y_i)$ such that,*

for period $t \in T(l)$, it is better for player i to report $(a_{i,t}, y_{i,t})$ truthfully: For all h_i^{main},

$$\begin{aligned} & \mathbb{E}\left[g_i(h_j^{\text{main}}, \hat{a}_{i,t}, \hat{y}_{i,t}) \mid h_i^{\text{main}}, (\hat{a}_{i,t}, \hat{y}_{i,t}) = (a_{i,t}, y_{i,t})\right] \\ > \; & \mathbb{E}\left[g_i(h_j^{\text{main}}, \hat{a}_{i,t}, \hat{y}_{i,t}) \mid h_i^{\text{main}}, (\hat{a}_{i,t}, \hat{y}_{i,t}) \neq (a_{i,t}, y_{i,t})\right] + \bar{\varepsilon} T^{-1}, \end{aligned} \tag{51}$$

where $(\hat{a}_{i,t}, \hat{y}_{i,t})$ is player i's report about $(a_{i,t}, y_{i,t})$ in the report block.

2. *For the round where player i sends the noisy cheap talk message, it is better to report $(m, g[i](m), g_2[i](m))$ truthfully:*

$$\begin{aligned} & \mathbb{E}\left[g_i(h_j^{\text{main}}, \hat{m}, \widehat{g[i](m)}, \widehat{g_2[i](m)}) \mid h_i^{\text{main}}, \left(\hat{m}, \widehat{g[i](m)}, \widehat{g_2[i](m)}\right) = (m, g[i](m), g_2[i](m))\right] \\ > \; & \mathbb{E}\left[g_i(h_j^{\text{main}}, \hat{m}, \widehat{g[i](m)}, \widehat{g_2[i](m)}) \mid h_i^{\text{main}}, \left(\hat{m}, \widehat{g[i](m)}, \widehat{g_2[i](m)}\right) \neq (m, g[i](m), g_2[i](m))\right] \\ & + \bar{\varepsilon} T^{-1}, \end{aligned} \tag{52}$$

where $\left(\hat{m}, \widehat{g[i](m)}, \widehat{g_2[i](m)}\right)$ is player i's report in the report block about the message and signals of the supplemental round.

3. *For the round where player i receives the noisy cheap talk message, it is better to report $(f[i](m), f_2[i](m))$ truthfully:*

$$\begin{aligned} & \mathbb{E}\left[g_i(h_j^{\text{main}}, \widehat{f[i](m)}, \widehat{f_2[i](m)}) \mid h_i^{\text{main}}, \left(\widehat{f[i](m)}, \widehat{f_2[i](m)}\right) = (f[i](m), f_2[i](m))\right] \\ > \; & \mathbb{E}\left[g_i(h_j^{\text{main}}, \widehat{f[i](m)}, \widehat{f_2[i](m)}) \mid h_i^{\text{main}}, \left(\widehat{f[i](m)}, \widehat{f_2[i](m)}\right) \neq (f[i](m), f_2[i](m))\right] \\ & + \bar{\varepsilon} T^{-1}, \end{aligned} \tag{53}$$

where $\left(\widehat{f[i](m)}, \widehat{f_2[i](m)}\right)$ is player i's report in the report block about the signals of the supplemental round.

Proof.

1. We show that

$$g_i(h_j^{\text{main}}, \hat{a}_{i,t}, \hat{y}_{i,t}) = -\mathbf{1}\{t_j(l) = t\} \left\| \mathbf{1}_{y_{j,t}} - \mathbb{E}[\mathbf{1}_{y_{j,t}} \mid \hat{a}_{i,t}, \hat{y}_{i,t}, a_{j,t}] \right\|^2$$

works.[58] To see this, consider the following two cases:

(a) If $t_j(l) \neq t$, then any report is optimal since $g_i(h_j^{\text{main}}, \hat{a}_{i,t}, \hat{y}_{i,t}) = 0$.

(b) If $t_j(l) = t$, then period t is not used for the construction of player j's continuation strategy. Hence, player i cannot learn $y_{j,t}$ from h_i^{main}. Hence, player i, after knowing $t_j(l) = t$ and $a_{j,t}$, wants to minimize

$$\min_{\hat{a}_{i,t},\hat{y}_{i,t}} \mathbb{E}\left[\left\|\mathbf{1}_{y_{j,t}} - \mathbb{E}[\mathbf{1}_{y_{j,t}} \mid \hat{a}_{i,t}, \hat{y}_{i,t}, a_{j,t}]\right\|^2 \mid a_{i,t}, y_{i,t}, a_{j,t}\right].$$

Assumption 5 implies that $(\hat{a}_{i,t}, \hat{y}_{i,t}) = (a_{i,t}, y_{i,t})$ is a unique minimizer.

We are left to show that there exists $\bar{\varepsilon} > 0$ such that, for any h_i^{main}, l and $t \in T(l)$, player i puts a belief at least $\bar{\varepsilon}T^{-1}$ on $t_j(l) = t$. Intuitively, this is true since the full support assumption of the signal distribution prevents player i from learning much about $t_j(l)$.

Formally, suppose that player i knows player j's past history before the lth review round and that player j's action plan $\alpha_j(l)$.

In addition, suppose that player i knows how many times player j observed each pair (a_j, y_j, φ_j) in the lth review round. Let $\#_j^*(l)(a_j, y_j, \varphi_j)$ denote this number and $\#_j^*(l)$ be a vector $\left(\#_j^*(l)(a_j, y_j, \varphi_j)\right)_{(a_j,y_j,\varphi_j)}$. Here, $\varphi_j \equiv (\Psi_j^{a(x)}, (E_j\Psi_i))$ is a statistics that player j constructs in the lth review round.

Further, suppose that player i knows what pair (a_j, y_j, φ_j) is observed in $t_j(l)$. That is, player i knows $\left(a_{j,t_j(l)}, y_{j,t_j(l)}, \varphi_{j,t_j(l)}\right) = (\bar{a}_j, \bar{y}_j, \bar{\varphi}_j)$.

Given player j's past history before the lth review round and player j's action plan $\alpha_j(l)$, how many times player j observed each pair (a_j, y_j, φ_j) in $T_j(l)$ determines player j's continuation strategy. Since this information can be calculated from $\#_j^*(l)$ and $\left(a_{j,t_j(l)}, y_{j,t_j(l)}, \varphi_{j,t_j(l)}\right) = (\bar{a}_j, \bar{y}_j, \bar{\varphi}_j)$, it suffices to show that, given player j's past

[58] Kandori and Matsushima (1994) use a similar reward to give a player the incentive to tell the truth about the history.

history before the lth review round and player j's action plan $\alpha_j(l)$, conditional on $\{a_{i,\tau}, y_{i,\tau}\}_{\tau \in T(l)}$, $\#_j^*(l)$ and $\left(a_{j,t_j(l)}, y_{j,t_j(l)}, \varphi_{j,t_j(l)}\right) = \left(\bar{a}_j, \bar{y}_j, \bar{\varphi}_j\right)$, player i puts a belief at least εT^{-1} on $t_j(l) = t$ for each t. For any t and $t' \in T(l)$, the likelihood ratio between $t_j(l) = t$ and $t_j(l) = t'$ is given by

$$\begin{aligned} & \frac{\Pr\left(t_j(l) = t \mid \{a_{i,\tau}, y_{i,\tau}\}_{\tau \in T(l)}, \alpha_j(l), \#_j^*(l), \left(a_{j,t_j(l)}, y_{j,t_j(l)}, \varphi_{j,t_j(l)}\right) = \left(\bar{a}_j, \bar{y}_j, \bar{\varphi}_j\right)\right)}{\Pr\left(t_j(l) = t' \mid \{a_{i,\tau}, y_{i,\tau}\}_{\tau \in T(l)}, \alpha_j(l), \#_j^*(l), \left(a_{j,t_j(l)}, y_{j,t_j(l)}, \varphi_{j,t_j(l)}\right) = \left(\bar{a}_j, \bar{y}_j, \bar{\varphi}_j\right)\right)} \\ = & \frac{\Pr\left(\#_j^*(l), \left(a_{j,t_j(l)}, y_{j,t_j(l)}, \varphi_{j,t_j(l)}\right) = \left(\bar{a}_j, \bar{y}_j, \bar{\varphi}_j\right) \mid \{a_{i,\tau}, y_{i,\tau}\}_{\tau \in T(l)}, \alpha_j(l), t_j(l) = t\right)}{\Pr\left(\#_j^*(l), \left(a_{j,t_j(l)}, y_{j,t_j(l)}, \varphi_{j,t_j(l)}\right) = \left(\bar{a}_j, \bar{y}_j, \bar{\varphi}_j\right) \mid \{a_{i,\tau}, y_{i,\tau}\}_{\tau \in T(l)}, \alpha_j(l), t_j(l) = t'\right)} \\ \in & \left[\min_{a_i, y_i, \bar{a}_j, \bar{y}_j, \bar{\varphi}_j} q(\bar{a}_j, \bar{y}_j, \bar{\varphi}_j \mid a_i, y_i, \alpha_j(l)), \frac{1}{\min_{a_i, y_i, \bar{a}_j, \bar{y}_j, \bar{\varphi}_j} q(\bar{a}_j, \bar{y}_j, \bar{\varphi}_j \mid a_i, y_i, \alpha_j(l))}\right]. \end{aligned}$$

Two remarks: First, we omit the conditioning on player j's history at the beginning of the lth review round for notational simplicity. Second, the last minimization with respect to $\bar{a}_j$ is taken subject to $\bar{a}_j \in \operatorname{supp}(\alpha_j(l))$.

Since player j's equilibrium action plan is *i.i.d.* in each round and we can assume the full support for y and φ_j from Assumption 3 and Lemma 3, there exists $\bar{\varepsilon} > 0$ such that

$$\begin{aligned} & \Pr\left(t_j(l) = t \mid \{a_{i,\tau}, y_{i,\tau}\}_{\tau \in T(l)}, \alpha_j(l), \#_j^*(l), \left(a_{j,t_j(l)}, y_{j,t_j(l)}, \varphi_{j,t_j(l)}\right) = \left(\bar{a}_j, \bar{y}_j, \bar{\varphi}_j\right)\right) \\ > & \ \bar{\varepsilon} \Pr\left(t_j(l) = t' \mid \{a_{i,\tau}, y_{i,\tau}\}_{\tau \in T(l)}, \alpha_j(l), \#_j^*(l), \left(a_{j,t_j(l)}, y_{j,t_j(l)}, \varphi_{j,t_j(l)}\right) = \left(\bar{a}_j, \bar{y}_j, \bar{\varphi}_j\right)\right) \end{aligned}$$

for all t and t'. Further, since $\alpha_j(l) \in A_j \cup \{\alpha_j^{\min\max}\}$, we can take $\bar{\varepsilon}$ independently from player j's history at the beginning of the lth review round and $\alpha_j(l)$. Since there exists at least one t with

$$\Pr\left(t_j(l) = t \mid \{a_{i,\tau}, y_{i,\tau}\}_{\tau \in T(l)}, \#_j^*(l), \left(a_{j,t_j(l)}, y_{j,t_j(l)}, \varphi_{j,t_j(l)}\right) = \left(\bar{a}_j, \bar{y}_j, \bar{\varphi}_j\right)\right) \geq T^{-1},$$

we are done.

2. From (13),

$$g_i(h_j^{\text{main}}, \hat{m}, \widehat{g[i](m)}, \widehat{g_2[i](m)}) = -\left\|\mathbf{1}_{f_2[j](m)} - \mathbb{E}[\mathbf{1}_{f_2[j](m)} \mid f[j](m), \hat{m}, \widehat{g[i](m)}, \widehat{g_2[i](m)}]\right\|^2$$

works. Since player j's continuation play is independent of $f_2[j](m)$, conditioning on h_i^{main} does not change the optimality.

3. From (12),

$$g_i(h_j^{\text{main}}, \widehat{f[i](m)}, \widehat{f_2[i](m)}) = -\left\|\mathbf{1}_{g_2[j](m)} - \mathbb{E}[\mathbf{1}_{g_2[j](m)} \mid m, g[j](m), \widehat{f[i](m)}, \widehat{f_2[i](m)}]\right\|^2$$

works. Since player j's continuation play is independent of $g_2[j](m)$, conditioning on h_i^{main} does not change the optimality.

■

Given this preparation, by backward induction, we construct $\pi_i^{\text{report}}\left(h_j^{r+1}, \hat{h}_i^{r+1}, r\right)$ for each r such that

$$\pi_i^{\text{report}}(x_j, h_j^{T_P+1} : \delta) = \sum_r \pi_i^{\text{report}}\left(h_j^{r+1}, \hat{h}_i^{r+1}, r\right)$$

makes it optimal to tell the truth in the report block and that $\sigma_i(x_i)$ is exactly optimal.

Formally, $\pi_i^{\text{report}}\left(h_j^{r+1}, \hat{h}_i^{r+1}, r\right)$ is the summation of the following rewards and punishments.

Punishment for a Lie One component of $\pi_i^{\text{report}}\left(h_j^{r+1}, \hat{h}_i^{r+1}, r\right)$ is the punishment for telling a lie. For round r corresponding to a review round, the punishment is the summation of the following three:

- The number indicating player i's lie about $\{a_{i,t}, y_{i,t}\}_{t\in T(r,k(r))}$:

$$\sum_{t\in T(r,k(r))} T^{-3} g_i(h_j^{\text{main}}, \hat{a}_{i,t}, \hat{y}_{i,t}). \tag{54}$$

- The number indicating player i's lie about $\#_i^r(k)$:

$$T^{-3} \times T^{\frac{3}{4}} \times \mathbf{1}\left\{\hat{\#}_i^r(k(r)) \neq \sum_{t \in T(r,k(r))} \mathbf{1}_{\hat{a}_{i,t},\hat{y}_{i,t}}\right\}, \tag{55}$$

  where $\mathbf{1}_{\hat{a}_{i,t},\hat{y}_{i,t}}$ is an $|A_i|\,|Y_i| \times 1$ vector such that the element corresponding to $(\hat{a}_{i,t}, \hat{y}_{i,t})$ is equal to 1 and the others are 0.

- The number indicating player i's lie about $\#_i^r$:

$$T^{-3} \times \mathbf{1}\left\{\hat{\#}_i^r \neq \sum_k \hat{\#}_i^r(k)\right\}. \tag{56}$$

If player i sends the noisy cheap talk message in round r, then player i reports $(m, g[i](m), g_2[i](m))$. Player j punishes player i if it is likely for player i to tell a lie by

$$T^{-3} g_i(h_j^{\text{main}}, \hat{m}, \widehat{g[i](m)}, \widehat{g_2[i](m)}). \tag{57}$$

If player i receives the noisy cheap talk message in round r, then player i reports $(f[i](m), f_2[i](m))$. Player j punishes player i if it is likely for player i to tell a lie by

$$T^{-3} g_i(h_j^{\text{main}}, \widehat{f[i](m)}, \widehat{f_2[i](m)}). \tag{58}$$

Cancel Out the Expected Punishment by Telling the Truth Note that even if player i tells the truth, punishment is positive for (54), (57) and (58) and the expectation of the punishment is different for different actions and messages.[59] We cancel out the differences in ex ante values of the punishment between different actions and messages: If player i is picked by the public randomization, then we add the following variables to π_i^{main}:

[59] On the equilibrium, (55) and (56) are 0 after any history.

- If round r is a review round, then

$$\sum_{t \in T(r)} \mathbf{1}\{t_j(l) = t\}\mathbf{1}\{t \in T\left(r, k(r)\right)\}T^{-3}\Pi_i(a_{j,t}, y_{j,t}). \tag{59}$$

- If player i sends the message in round r, then

$$T^{-3}\Pi_i(f\left[j\right](m)). \tag{60}$$

- If player i receives the message in round r, then

$$T^{-3}\Pi_i(m). \tag{61}$$

Here, $\Pi_i(f\left[j\right](m))$ ($\Pi_i(m)$, respectively) is defined so that the differences in (57) ((58), respectively) among messages are canceled out ex ante before sending (receiving, respectively) the message, as we define $\Pi_i(a_{j,t}, y_{j,t})$ in Lemma 7. Lemma 2 implies that the identifiability to create such a reward is satisfied. In addition, we abuse notation since we use the information in the report block such as $T\left(r, k(r)\right)$ to define π_i^{main}.

Then, in each period of the main block, given truthtelling in the report block, before taking an action or sending a message, the ex ante punishments from (54), (55), (56), (57), (58), (59), (60) and (61) are zero.

Reward for the Optimal Action Another component of $\pi_i^{\text{report}}\left(h_j^{r+1}, \hat{h}_i^{r+1}, r\right)$ is the reward for taking an equilibrium action in round r (or, punishment for not taking an equilibrium action). From $\left\{\#_i^{\tilde{r}}\right\}_{\tilde{r} \leq r-1}$, we can calculate $\mathfrak{h}_i^r = \left\{\#_i^{\tilde{r}}\right\}_{\tilde{r} \leq r-1}$. Let $\hat{\mathfrak{h}}_i^r$ be player j's inference of $\mathfrak{h}_i^r$ based on player i's reports $\left\{\hat{\#}_i^{\tilde{r}}\right\}_{\tilde{r} \leq r-1}$.

If round r corresponds to a review round, then based on the reports $\hat{\mathfrak{h}}_i^r$ and $\hat{\#}_i^r$, player j gives the reward

$$f_i(\hat{\mathfrak{h}}_i^r, \hat{\#}_i^r, \alpha_j(r)), \tag{62}$$

which is to be determined. Here, $\alpha_j(r)$ is player j's strategy in round r.

If round r corresponds to a round where player i sends the message m, then based on the reports $\hat{\mathfrak{h}}_i^r$ and $\hat{m}$, player j gives the reward

$$f_i(\hat{\mathfrak{h}}_i^r, \hat{m}), \tag{63}$$

which is to be determined.

We will take f_i such that

$$f_i(\hat{\mathfrak{h}}_i^r, \hat{\#}_i^r, \alpha_j(r)), f_i(\hat{\mathfrak{h}}_i^r, \hat{m}) \in [-T^{-r-5}, T^{-r-5}] \tag{64}$$

for all $\hat{\mathfrak{h}}_i^r$, $\hat{\#}_i^r$, $\alpha_j(r)$ and $\hat{m}$.

Incentive to Tell the Truth Before specifying f_i, we establish player i's incentive to tell the truth. For the reports about the last round, all the reports about the previous rounds are sunk. Hence, what the reports affect is the punishment and f_i for the last round. From (64), the effect on f_i is bounded by $O(T^{-5})$ while the marginal effect on punishment from telling a lie is at least $O(T^{-4})$ from (54), (55), (56), (57), (58) and Lemma 9. Hence, truthtelling is strictly optimal.

Given the incentive to tell the truth about the last round, the same argument holds for the second last round, and so on. By induction, we establish player i's incentive to tell the truth for all the rounds.

Ex Ante Expected Punishment Given the truthtelling incentive and (59), (60) and (61), in each period of each main block, before taking an action or sending a message, the ex ante punishments from (54), (55), (56), (57), (58), (59), (60) and (61) are zero.

Determination of f_i We determine f_i by backward induction. For round r corresponding to a review round, given $\hat{\mathfrak{h}}_i^r$ and $\alpha_j(r)$, we can calculate the ex ante continuation value of $(T(r, a_i))_{a_i}$. To do so, we assume that player i follows the equilibrium path from the next round if $\hat{\mathfrak{h}}_i^r$ is an on-path history and $(T(r,a_i))_{a_i}$ is a strategy taken on the equilibrium path. If $\hat{\mathfrak{h}}_i^r$ is an off-path history or $(T(r,a_i))_{a_i}$ is a deviation, then we assume that player i takes a best response from the next round. Here, we take f_i for round $\tilde{r} \geq r+1$ into account. This calculation is well defined since

- (23) implies that we can ignore discounting.[60]
- f_i for round $\tilde{r} \geq r+1$ has been defined since we proceed backward.
- Given the truthtelling incentive, the ex ante punishments from (54), (55), (56), (57), (58), (59), (60) and (61) are zero and can be ignored.
- The set of $(T(r,a_i))_{a_i}$'s that player i should take with positive probability on the equilibrium path is determined by a summary of player i's history $\mathfrak{h}_i^r \equiv \{\#_i^{\tilde{r}}\}_{\tilde{r} \leq r-1}$.
- Player j determines her continuation strategy ($\sigma_j(x_j)$, π_i^{main} and f_i for round $\tilde{r} \geq r+1$) treating each period within a round identically. Hence, given the truthtelling strategy in the report block, even after a deviation, $\mathfrak{h}_i^r$ and $\#_i^r$ are enough to calculate the best responses and the continuation value at the beginning of the next round. To calculate the ex ante value of taking $(T(r,a_i))_{a_i}$, we take the expectation of this continuation value using the conditional distribution of $\#_i^r$ given $(T(r,a_i))_{a_i}$. Since player j's strategy treats each period in round r identically, the timing of taking a_i does not change the expectation as long as $(T(r,a_i))_{a_i}$ is fixed.

Let $v(\hat{\mathfrak{h}}_i^r, \alpha_j(r), (T(r,a_i))_{a_i})$ denote this value. Let $(T^*(r,a_i))_{a_i}$ be $(T(r,a_i))_{a_i}$ that it gives the lowest ex ante value within those taken with positive probability in equilibrium.

We define f_i such that, if $\hat{\mathfrak{h}}_i^r$ is an on-path history,

- if $\left(\hat{T}(r,a_i)\right)_{a_i}$ is not an equilibrium strategy given $\hat{\mathfrak{h}}_i^r$, then

$$f_i(\hat{\mathfrak{h}}_i^r, \hat{\#}_i^r, \alpha_j(r)) = -T^{-r-5}, \tag{65}$$

[60] Since a player takes a mixed strategy when she minimaxes the opponent when we consider non-prisoners'-dilemma games in the Supplemental Materials, it is important to cancel out discounting so that this value is well defined without reporting what action is taken for each period.

and

- if $\left(\hat{T}(r,a_i)\right)_{a_i}$ is an equilibrium strategy given $\hat{\mathfrak{h}}_i^r$, then

$$f_i(\hat{\mathfrak{h}}_i^r, \hat{\#}_i^r, \alpha_j(r)) = 2\left(v(\hat{\mathfrak{h}}_i^r, \alpha_j(r), (T^*(r,a_i))_{a_i}) - v(\hat{\mathfrak{h}}_i^r, \alpha_j(r), \left(\hat{T}(r,a_i)\right)_{a_i})\right) \tag{66}$$

so that player i is indifferent between any equilibrium strategy. The term 2 represents

$$\frac{1}{\Pr(\text{player } i \text{ is picked by the public randomization})}. \tag{67}$$

On the other hand, if $\hat{\mathfrak{h}}_i^r$ is an off-path history, then f_i is defined to be 0.

We can take $f_i(\hat{\mathfrak{h}}_i^r, \hat{\#}_i^r, \alpha_j(r))$ satisfying (64) since (i) the original strategy is almost optimal, (ii) f_i for round $\tilde{r} \geq r+1$ is bounded by $[-T^{-r-6}, T^{-r-6}]$,[61] (iii) we have established the incentive to tell the truth, and (iv) from (iii) and Π_i, the ex ante punishments from (54), (55), (56), (57), (58), (59), (60) and (61) are zero.

(65) is enough to discourage any deviation after any history on the equilibrium path, considering all the continuation strategies after a deviation, that is, $\sigma(x)$ and the reward functions are a Nash equilibrium, by the following reasons: Since a strategy which takes non-constant actions in a review round is a deviation, the original strategy is almost optimal conditional on $\alpha_j(r)$, and the punishment in (65) for the current round is sufficiently larger than f_i for the later rounds, (65) incentivizes player i to take a constant action in each review round in equilibrium. Hence, the ex ante optimality at the beginning of each review round established by (66) is sufficient for the sequential optimality.

For round r where player i sends a message m, we replace $(T(r,a_i))_{a_i}$ with the set of possible messages.

We can proceed until the first round and show the optimality of $\sigma_i(x_i)$ recursively.

Finally, without the reward in the report block, for all $x \in \{G,B\}^2$, $\sigma_i(x_i)$ gives a payoff $\underline{v}_i$ for $x_j = B$ and $\bar{v}_i$ for $x_j = G$. In this section, we have established the exact optimality of

[61] For the last round, f_i for round $\tilde{r} \geq r+1$ does not exist.

$\sigma_i(x_i)$ conditional on x_j. Since the summation of the reward in the report block is bounded by T^{-1}, for all $x \in \{G, B\}^2$, $\sigma_i(x_i)$ is optimal against $\sigma_j(x_j)$ and gives a payoff close to $\underline{v}_i$ for $x_j = B$ and $\bar{v}_i$ for $x_j = G$. Since $\sigma_i(x_i)$ is optimal conditional on x_j, it is optimal for both players to send x_i truthfully in the coordination block (although player 2, the second sender, knows x_1 when she sends x_2).

References

[1] Abreu, D., P. Milgrom and D. Pearce (1991): "Information and Timing in Repeated Partnerships," *Econometrica*, 59, 1713-1733.

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

[3] Bhaskar, V. and I. Obara (2002): "Belief-Based Equilibria in the Repeated Prisoners' Dilemma with Private Monitoring," *Journal of Economic Theory*, 102, 49-69.

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

[5] Deb, D. (2011): "Cooperation and Community Responsibility: A Folk Theorem for Repeated Random Matching Games," *mimeo.*

[6] Ely, J., J. Hörner and W. Olszewski (2004): "Dispensability of Public Randomization Device," *mimeo.*

[7] Ely, J., J. Hörner and W. Olszewski (2005): "Belief-Free Equilibria in Repeated Games," *Econometrica*, 73, 377-415.

[8] Ely, J. and J. Välimäki (2002): "A Robust Folk Theorem for the Prisoners' Dilemma," *Journal of Economic Theory*, 102, 84-105.

[9] Fong, K., O. Gossner, J. Hörner and Y. Sannikov (2010): "Efficiency in a Repeated Prisoners' Dilemma with Imperfect Private Monitoring," *mimeo*.

[10] Fuchs, W. (2007): "Contracting with Repeated Moral Hazard and Private Evaluations," *American Economic Review*, 97, 1432-1448.

[11] Fudenberg, D. and D. Levine (2002): "The Nash-Threats Folk Theorem with Communication and Approximate Common Knowledge in Two Player Games," *Journal of Economic Theory*, 132, 461-473.

[12] Fudenberg, D., D. Levine and E. Maskin (1994): "The Folk Theorem with Imperfect Public Information," *Econometrica*, 62, 997-1040.

[13] Fudenberg, D. and E. Maskin (1986): "The Folk Theorem in Repeated Games with Discounting and with Incomplete Information," *Econometrica*, 54, 533-554.

[14] Harrington, J. E. Jr. and A. Skrzypacz (2007): "Collusion under Monitoring of Sales," *RAND Journal of Economics*, 38, 314-331.

[15] Harrington, J. E. Jr. and A. Skrzypacz (2011): "Private Monitoring and Communication in Cartels: Explaining Recent Collusive Practices," *American Economic Review*, forthcoming.

[16] Hörner, J. and W. Olszewski (2006): "The Folk Theorem for Games with Private Almost-Perfect Monitoring," *Econometrica*, 74, 1499-1544.

[17] Hörner, J. and W. Olszewski (2009): "How Robust is the Folk Theorem with Imperfect Public Monitoring?," *Quarterly Journal of Economics*, 124, 1773-1814.

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

[19] Kandori, M. (2010): "Weakly Belief-Free Equilibria in Repeated Games with Private Monitoring," *Econometrica*, 79, 877-892.

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

[21] Kandori, M. and I. Obara (2006): "Efficiency in Repeated Games Revisited: the Role of Private Strategies," *Econometrica*, 72, 499-519.

[22] Kandori, M. and I. Obara (2010): "Towards a Belief-Based Theory of Repeated Games with Private Monitoring: An Application of POMDP," *mimeo.*

[23] Lehrer, E. (1990): "Nash Equilibria of n-Player Repeated Games with Semi-Standard Information," *International Journal of Game Theory*, 19, 191–217.

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

[25] Mailath, G. and S. Morris (2006): "Coordination Failure in Repeated Games with Almost-Public Monitoring," *Theoretical Economics*, 1, 311-340.

[26] Mailath, G. and L. Samuelson (2006): *Repeated Games and Reputations: Long-Run Relationships.* Oxford University Press, New York, NY.

[27] Matsushima, H. (2004): "Repeated Games with Private Monitoring: Two Players," *Econometrica*, 72, 823-852.

[28] Miyagawa, E., Y. Miyahara and T. Sekiguchi (2008): "The Folk Theorem for Repeated Games with Observation Costs," *Journal of Economic Theory*, 139, 192-221.

[29] Obara, I. (2009): "Folk Theorem with Communication," *Journal of Economic Theory*, 144, 120-134.

[30] Phelan, C. and A. Skrzypacz (2011): "Beliefs and Private Monitoring," *mimeo*

[31] Piccione, M. (2002): "The Repeated prisoners' Dilemma with Imperfect Private Monitoring," *Journal of Economic Theory*, 102, 70-83.

[32] Radner, R. (1985): "Repeated Principal-Agent Games with Discounting," *Econometrica*, 53, 1173-1198.

[33] Sekiguchi, T. (1997): "Efficiency in Repeated prisoners' Dilemma with Private Monitoring," *Journal of Economic Theory*, 76, 345-361.

[34] Stigler, G. (1964): "A Theory of Oligopoly," Journal of Political Economy, 72, 44-61.

[35] Sugaya, T. (2010a): "Folk Theorem in a Repeated Prisoners' Dilemma without Conditional Independence," *mimeo.*

[36] Sugaya, T. (2010b): "Generic Characterization of the Set of Belief-Free Review-Strategy Equilibrium Payoffs," *mimeo.*

[37] Sugaya, T. and S. Takahashi (2011): "Coordination Failure in Repeated Games with Private Monitoring," *mimeo.*

[38] Takahashi, S. (2010): "Community Enforcement when Players Observe Partners' Past Play," *Journal of Economic Theory*, 145, 42-62.

[39] Yamamoto, Y. (2007): "Efficiency Results in N Player Games with Imperfect Private Monitoring," *Journal of Economic Theory*, 135, 382–413.

[40] Yamamoto, Y. (2009): "A Limit Characterization of Belief-Free Equilibrium Payoffs in Repeated Games," *Journal of Economic Theory*, 144, 802-824.

[41] Yamamoto, Y. (2011): "Characterizing Belief-Free Review-Strategy Equilibrium Payoffs under Conditional Independence," *mimeo.*