# Common Learning and Cooperation in Repeated Games*

Takuo Sugaya† and Yuichi Yamamoto‡

First Draft: December 10, 2012
This Version: October 15, 2019

**Abstract**

We study repeated games in which players learn the unknown state of the world by observing a sequence of noisy *private* signals. We find that for generic signal distributions, the folk theorem obtains using ex-post equilibria. In our equilibria, players *commonly* learn the state, that is, the state becomes asymptotic common knowledge.

*Journal of Economic Literature* Classification Numbers: C72, C73.

Keywords: repeated game, private monitoring, incomplete information, ex-post equilibrium, individual learning.

*The authors thank Michihiro Kandori, George Mailath, Stephen Morris, Andy Skrzypacz and seminar participants at various places for helpful comments.

†Stanford Graduate School of Business. Email: tsugaya@stanford.edu

‡Department of Economics, University of Pennsylvania. Email: yyam@sas.upenn.edu

# 1 Introduction

In many economic activities, agents face uncertainty about the underlying payoff structure, and experimentation is useful to resolve such a problem. Suppose that two firms enter a new market. The firms are not familiar with the structure of the market, and in particular do not know how profitable the market is (e.g., the intercept of the demand function). The firms interact repeatedly; every period, each firm chooses a price and then privately observes its sales level, which is stochastic due to an i.i.d. demand shock. Actions (prices) are perfectly observable. In this situation, the firms can eventually learn the true profitability of the market through sales; they may conclude that the market is profitable if they observe high sales frequently. However, since sales are private information, a firm faces uncertainty about whether the rival firm also believes that the market is profitable. Such *higher-order beliefs* have a significant impact on the firms' incentives: For example, suppose that choosing a high price is a "risky" action, in the sense that it yields a high profit only if the market is profitable enough *and* the rival firm also chooses a high price. Then even when a firm believes that the market is profitable, if it believes that the rival firm is pessimistic about the market profitability (and hence will choose a low price likely), it may prefer choosing a low price rather than a high price. Note also that each firm can manipulate the rival firm's belief (both the first and higher-order beliefs) via a *signaling effect*: Even if firm A believes that the market is not very profitable, it may still be tempted to choose a high price today, because by doing so, firm B updates the posterior upwards and starts to choose a high price in later periods, which is beneficial for firm A. Can the firms sustain collusion in such a situation? I.e., is there an equilibrium in which they can coordinate on the high price if the market is profitable, and on the low price if not? More generally, does a long-run relationship facilitate cooperation, when players *privately* learn the unknown economic state?

To address this question, we develop a general model of *repeated games with individual learning*. In our model, Nature moves first and chooses the state of the world $\omega$ (e.g., the market profitability in the duopoly market). The state is fixed throughout the game and is not observable to players. Then players play an infinitely repeated game. Each period, players observe private signals, whose distribution depends on the state. A player's stage-game payoff depends both

on actions and on her private signal, so the state (indirectly) influences expected payoffs through the signal distribution.

In general, when players have private information about the economic state, they can effectively coordinate their play if they *commonly learn* the state so that the state becomes almost common knowledge in the long run. Cripps, Ely, Mailath, and Samuelson (2008) show that common learning indeed occurs, if players learn the state from i.i.d. private signals. Unfortunately, their result does not apply to our setup, because (i) the signal distribution is influenced by actions, which are endogenously determined in equilibrium, and (ii) a player learns the state not only from her private signals, but from the opponents' actions. Accordingly, in our model, it is not obvious if common learning occurs. Another complication in our model is that while actions are perfectly observable, a player needs to rely on her private signals in order to detect the opponents' deviations, because in general the opponents choose different actions depending on their signals in equilibrium. In this sense, our model is a variant of repeated games with private monitoring, and it is well-known that finding an equilibrium in such a model is a hard problem (see Sugaya (2019), for example).

Despite such complications, we find that there indeed exist equilibria in which players commonly learn the state and obtain Pareto-efficient payoffs state by state. More generally, we find that the folk theorem holds so that any feasible and individually rational payoff (not only efficient outcomes) can be achievable as an equilibrium payoff. Our solution concept is an ex-post equilibrium, in that our equilibrium strategy is a sequential equilibrium regardless of the state; so it is an equilibrium even if the initial prior changes.[1] For a fixed discount factor $\delta$, the set of ex-post equilibrium payoffs is smaller than the set of sequential equilibrium payoffs, because providing ex-post incentives is more costly in general. However, it turns out that in our model, this cost becomes almost negligible as the discount factor approaches one, and accordingly we can obtain the folk theorem using ex-post equilibria.

[1] Some recent papers use ex-post equilibria in different settings of repeated games, such as perfect monitoring and fixed states (Hörner and Lovo (2009) and Hörner, Lovo, and Tomala (2011)), public monitoring and fixed states (Fudenberg and Yamamoto (2010) and Fudenberg and Yamamoto (2011a)), private monitoring and fixed states (Yamamoto (2014)), and changing states with an i.i.d. distribution (Miller (2012)). Note also that there are many papers working on ex-post equilibria in undiscounted repeated games; see Koren (1992) and Shalev (1994), for example.

To establish the folk theorem, we need the following two conditions. The first condition is the *statewise full-rank* condition, which requires that there be an action profile such that different states generate different signal distributions, even if someone unilaterally deviates. This condition ensures that each player can learn the true state from private signals, and that no one can stop the opponents' state learning. The second condition is the *correlated learning* condition. Roughly, it requires that signals be correlated across players, so that a player's signal is informative about the opponents'. These conditions are not only sufficient, but "almost necessary" for our result. Indeed, if the statewise full-rank condition does not hold, one can obtain a payoff significantly higher than the minimax payoff, by preventing the opponents' state learning. Also, if the correlated learning condition does not hold, we can construct an example in which the folk theorem cannot be obtained by ex-post equilibria. See the working paper version (Sugaya and Yamamoto (2019)) for more details.

Our proof of the folk theorem is constructive, and it builds on the idea of block strategies of Hörner and Olszewski (2006) and Wiseman (2012). For the sake of exposition, suppose for now that there are only two players and two states, $\omega_1$ and $\omega_2$. In our equilibrium, the infinite horizon is divided into a sequence of *blocks*. At the beginning of the block, each player $i$ chooses a *state-specific* plan about whether to reward or punish the opponent: Her plan is either "reward the opponent at both states," "punish the opponent at both states," "reward at state $\omega_1$ but punish at $\omega_2$," or "reward at state $\omega_2$ but punish at $\omega_1$." As will be explained shortly, the use of state-specific punishments is crucial in order to provide appropriate incentives in our environment.

In the first $T$ periods of the block, player 1 collects private signals and makes an inference $\omega(1)$ about the state $\omega$. Similarly, in the next $T$ periods, player 2 makes an inference $\omega(2)$ about the state. We take $T$ sufficiently large, so that each player $i$'s inference $\omega(i)$ matches the true state almost surely. Then in the next period, each player reports her inference $\omega(i)$ using actions, and check if they indeed agree on the state. Then depending on the reported information and on the plan chosen at the beginning of the block, they adjust the continuation play in the rest of the block. For example, if both players report $\omega_1$ and plan to reward each other at $\omega_1$, they will choose an action profile which yields high payoffs to both players at $\omega_1$. At the end of the block, (again, via actions) players report

their private signals during the learning phase in earlier periods; this information is used to make a minor modification to the continuation play (the punishment plan for the next block), which helps to provide right incentives. Once the block is over, a new block starts and players behave as above again.

It is important that players make the inference $\omega(i)$ based only on the signals during the current block; it does not depend on the signals in the previous blocks. This property ensures that even if someone makes a wrong inference (i.e., $\omega(i)$ does not match the true state), it does not have a long-run impact on payoffs. Indeed, in the next block, players can learn the true state with high probability and adjust the continuation play. This implies that even if a player deviates during the learning phase, its impact on a long-run payoff is not very large, which helps to deter such a deviation.

We find that this "learning, communication, and coordination" mechanism works effectively and approximates the Pareto-efficient frontier. Also, common learning occurs in this equilibrium. A key is that players communicate truthfully in our equilibrium, which makes (a piece of) their private information public and facilitates common learning. So in our equilibrium, a signaling effect helps to achieve common learning.

A critical step in our proof is to show that it is indeed possible to provide appropriate incentives for such truthful communication.[2] To provide such truthful incentives, signal correlation plays a crucial role. Recall that player $i$ makes an inference $\omega(i)$ using private signals pooled over the $T$-period interval. Since signals are correlated across players, the opponent's signal frequency $f_{-i}$ during this interval is informative about player $i$'s signal frequency $f_i$, and hence informative about player $i$'s inference $\omega(i)$. This suggests that the opponent can statistically distinguish player $i$'s misreport. A similar idea appears in the mechanism design literature (e.g., Crémer and Mclean (1988)), but a new complication is that the unknown state $\omega$ influences the signal correlation, which makes signals ambiguous. For example, there may be player $i$'s signal which is highly correlated with the opponent's signal $z_{-i}$ conditional on the state $\omega_1$, but is correlated with a different signal $\tilde{z}_{-i}$ conditional on the state $\omega_2$.

[2]Allowing cheap-talk communication does not simplify our analysis, due to this problem; we need to find a mechanism under which players report truthfully in the cheap-talk communication stage, and it is essentially the same as the problem we consider here.

To deter player $i$'s misreport using such ambiguous signals, state-contingent punishments are helpful.[3] A rough idea is that the opponent interprets her signal frequency $f_{-i}$ *taking a state* $\omega$ *as given*, and decides whether to punish player $i$ or not for that state $\omega$. For example, suppose that the opponent's signal frequency $f_{-i}$ is typical of the state $\omega_1$, i.e., it is close to the true signal distribution at $\omega_1$. Then *conditional on the state* $\omega_1$, the opponent believes that player $i$'s observation is also typical of the state $\omega_1$ and hence $i$'s inference is $\omega(i) = \omega_1$. On the other hand, *conditional on the state* $\omega_2$, the opponent *may not* believe that player $i$'s inference is $\omega(i) = \omega_1$, since signals are interpreted differently at different states. Suppose now that player $i$ reports $\omega(i) = \omega_1$. Should the opponent punish player $i$? The point is that this report is consistent with the opponent's signals conditional on the state $\omega_1$, but not conditional on $\omega_2$. This suggests that the opponent should punish player $i$ *only at the state* $\omega_2$, by playing a continuation strategy which yields a low payoff to player $i$ conditional on the state $\omega_2$ but a high payoff conditional on $\omega_1$. That is, the opponent should choose the plan "reward player $i$ at $\omega_1$ but punish at $\omega_2$" more likely in the next block.

In the proof, we carefully construct such a state-contingent punishment mechanism so that player $i$'s misreport is indeed deterred. In particular, we find that there is a punishment mechanism such that

(i) If everyone reports truthfully, the probability of a punishment being triggered is almost negligible.

(ii) The truthful report is ex-post incentive compatible, that is, regardless of the true state $\omega$ and the true inference $\omega(i)$, reporting $\omega(i)$ truthfully is a best reply for each player $i$.

The first property ensures that even though a punishment destroys the total welfare (players choose inefficient actions once it is triggered), the equilibrium payoff can still approximate the Pareto-efficient outcome.[4] The second property implies that

[3] For this idea to work, it is crucial that players' signals are correlated *conditional on each state* $\omega$. Indeed, if not and signals are independent at some state $\omega$, they are uninformative about the opponent's observation conditional on that state, and hence not useful to detect the opponent's misreport.

[4] Fudenberg, Levine, and Maskin (1994) show that this inefficiency can be avoided if continuation payoffs take the form of "utility transfers." Unfortunately this technique does not seem to apply to our setup, because players condition their play on their private signals.

any misreport is not profitable, regardless of player $i$'s belief about the state $\omega$. This in particular implies that player $i$'s history in the previous blocks, which influences her belief about $\omega$, is irrelevant to her incentive in the current block; her incentive is solely determined by her history within the current block. This allows us to use a recursive technique to construct an equilibrium in the infinite-horizon game.

The design of the state-contingent punishment mechanism is a bit complicated, because player $i$'s belief about the opponent's signal frequency $f_{-i}$ is also influenced by the unknown state $\omega$. For example, suppose that player $i$'s signal frequency $f_i$ during the $T$ period interval is typical of the state $\omega_1$, so that her inference is $\omega(i) = \omega_1$. With such an observation $f_i$, *conditional on the state* $\omega_1$, she believes that the opponent believes that player $i$'s inference is $\omega(i) = \omega_1$, and hence the truthful report of $\omega(i) = \omega_1$ is a best reply. However, *conditional on the state* $\omega_2$, she *need not* believe that the opponent believes $\omega(i) = \omega_1$ in general. So to satisfy the property (ii) above, we need to carefully design a (state-contingent) punishment mechanism for the state $\omega_2$, that is, reporting $\omega(i) = \omega_1$ must be a best reply for player $i$ at $\omega_2$ even though she does not expect the opponent to believe $\omega(i) = \omega_1$. More generally, we need to find a mechanism with which for each given observation $f_i$, player $i$'s best reply does not depend on the state (the truthful report of $\omega(i)$ must be a best reply at *both* states), even though her belief about the opponent's belief depends on the state. One way to solve this problem is to let the opponent make player $i$ indifferent over all reports, regardless of the observation $f_{-i}$; then the truthful report of $\omega(i)$ is always a best reply for player $i$. But it turns out that such a mechanism does not satisfy the property (i) above and causes inefficiency, that is, a punishment is triggered with positive probability and destroys the total welfare even on the equilibrium path.[5] To avoid such inefficiency while maintaining truthful incentives, we consider a mechanism in which the opponent makes player $i$ indifferent only after *some* (but not all) observations $f_{-i}$. It turns out that this idea "almost" solves our problem, that is, it allows us to construct a mechanism in which the truthful report of the summary inference $\omega(i)$ is an *approximate* best reply regardless of the past history, while minimizing the welfare destruction. Of course, this is not an exact solution to our problem, as we

[5]This is similar to the fact that belief-free equilibria of Ely, Hörner, and Olszewski (2005) cannot attain the Pareto-efficient outcome when monitoring is imperfect.

need the truthful report to be an *exact* best reply. To fix this problem, in the last step of the proof, we modify the equilibrium strategy a bit; we let players reveal her signal sequence during the learning phase (this is different from $\omega(i)$, which is just a summary statistics of the observed signals) at the end of each block, and use this information to provide an extra incentive to report the summary inference $\omega(i)$ truthfully. See Section 4.5 for more details.

Fudenberg and Yamamoto (2010) also use the idea of state-contingent punishments, but their proof is not constructive. In particular, both state learning process and intertemporal incentives are implicitly described through the motion of continuation payoffs. The interaction of these two forces complicates the motion of continuation payoffs, which makes it difficult to see how players learn the state in equilibrium, and how they use this information to punish a deviator. In contrast, our proof is constructive, and we explicitly describe how each player learns the state in each block and chooses a state-contingent punishment plan. We hope that this helps to understand the role of state-contingent punishment in a more transparent way.

Throughout this paper, we assume that actions are perfectly observable. But this assumption is not crucial; in the working paper version (Sugaya and Yamamoto (2019)), we extend the analysis to the case in which actions are not observable. In this new setup, players need to monitor the opponents' actions only through noisy private signals, whose distribution is influenced by the unknown state $\omega$. So it is a repeated game with *private monitoring* and *unknown monitoring structure*. We find that the folk theorem still holds when the identifiability conditions are strengthened. This result generalizes various efficiency theorems for repeated games with private monitoring[6] (in particular the folk theorem of Sugaya (2019)) to the case in which the monitoring structure is unknown.

To the best of our knowledge, this is the first paper which considers common

[6]For example, the efficient outcome is approximately achieved in the prisoner's dilemma, when observations are nearly perfect (Sekiguchi (1997), Bhaskar and Obara (2002), Piccione (2002), Ely and Välimäki (2002), Yamamoto (2007), Yamamoto (2009), Hörner and Olszewski (2006), Chen (2010), and Mailath and Olszewski (2011)), nearly public (Mailath and Morris (2002), Mailath and Morris (2006), and Hörner and Olszewski (2009)), statistically independent (Matsushima (2004), Yamamoto (2012)), and even fully noisy and correlated (Kandori (2011), Fong, Gossner, Hörner and Sannikov (2011), Sugaya (2012), and Sugaya (2019)). Kandori (2002) and Mailath and Samuelson (2006) are excellent surveys. See also Lehrer (1990) for the case of no discounting, and Fudenberg and Levine (1991) for the study of approximate equilibria with discounting.

learning with strategic players.[7] Cripps, Ely, Mailath, and Samuelson (2008) and Cripps, Ely, Mailath, and Samuelson (2013) consider the case in which players are not strategic, i.e., players observe private signals about the state each period, without taking actions. They find that common learning occurs when signals are i.i.d., but it does not occur in general for non-i.i.d. signals. Our work extends their analysis by considering strategic players; now signal distributions are non-i.i.d., and *endogenously* determined by players' equilibrium play. It turns out that players' strategic behavior has a substantial impact on the joint learning outcome; we find that with strategic players, the negative result overturns and common learning occurs in general, thanks to the signaling effect discussed above.

Our work belongs to the literature on learning in repeated games. Most of the existing work assumes that players observe *public* (or almost public) signals about the state, and focuses on equilibria in which players ignore private information. (Wiseman (2005), Wiseman (2012), Fudenberg and Yamamoto (2010), Fudenberg and Yamamoto (2011a)). An exception is Yamamoto (2014), who considers the case in which players learn from private signals only. The difference from this paper is that he focuses on *belief-free equilibria*, which are a subset of sequential equilibria. An advantage of belief-free equilibrium is its tractability; it does not require players' coordination, and a player's higher-order belief is payoff-irrelevant. But unfortunately, its payoff set is bounded away from the Pareto-efficient frontier in general, due to the lack of coordination. In order to avoid such inefficiency, we consider sequential equilibria in which players coordinate their play through communication. As noted earlier, a player's best reply in such communication is very sensitive to her higher-order belief (her belief about the opponent's signals), which makes our analysis quite different from the ones in the literature.

# 2 Repeated Games with Individual Learning

Given a finite set $X$, let $\triangle X$ be the set of probability distributions over $X$. Given a subset $W$ of $\mathbb{R}^n$, let co$W$ denote the convex hull of $W$.

[7] A recent paper by Basu, Chatterjee, Hoshino, and Tamuz (2017) considers a similar question, but their analysis is quite different from ours because (i) they impose a special assumption on the payoff function (there are only two actions, and one of them is a dominant action) and (ii) they assume conditionally independent signals.

We consider an $N$-player infinitely repeated game, in which the set of players is denoted by $I = \{1, \cdots, N\}$. At the beginning of the game, Nature chooses the state of the world $\omega$ from a finite set $\Omega$. Assume that players cannot observe the true state $\omega$, and let $\mu \in \triangle\Omega$ denote their common prior over $\omega$.[8] Throughout the paper, we assume that the game begins with symmetric information: Each player's initial belief about $\omega$ is equal to the prior $\mu$. But it is straightforward to extend our analysis to the asymmetric-information case as in Fudenberg and Yamamoto (2011a).[9]

Each period, players move simultaneously, and each player $i \in I$ chooses an action $a_i$ from a finite set $A_i$. The chosen action profile $a \in A \equiv \times_{i \in I} A_i$ is publicly observable, and in addition, each player $i$ receives a private signal $z_i$ about the state $\omega$ from a finite set $Z_i$. The distribution of the signal profile $z \in Z \equiv \times_{i \in I} Z_i$ depends on the state of the world $\omega$ and on the action profile $a \in A$, and is denoted by $\pi^\omega(\cdot|a) \in \triangle Z$. Let $\pi_i^\omega(\cdot|a)$ denote the marginal distribution of player $i$'s signal $z_i$ given $\omega$ and $a$, that is, $\pi_i^\omega(z_i|a) = \sum_{z_{-i} \in Z_{-i}} \pi^\omega(z|a)$. Likewise, let $\pi_{-i}^\omega(\cdot|a)$ be the marginal distribution of the opponents' signals $z_{-i}$. Player $i$'s payoff is $u_i^\omega(a, z_i)$, so her expected payoff given the state $\omega$ and the action profile $a$ is $g_i^\omega(a) = \sum_{z_i \in Z_i} \pi_i^\omega(z_i|a) u_i^\omega(a, z_i)$.[10] Let $g^\omega(a) = (g_i^\omega(a))_{i \in I}$ be the payoff vector given $\omega$ and $a$. As usual, we write $\pi^\omega(\alpha)$ and $g_i^\omega(\alpha)$ for the signal distribution and the expected payoff when players play a mixed action profile $\alpha \in \times_{i \in I} \triangle A_i$. Similarly, we write $\pi^\omega(a_i, \alpha_{-i})$ and $g_i^\omega(a_i, \alpha_{-i})$ for the signal distribution and the expected payoff when players $-i$ play a mixed action $\alpha_{-i} \in \times_{j \neq i} \triangle A_j$.

As emphasized in the introduction, uncertainty about the payoff functions is common in applications. Examples that fit our model include:

- Oligopoly market with unknown demand function. Often times, firms do

[8] Because our arguments deal only with ex-post incentives, they extend to games without a common prior. However, as Dekel, Fudenberg, and Levine (2004) argue, the combination of equilibrium analysis and a non-common prior is hard to justify.

[9] Specifically, all the results in this paper extend to the case in which each player $i$ has initial private information $\theta_i$ about the true state $\omega$, where the set $\Theta_i$ of player $i$'s possible private information is a partition of $\Omega$. Given the true state $\omega \in \Omega$, player $i$ observes $\theta_i^\omega \in \Theta_i$, where $\theta_i^\omega$ denotes $\theta_i \in \Theta_i$ such that $\omega \in \theta_i$. In this setup, private information $\theta_i^\omega$ allows player $i$ to narrow down the set of possible states; for example, player $i$ knows the state if $\Theta_i = \{(\omega_1), \cdots, (\omega_o)\}$.

[10] If there are $\omega \in \Omega$ and $\tilde{\omega} \neq \omega$ such that $u_i^\omega(a, z_i) \neq u_i^{\tilde{\omega}}(a, z_i)$ for some $a_i \in A_i$ and $z \in Z$, then it might be natural to assume that player $i$ does not observe the realized value of $u_i$ as the game is played; otherwise players might learn the true state from observing their realized payoffs. Since we consider ex-post equilibria, we do not need to impose such a restriction.

not have precise information about the market structure, and such a situation is a special example of our model. To see this, let $I$ be the set of firms, $a_i$ be firm $i$'s price, and $z_i$ be firm $i$'s sales level. The distribution $\pi^\omega(\cdot|a)$ of sales levels depends on the unknown state $\omega$, which means that the firms do not know the true distribution of the sales level.

- Team production and private benefit. Consider agents working on a joint project who do not know the profitability of the project; they may learn the true profitability through their experience over time. To describe such a situation, let $I$ be the set of agents, $a_i$ be agent $i$'s effort level, and $z_i$ be agent $i$'s private profit from the project. The distribution $\pi^\omega(\cdot|a)$ of private profits depends on the unknown state $\omega$, so the agents learn the true distribution through their observations over time.

In the infinitely repeated game, players have a common discount factor $\delta \in (0,1)$. Let $(a^\tau, z_i^\tau) \in A \times Z_i$ be player $i$'s private observation in period $\tau$, and let $h_i^t = (a^\tau, z_i^\tau)_{\tau=1}^t$ be player $i$'s private history until period $t \geq 1$. Let $h_i^0 = \emptyset$, and for each $t \geq 0$, and let $H_i^t$ be the set of all private histories $h_i^t$. Let $h^t = (h_i^t)_{i \in I}$ denote a profile of $t$-period private histories, and $H^t$ be the set of all history profiles $h^t$. A strategy for player $i$ is defined to be a mapping $s_i : \bigcup_{t=0}^\infty H_i^t \to \triangle A_i$. Let $S_i$ be the set of all strategies for player $i$, and let $S = \times_{i \in I} S_i$.

The feasible payoff set for a given state $\omega$ is defined as

$$V(\omega) \equiv \text{co}\{g^\omega(a) | a \in A\},$$

that is, $V(\omega)$ is the convex hull of possible stage-game payoff vectors at the state $\omega$. Then the feasible payoff set for the overall game is defined as

$$V \equiv \times_{\omega \in \Omega} V(\omega).$$

Thus each feasible payoff vector $v \in V$ specifies payoffs for each player and for each state, i.e., $v = ((v_1^\omega, \cdots, v_N^\omega))_{\omega \in \Omega}$. Note that a given $v \in V$ may be generated using different action distributions at different states $\omega$. We will show that there are equilibria which approximate payoffs in $V$ if the state is statistically identified by private signals so that players learn it over time.

Player $i$'s minimax payoff for a given state $\omega$ is defined as

$$m_i^\omega \equiv \min_{\alpha_{-i}} \max_{a_i} g_i^\omega(a_i, \alpha_{-i}).$$

Let $\underline{\alpha}^{\omega}(i)$ denote the (possibly mixed) minimax action profile against player $i$ conditional on $\omega$. Let $V^*$ be the set of feasible and individually rational payoffs, that is,

$$V^* \equiv \{v \in V | v_i^{\omega} \geq m_i^{\omega} \ \ \forall i \forall \omega\}.$$

Here the individual rationality is imposed state by state; i.e., $V^*$ is the set of feasible payoffs such that each player obtains at least her minimax payoff for each state $\omega$.[11] Throughout the paper, we assume that the set $V^*$ is full dimensional:

**Condition 1. (Full Dimension)** $\dim V^* = |I| \times |\Omega|$.

# 3 The Folk Theorem with Individual Learning

In this section, we will present our main result, the folk theorem for games with individual learning. In our equilibrium, common learning occurs, so that the state becomes approximate common knowledge, even though players learn the state from private signals.

We will use an ex-post equilibrium as an equilibrium concept:

[11] If there are only two players and our Condition 2 holds, the minimax payoff $m_i^{\omega}$ indeed characterizes player $i$'s minimum equilibrium payoff in the limit as $\delta \to 1$. Precisely, we can show that for any $v_i < \sum_{\omega \in \Omega} \mu(\omega) m_i^{\omega}$, there is $\overline{\delta} \in (0,1)$ such that for any $\delta \in (\overline{\delta}, 1)$, player $i$'s expected payoff (here we consider the expected payoff given the initial prior $\mu$) is at least $v_i$ for all Nash equilibria. For simplicity, suppose that there are only two states, $\omega$ and $\tilde{\omega}$. (It is not difficult to extend the argument to the case with more than two states.) Fix an arbitrary Nash equilibrium $\sigma$. Let $a^*$ be as in Condition 2, and let $\sigma_i^T$ be player $i$'s strategy with the following form:

- Play $a^*$ for the first $T$ periods, and make an inference $\omega(i)$ as in Lemma 1.
- In each period $t > T$, choose $a_i \in \arg\max g_i^{\omega(i)}(\tilde{a}_i, \alpha_{-i}|_{\omega(i), h_i^{t-1}})$ where $\alpha_{-i}|_{\omega^*, h_i^{t-1}}$ is the distribution of the opponent's actions conditional on the history $h_i^{t-1}$ and the true state $\omega^*$.

From Lemma 1 (i) and (ii), the probability that $\omega(i)$ coincides with the true state is at least $1 - 2\exp(-T^{\frac{1}{2}})$, regardless of the opponent's play. Hence if player $i$ deviates to $\sigma_i^T$, her payoff is at least

$$(1-\delta^T)\underline{g}_i + \delta^T \sum_{\omega^* \in \Omega} \mu(\omega^*) \left\{ \left(1 - 2\exp(-T^{\frac{1}{2}})\right) m_i^{\omega^*} + 2\exp(-T^{\frac{1}{2}})\underline{g}_i \right\}$$

where $\underline{g}_i = \min_{\omega, a} g_i^{\omega}(a)$. Player $i$'s equilibrium payoff is at least this deviation payoff, which approximates $\sum_{\omega \in \Omega} \mu(\omega) m_i^{\omega}$ when we take $\delta \to 1$ and then $T \to \infty$. This proves the above claim.

When there are more than two players, player $i$'s minimum equilibrium payoff can be below $\sum_{\omega \in \Omega} \mu(\omega) m_i^{\omega}$ even in the limit as $\delta \to 1$. This is because the opponents may be able to use correlated actions to punish player $i$, when private signals are correlated.

**Definition 1.** A strategy profile $s$ is an *ex-post equilibrium* if it is a sequential equilibrium in the infinitely repeated game in which $\omega$ is common knowledge for each $\omega$.

In an ex-post equilibrium, after every history $h^t$, player $i$'s continuation play is a best reply regardless of the true state $\omega$. Hence these equilibria are robust to a perturbation of the initial prior, that is, an ex-post equilibrium is a sequential equilibrium given any initial prior.

We will provide a set of conditions under which the folk theorem is established using ex-post equilibria. Our first condition is the statewise full-rank condition of Yamamoto (2014), which requires that there be an action profile such that each player $i$ can learn the true state $\omega$ from her private signal $z_i$:

**Condition 2. (Statewise Full Rank)** There is an action profile $a^* \in A$ such that $\pi_i^{\omega}(\cdot|a_j, a^*_{-j}) \neq \pi_i^{\tilde{\omega}}(\cdot|a_j, a^*_{-j})$ for each $i$, $j \neq i$, $a_j$, $\omega$, and $\tilde{\omega} \neq \omega$.

Intuitively, the statewise full rank implies that player $i$ can statistically distinguish $\omega$ from $\tilde{\omega}$ through her private signal $z_i$, even if someone else unilaterally deviates from $a^*$.[12] We fix this profile $a^*$ throughout the paper. Note that Condition 2 is satisfied for generic signal structures if $|Z_i| \geq 2$ for each $i$.

Our next condition is about the correlation of players' private signals. The following notation is useful. Let $\pi^{\omega}(z_{-i}|a, z_i)$ denote the conditional probability of $z_{-i}$ given that the true state is $\omega$, players play an action profile $a$, and player $i$ observes $z_i$; i.e.,

$$\pi^{\omega}(z_{-i}|a, z_i) = \frac{\pi^{\omega}(z|a)}{\pi_i^{\omega}(z_i|a)}.$$

Let $\pi^{\omega}(z_{-i}|a, z_i) = 0$ if $\pi_i^{\omega}(z_i|a) = 0$. Then let $C_i^{\omega}(a)$ be the matrix such that the rows are indexed by the elements of $Z_{-i}$, the columns are indexed by the elements of $Z_i$, and the $(z_{-i}, z_i)$-component is $\pi^{\omega}(z_{-i}|a, z_i)$. Intuitively, the matrix $C_i^{\omega}(a)$ maps player $i$'s observations to her estimate (expectation) of the opponents'

[12] This condition is stronger than necessary. For example, our proof extends with no difficulty as long as for each $(i, \omega, \tilde{\omega})$ with $\omega \neq \tilde{\omega}$, there is an action profile $a$ such that $\pi_i^{\omega}(\cdot|a'_j, a_{-j}) \neq \pi_i^{\tilde{\omega}}(\cdot|a'_j, a_{-j})$ for each $j \neq i$ and $a'_j$. That is, each player may use different action profiles to distinguish different pairs of states. But it significantly complicates the notation with no additional insights. Also, while Condition 2 requires that all players can learn the state from private signals, it is easy to see that our proof is valid as long as there are at least two players who can distinguish the state.

observations *conditional on the true state being* $\omega$. To get the precise meaning, suppose that players played an action profile $a$ for $T$ periods, and player $i$ observed a signal sequence $(z_i^1, \cdots, z_i^T)$. Let $f_i \in \triangle Z_i$ denote the corresponding signal frequency, i.e., let $f_i = (f_i[z_i])_{z_i \in Z_i}$ where $f_i[z_i] = \frac{|\{t \leq T | z_i^t = z_i\}|}{T}$ for each $z_i$. Given this observation $f_i$ (and given the true state being $\omega$), the conditional expectation of the opponents' signal frequency during these $T$ periods is represented by $C_i^\omega(a) f_i$. So the matrix $C_i^\omega(a)$ converts player $i$'s signal frequency $f_i$ to her estimate of the opponents' signal frequencies, when the state $\omega$ is given.

**Condition 3. (Correlated Learning)** $C_i^\omega(a^*)\pi_i^{\tilde{\omega}}(a^*) \neq \pi_{-i}^\omega(a^*)$ for each $i$ and for each $(\omega, \tilde{\omega})$ with $\omega \neq \tilde{\omega}$.

Roughly, this condition requires that signals are correlated across players,[13] so that if a player observes some "unusual" signal frequency, then she believes that the opponent's observation is also "unusual." To better understand, suppose that players played $a^*$ for a while and player $i$'s signal frequency was exactly the expected distribution $\pi_i^{\tilde{\omega}}(a^*)$ for some state $\tilde{\omega}$. Note that this signal frequency is "unusual" if the true state were $\omega \neq \tilde{\omega}$. Condition 3 requires that in this case, player $i$ believes that conditional on the state $\omega$, the opponent's signal frequency is also "unusual" and different from the ex-ante distribution $\pi_{-i}^\omega(a^*)$. This condition holds for generic signal structures, since it can be satisfied by (almost all) small perturbations of the matrix $C_i^\omega(a^*)$.[14]

The assumption above is weaker than that of Wiseman (2012), who prove the folk theorem for the case in which signals are almost public (i.e., highly correlated). He assumes that there is a cheap-talk game after each stage game, and considers an equilibrium in which players report their private signals truthfully in this cheap-talk game. High correlation of signals is useful to provide incentives for the truthful report. The idea is that when signals are highly correlated, a

[13] Condition 3 is only about signal correlation conditional on the action $a^*$; we allow the signals to be independent across players if someone deviates from $a^*$. In the proof of the folk theorem, we will construct an equilibrium in which such a deviation is not profitable. A key is that actions are observable, so players can detect such a deviation for sure and punish a deviator. See the proof of Lemma 3 for more details.

[14] Condition 3 does not hold if signals are conditionally independent, in that $\pi^\omega(z|a) = \prod_{i \in I} \pi_i^\omega(z_i|a)$ for all $\omega$ and $a$. In the working paper version (Sugaya and Yamamoto (2019)), we present an example with conditionally independent signals in which ex-post equilibria cannot approximate the Pareto-efficient frontier.

player's signal today is very informative about the opponent's observation, so any misreport can be easily detected. In contrast, our assumption (Conditon 3) needs only a minimal amount of correlation. In the proof, we consider an equilibrium in which players report their inference, which is a *summary statistic* of observations over $T$ periods. When $T$ is large enough and signals are correlated, the law of large numbers ensures that players have accurate information about the opponent's inference after *some* (but not all) histories. As we will show, this property is enough to provide appropriate incentives.

We can prove that a folk theorem under Conditions 1 through 3. However, the proof for a general case is fairly complex, so in this paper, we will focus on the special case in which (i) there are only two players and two states and (ii) each player has sufficiently many actions. These extra assumptions significantly simplifies the notation and the proof, and we believe that this is the best way to illustrate our key ideas; in the proof, we will show that there are equilibria in which players learn the state from signals and communicate it via actions. See the working paper version (Sugaya and Yamamoto (2019)) for how this idea extends to a general case.

The following is the formal statement of our folk theorem. It asserts that there are ex-post equilibria in which players eventually obtain payoffs as if they knew the true state and played an equilibrium for that state.

**Proposition 1.** *Suppose that* $|I| = |\Omega| = 2$ *and* $|A_i| \geq |Z_i|$.[15] *Suppose also that Conditions 1 through 3 hold. Then the folk theorem holds, i.e., for any* $v \in intV^*$*, there is* $\overline{\delta} \in (0,1)$ *such that for any* $\delta \in (\overline{\delta}, 1)$*, there is an ex-post equilibrium with payoff* $v$*.*

Proposition 1 is about players' equilibrium payoffs, and it does not say anything about how players' beliefs about the state change over time. In the working paper version (Sugaya and Yamamoto (2019)), we show that in any equilibrium constructed in the proof of Proposition 1, the state asymptotically becomes common knowledge among players. A rough idea is as follows.

In the proof of Proposition 1, we consider an equilibrium in which (i) players learn the state from private signals, (ii) report these signals truthfully via actions,

[15]This assumption greatly simplifies the structure of the "detailed report round" which will appear in the proof.

ad then (iii) adjust the continuation play depending on the report. Thanks to the communication in Phase (ii), all private signals in Phase (i) become public information. (This is what we meant by the signalling effect in the introduction.) Also, in our equilibrium, signals observed in Phases (ii) and (iii) do not influence the continuation play at all, that is, there is no correlation between these signals and the chosen actions. So the only way to learn the opponent's signals observed in Phases (ii) and (iii) is to use correlation of private signals. As shown in Cripps, Ely, Mailath, and Samuelson (2008), in such a situation, the state asymptotically becomes common knowledge. Hence the result follows.

# 4 Proof of Proposition 1

Fix an arbitrary payoff vector $v \in \text{int}V^*$. We will construct an ex-post equilibrium with payoff $v$, by extending the idea of block strategies of Hörner and Olszewski (2006). A key difference from Hörner and Olszewski (2006) is that in our equilibrium, each player makes an inference about the state $\omega$ from private signals, and publicly reports it in order to coordinate the continuation play. A crucial step in the proof is how to induce the truthful report of the inference.

For each state $\omega$, we choose four values, $\underline{v}_1^\omega$, $\underline{v}_2^\omega$, $\overline{v}_1^\omega$, and $\overline{v}_2^\omega$, as in Figure 1. That is, we choose these values so that the rectangle $\times_{i\in I}[\underline{v}_i^\omega, \overline{v}_i^\omega]$ is in the interior of the feasible and individually rational payoff set for $\omega$, and contains the payoff $v$. Looking ahead, these values are "target payoffs" in our equilibrium: We will construct an equilibrium in which player $i$'s payoff in the continuation game conditional on the state $\omega$ is $\underline{v}_i^\omega$ if the opponent plans to punish her, and $\overline{v}_i^\omega$ if the opponent plans to reward her.

For each state $\omega$, we take four action profiles, $a^{\omega,GG}$, $a^{\omega,GB}$, $a^{\omega,BG}$, and $a^{\omega,BB}$ such that the corresponding stage-game payoffs surround the rectangle, as in Figure 1. Formally, choose these profiles so that[16]

$$\max\{g_1^\omega(a^{\omega,BB}), g_1^\omega(a^{\omega,GB})\} < \underline{v}_1^\omega < \overline{v}_1^\omega < \min\{g_1^\omega(a^{\omega,GG}), g_1^\omega(a^{\omega,BG})\}$$

[16]For some payoff function, such action profiles $a^{\omega,x^\omega}$ may not exist. In this case, as in Hörner and Olszewski (2006), we take action sequences $(a^{\omega,x^\omega}(1), \cdots, a^{\omega,x^\omega}(n))$ instead of action profiles; the rest of the proof extends to this case with no difficulty.

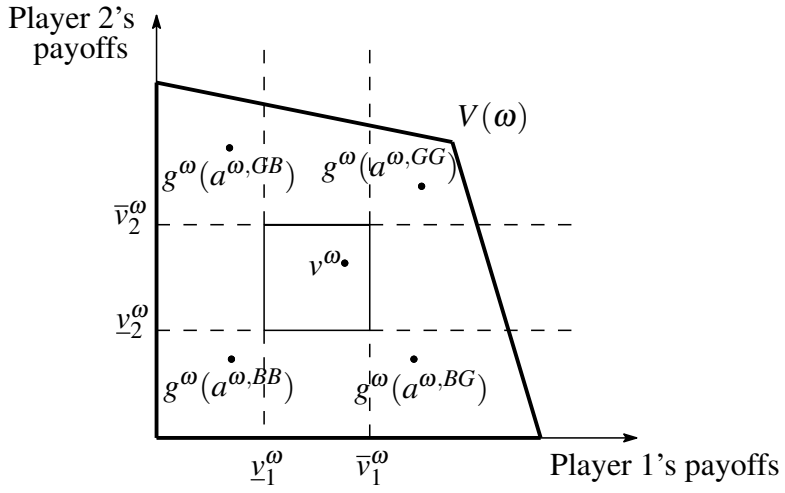

Figure 1: Actions $a^{\omega,GG}$, $a^{\omega,GB}$, $a^{\omega,BG}$, and $a^{\omega,BB}$

and

$$\max\{g_2^\omega(a^{\omega,BB}), g_2^\omega(a^{\omega,BG})\} < \underline{v}_2^\omega < \overline{v}_2^\omega < \min\{g_2^\omega(a^{\omega,GG}), g_2^\omega(a^{\omega,GB})\}.$$

Intuitively, the $i$th capital letter in the superscript ($G$ for good, and $B$ for bad) describes whether player $i$ plans to reward or punish the opponent. Player $i$'s payoff is above $\overline{v}_i^\omega$ when the opponent rewards her, and is below $\underline{v}_i^\omega$ when the opponent punishes her. Note that the definition of these action profiles is very similar to that in Hörner and Olszewski (2006).

Then we pick $\varepsilon > 0$ sufficiently small so that all the following conditions hold:

- For each $\omega$,

$$\max\{g_1^\omega(a^{\omega,GB}), g_1^\omega(a^{\omega,BB}), m_1^{\omega_1}\} < \underline{v}_1^\omega - \varepsilon, \tag{1}$$

$$\max\{g_2^\omega(a^{\omega,BG}), g_2^\omega(a^{\omega,BB}), m_2^{\omega_2}\} < \underline{v}_2^\omega - \varepsilon, \tag{2}$$

$$\min\{g_1^\omega(a^{\omega,GG}), g_1^\omega(a^{\omega,BG})\} > \overline{v}_1^\omega + 2\varepsilon, \tag{3}$$

$$\min\{g_2^\omega(a^{\omega,GG}), g_2^\omega(a^{\omega,GB})\} > \overline{v}_2^\omega + 2\varepsilon. \tag{4}$$

- For each $\omega$ and $\tilde{\omega} \neq \omega$,

$$|\pi_{-i}^\omega(a^*) - C_i^\omega(a^*)\pi_i^{\tilde{\omega}}(a^*)| > 2\sqrt{\varepsilon}. \tag{5}$$

- For each $\omega$, $\tilde{\omega} \neq \omega$, and $f_i \in \triangle Z_i$ with $|\pi_i^{\tilde{\omega}}(a^*) - f_i| < \varepsilon$,

$$|C_i^\omega(a^*)\pi_i^{\tilde{\omega}}(a^*) - C_i^\omega(a^*)f_i| < \sqrt{\varepsilon}. \tag{6}$$

Note that (5) indeed holds for small $\varepsilon$, thanks to Condition 3. Similarly, (1) through (4) follow from the definition of $a^{\omega,GG}$, $a^{\omega,GB}$, $a^{\omega,BG}$, and $a^{\omega,BB}$, and the fact that $\underline{v}_i^\omega$ is larger than the minimax payoff $m_i^\omega$. (5) follows from Condition 3. (6) simply says that if an observation $f_i$ is close to $\pi_i^{\tilde{\omega}}(a^*)$, then the posterior belief $C_i^\omega(a^*)f_i$ is close to $C_i^\omega(a^*)\pi_i^{\tilde{\omega}}(a^*)$. This inequality holds for any small $\varepsilon$, because $\frac{\sqrt{\varepsilon}}{\varepsilon} \to \infty$ as $\varepsilon \to 0$. In the rest of the proof, we fix this parameter $\varepsilon$.

## 4.1 Automaton with State-Contingent Punishment

In our equilibrium, the infinite horizon is divided into a series of *blocks* with length $T_b$, where a parameter $T_b$ is to be specified. Each player $i$'s equilibrium strategy is described as an automaton strategy over blocks. At the beginning of the block, she chooses an automaton state $x_i$ from the set $X_i = \{GG, GB, BG, BB\}$. (So there are four possible automaton states.) This automaton state $x_i$ determines her play during the block; player $i$ with an automaton state $x_i$ plays a block strategy $s_i^{x_i}$ (to be specified). See Figure 2.

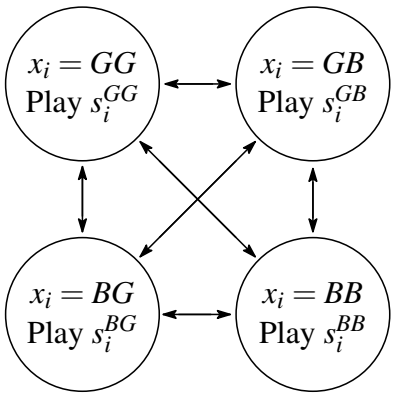


Figure 2: Automaton

The automaton state $x_i$ can be interpreted as player $i$'s *state-contingent plan* about whether to reward or punish the opponent. To be more precise, note that each automaton state $x_i$ consists of two components, and let $x_i^{\omega_1} \in \{G, B\}$ denote the first component and $x_i^{\omega_2} \in \{G, B\}$ denote the second. The first component $x_i^{\omega_1}$ represents player $i$'s plan about whether to punish the opponent *if the true state were* $\omega_1$. Similarly, the second component $x_i^{\omega_2}$ represents her plan *if the true state*

*were* $\omega_2$. For example, if player $i$'s automaton state is $x_i = GB$, then during the current block, she rewards the opponent at state $\omega_1$ and punishes the opponent at state $\omega_2$. (In other words, we will choose the corresponding block strategy $s_i^{GB}$ so that it yields a high payoff to the opponent conditional on $\omega_1$ but a low payoff conditional on $\omega_2$.) Likewise, If $x_i = BG$, she punishes the opponent at state $\omega_1$ but rewards at state $\omega_2$. If $x_i = GG$, she rewards the opponent at both states. If $x_i = BB$, she punishes the opponent at both states.

After the block, each player $i$ chooses a new automaton state (plan) $\tilde{x}_i = (\tilde{x}_i^{\omega_1}\tilde{x}_i^{\omega_2})$ for the next block. Specifically, for each state $\omega$, the new plan for the state $\omega$ is determined by a *transition rule* $\rho_i^{\omega}(\cdot|x_i^{\omega}, h_i^{T_b}) \in \triangle\{G,B\}$; that is, given the current plan $x_i^{\omega}$ and the current block history $h_i^{T_b}$, player $i$ randomly selects a new plan $\tilde{x}_i^{\omega} \in \{G,B\}$ according to this distribution $\rho_i^{\omega}$. Note that the current plan $x_i^{\tilde{\omega}}$ for state $\tilde{\omega}$ does not directly influence the new plan $\tilde{x}_i^{\omega}$ for state $\omega \neq \tilde{\omega}$.

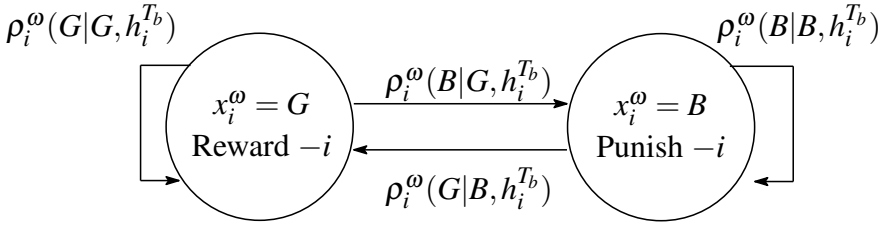


Figure 3: Transition of $x_i^{\omega}$

In what follows, we will carefully choose the block strategies $s_i^{GG}$, $s_i^{GB}$, $s_i^{BG}$, and $s_i^{BB}$ and the transition rules $\rho_i^{\omega_1}$ and $\rho_i^{\omega_2}$ so that the resulting automaton strategy is indeed an equilibrium.

## 4.2 Block Strategy $s_i^{x_i}$

### 4.2.1 Brief Description

Let $T_b = 2T + 1 + T^2 + 4T$, where $T > 0$ is to be specified. As noted, we regard the infinite horizon as a sequence of blocks with length $T_b$. Each block is further divided into four parts: The first $2T$ periods of the block are the *learning round*. The next period is the *summary report round*, and then the next $T^2$ periods are the *main round*. The remaining $4T$ periods are the *detailed report round*. See Figure 4.

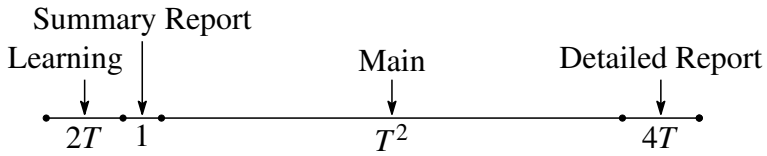

Figure 4: Structure of the block. Time goes from left to right.

As will be explained, we will choose $T$ sufficiently large, so that the main round is much longer than the other rounds. Thus, the average payoff during the block is approximately the payoff during the main round. In other words, the payoffs during the learning round and the two report rounds are almost negligible.

The role of each round is roughly as follows.

**Learning Round:** The first $T$ periods of the learning round are player 1's learning round, in which player 1 collects private signals and makes an inference $\omega(1)$ about the true state $\omega$. The next $T$ periods are player 2's learning round, in which player 2 makes an inference $\omega(2)$ about the state. During the learning round, players play the action profile $a^*$, so Condition 2 ensures that players can indeed distinguish the state statistically. Player $i$'s inference $\omega(i)$ takes one of three values: $\omega_1$, $\omega_2$, or $\emptyset$. Roughly, she chooses $\omega(i) = \omega_1$ if the signal frequency during her learning round is close to the true distribution $\pi_i^{\omega_1}(a^*)$ at $\omega_1$, and $\omega(i) = \omega_2$ if it is close to the true distribution $\pi_i^{\omega_2}(a^*)$ at $\omega_2$. Otherwise, she chooses a "null" inference $\omega(i) = \emptyset$. More details will be given in the next subsubsection. Let $T(i)$ denote the set of the periods included in player $i$'s learning round. That is, $T(1) = \{1, \cdots, T\}$ and $T(2) = \{T+1, \cdots, 2T\}$.

**Summary Report Round:** The next period is the summary report round, in which each player $i$ publicly reports her inference $\omega(i)$ using her action. For simplicity, we assume that each player has at least three actions, so that she can indeed represent $\omega(i) \in \{\omega_1, \omega_2, \emptyset\}$ by one-shot actions.[17] This "communication" allows

[17]This assumption is not essential. If there is a player who has only two actions, we can modify the structure of the block, so that the summary report round consists of two periods and each player represents her inference by a sequence of actions. The rest of the proof remains the same. (When the summary report round consists of two periods, each player can obtain partial information about the opponent's inference $\omega(-i)$ after the first period of the summary report round. But this information does not influence players' incentives, that is, the truthful report of $\omega(i)$ is still a best

players to coordinate their continuation play. Note that $\omega(i)$ is just a summary statistic of player $i$'s observation during the learning round, and hence this round is called "summary report."

**Main Round:** The next $T^2$ periods are the main round, in which players coordinate their play depending on the information revealed in the summary report round. If players report the same state $\omega$ in the summary report round, then players play the block strategy of Hörner and Olszewski (2006) during the main round:

- If both players report the same state $\omega$ in the summary report round, then in the first period of the main round, they "communicate" again and each player $i$ reports her current plan $x_i^\omega \in \{G,B\}$ for this state $\omega$. After that, players choose the action profile $a^{\omega,x^\omega}$ until the main round ends, where $x^\omega = (x_1^\omega x_2^\omega)$ is the reported plan. (Recall that this action profile $a^{\omega,x^\omega}$ is chosen as in Figure 1.) If someone (say player $i$) deviates from this action profile $a^{\omega,x^\omega}$, she will be minimaxed by $\underline{\alpha}^\omega(i)$. That is, players minimax the deviation, assuming that the summary report $\omega$ is the true state.

So if players report the same state $\omega$ in the summary report round, they coordinate their play during the main round and choose an action profile which is consistent with the current plan. By the definition of the action $a^{\omega,x^\omega}$, each player $i$ obtains a payoff higher than $\overline{v}_i^\omega$ if the opponent plans to reward her (i.e., $x_{-i}^\omega = G$), and a payoff lower than $\underline{v}_i^\omega$ if the opponent plans to punish her (i.e., $x_{-i}^\omega = B$).

If players' reports in the summary report round do not coincide, or if someone reports the null inference $\omega(i) = \emptyset$, they adjust their play in the following way:

- If one player reports $\omega$ but the other reports $\emptyset$, then the play during the main round is the same as above. (Intuitively, reporting $\omega(i) = \emptyset$ is treated as an abstention.)

- If both players report $\emptyset$, then the play during the main round is the same as the case in which both players report $\omega_1$.

- If one player reports $\omega_1$ while the other reports $\omega_2$, then each player $i$ reveals $x_i^{\omega(i)}$ in the first period of the main round, and then chooses the minimax

reply. This is so because in our equilibrium, the truthful report of $\omega(i)$ is a best reply, regardless of the opponent's inference $\omega(-i)$.)

action $\underline{\alpha}_i^{\omega(i)}(-i)$, where $\omega(i)$ denotes the state reported by player $i$. That is, each player minimaxes the opponent, assuming that her own summary report is the true state.

**Detailed Report Round:** The remaining $4T$ periods of the block are the detailed report round. Recall that in the summary report round, each player reports only $\omega(i)$, which is a *summary statistic* of her observation during the learning round. Now, in the detailed report round, each player reports her *full history* during the learning round. Specifically, in the first $T$ periods, player 1 reports her observation $(z_1^t)_{t\in T(1)}$ during her own learning round. The assumption $|A_i| \geq |Z_i|$ ensures that players can reveal her signal $z_i$ by choosing one action, so she can indeed report her signal sequence $(z_1^t)_{t\in T(1)}$ using $T$ periods. In the next $T$ periods, player 2 reports her observation $(z_2^t)_{t\in T(2)}$ during her own learning round. After that, player 1 reports her observation $(z_1^t)_{t\in T(2)}$ during the opponent's learning round, and then player 2 reports $(z_2^t)_{t\in T(1)}$. This information (the detailed report) can be used to double-check whether the opponent's summary report earlier was truthful or not, and it influences the choice of the new automaton state $\tilde{x}_i$ for the next block. We will explain more on this later.

For each automaton state $x_i$, let $s_i^{x_i}$ be the block strategy which chooses actions as described above. That is, $s_i^{x_i}$ chooses the action $a_i^*$ and makes the inference $\omega(i)$ in the learning round; reports the summary inference $\omega(i)$ in the summary report round; coordinates the play as above in the main round; and then reports the actual signal sequence $(z_i^t)_{t\in T(i)}$ in the detailed report round. The definition of $s_i^{x_i}$ here is informal, because we have not explained how player $i$ forms $\omega(i)$.

**Remark 1.** Why do we want to have a learning round for each player $i$ separately? A point is that with this structure, player $i$'s inference $\omega(i)$ (the first-order belief in some sense) and her belief about the opponent's inference $\omega(-i)$ (the second-order belief in some sense) are made from different information sources: Player $i$'s first-order belief depends only on her history during her own learning round, while her second-order belief depends only on her history during the *opponent's* learning round. This means that even if player $i$ reports $\omega(i)$ in the summary report round, it does not reveal her second-order belief to the opponent. This property is crucial in order to provide appropriate incentives for the detailed report round.

See the proof of Lemma 3 for more details.

### 4.2.2 Inference Rule

To complete the definition of the block strategy $s_i^{x_i}$, we will explain how each player $i$ forms the inference $\omega(i)$ during her learning round.

Recall that player $i$'s learning round consists of $T$ periods. Let $h_i^T$ denote player $i$'s history during this round, and let $H_i^T$ denote the set of all such histories. Player $i$'s *inference rule* is defined as a mapping $P : H_i^T \to \triangle\{\omega_1, \omega_2, \emptyset\}$. That is, given a private history $h_i^T$, player $i$ (randomly) chooses the inference $\omega(i)$ from the set $\{\omega_1, \omega_2, \emptyset\}$, according to the distribution $P(\cdot|h_i^T)$. It is important that we allow player $i$ to choose $\omega(i)$ randomly; this property is needed in order to prove Lemma 1 below.

Given an inference rule $P$, let $\hat{P}(\cdot|\omega, a^1, \cdots, a^T)$ denote the conditional distribution of $\omega(i)$ induced by $P$ given that the true state is $\omega$ and players play the action sequence $(a^1, \cdots, a^T)$ during player $i$'s learning round. That is,

$$\hat{P}(\cdot|\omega, a^1, \cdots, a^T) = \sum_{h_i^T \in H_i^T} \Pr(h_i^T|\omega, a^1, \cdots, a^T) P(\cdot|h_i^T)$$

where $\Pr(h_i^T|\omega, a^1, \cdots, a^T)$ denotes the probability of $h_i^T$ when the true state is $\omega$ and players play $(a^1, \cdots, a^T)$. Likewise, for each $t \in \{0, \cdots, T-1\}$ and $h^t$, let $\hat{P}(\cdot|\omega, h_{-i}^t, a^{t+1}, \cdots, a^T,)$ be the conditional distribution of $\omega(i)$ given that the true state is $\omega$, the opponent's history up to the $t$th period is $h_{-i}^t = (a^\tau, z_{-i}^\tau)_{\tau=1}^t$, and players play $(a^{t+1}, \cdots, a^T)$ thereafter. Given $h_i^T$, let $f_i(h_i^T) \in \triangle Z_i$ denote player $i$'s signal frequency induced by $h_i^T$. That is, $f_i(h_i^T)[z_i] = \frac{|\{t|z_i^t = z_i\}|}{T}$ for each $z_i$.

The following lemma shows that there is an inference rule $P$ which satisfies some useful properties. The proof is similar to Fong, Gossner, Hörner and Sannikov (2011) and Sugaya (2019), and can be found in Appendix A. Recall that $\varepsilon$ has been fixed so that (1)-(6) hold.

**Lemma 1.** *Suppose that Condition 2 holds. Then there is $\overline{T}$ such that for any $T > \overline{T}$, there is an inference rule $P : H_i^T \to \triangle\{\omega_1, \omega_2, \emptyset\}$ which satisfies the following properties:*

*(i) If players do not deviate from $a^*$, the inference $\omega(i)$ coincides with the true*

*state with high probability: For each* $\omega$*,*

$$\hat{P}(\omega(i)=\omega|\omega,a^*,\cdots,a^*)\geq 1-\exp(-T^{\frac{1}{2}}).$$

*(ii) Regardless of the past history, the opponent's deviation cannot manipulate player i's inference with high probability: For each* $\omega$*,* $t\in\{0,\cdots,T-1\}$*,* $h^t$*,* $(a^\tau)_{\tau=t+1}^T$*, and* $(\tilde{a}^\tau)_{\tau=t+1}^T$ *such that* $a_i^\tau=\tilde{a}_i^\tau=a_i^*$ *for all* $\tau\geq t+1$*,*

$$|\hat{P}(\cdot|\omega,h_{-i}^t,a^{t+1},\cdots,a^T)-\hat{P}(\cdot|\omega,h_{-i}^t,\tilde{a}^{t+1},\cdots,\tilde{a}^T)|\leq\exp(-T^{\frac{1}{2}}).$$

*(iii) Suppose that no one deviates from* $a^*$*. Then player i's inference is* $\omega(i)=\omega$*, only if her signal frequency is close to the true distribution* $\pi_i^\omega(a^*)$ *at* $\omega$*: For all* $h_i^T=(a^t,z_i^t)_{t=1}^T$ *such that* $a^t=a^*$ *for all t and such that* $P(\omega(i)=\omega|h_i^T)>0$*,*

$$|\pi_i^\omega(a^*)-f_i(h_i^T)|<\varepsilon.$$

Clause (i) ensures that state learning is almost perfect. Clause (ii) asserts that state learning is robust to the opponent's deviation *after every history*. To see its precise meaning, suppose that the first $t$ periods of player $i$'s learning round are over, and the opponent's history during these periods was $h_{-i}^t$. The opponent can deviate in the remaining periods, but clause (ii) implies that it cannot influence player $i$'s inference much. Note that both clauses (i) and (ii) are natural consequences of Condition 2, which guarantees that player $i$ can learn the true state even if someone else unilaterally deviates. Clause (iii) implies that player $i$ makes the inference $\omega(i)=\omega$ only when her signal frequency is close to the true distribution $\pi_i^\omega(a^*)$ at state $\omega$. So if player $i$'s signal frequency is not close to $\pi_i^{\omega_1}(a^*)$ or $\pi_i^{\omega_2}(a^*)$, her inference must be $\omega(i)=\emptyset$. (On the other hand, as can be seen from the proof of the lemma, player $i$ mixes $\omega(i)=\omega$ and $\omega(i)=\emptyset$ if her signal frequency is close to $\pi_i^\omega(a^*)$. See Figure 5.)

Clause (iii) is useful when we derive a bound on player $i$'s higher-order belief (i.e., player $i$'s belief about the opponent's signal frequency $f_{-i}$, which is informative about player $i$'s inference $\omega(i)$ about the state). Let $\Pr(f_{-i}|\omega,a^*,\cdots,a^*,f_i)$ denote the probability of the opponent's signal frequency being $f_{-i}$, given that the true state is $\omega$, players play $a^*$ for $T$ periods, and player $i$'s signal frequency during these periods is $f_i$. Then we have the following lemma:

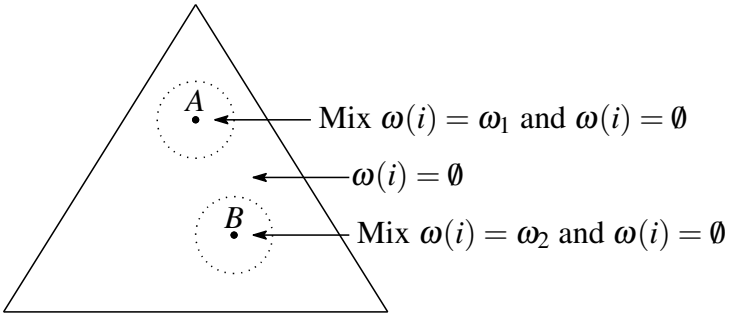


Figure 5: The triangle is the set of signal frequencies, $\triangle Z_i$. The point $A$ denotes $\pi_i^{\omega_1}(a^G)$, while $B$ denotes $\pi_i^{\omega_2}(a^G)$.

**Lemma 2.** *Suppose that Condition 3 holds. Then there is $\overline{T}$ such that for any $T > \overline{T}$, $\omega$, $\tilde{\omega} \neq \omega$, and $h_i^T$ such that $|f_i(h_i^T) - \pi_i^{\tilde{\omega}}(a^*)| < \varepsilon$, we have*

$$\sum_{f_{-i}:|f_{-i}-\pi_{-i}^{\omega}(a^*)|<\varepsilon} \Pr(f_{-i}|\omega, a^*, \cdots, a^*, f_i(h_i^T)) < \exp(-T^{\frac{1}{2}}).$$

Roughly, this lemma implies that if player $i$ has the inference $\omega(i) = \tilde{\omega}$ (which is unusual conditional on the state $\omega \neq \tilde{\omega}$), then she believes that conditional on the state $\omega$, the opponent's observation is also unusual and not close to the ex-ante distribution $\pi_{-i}^{\omega}(a^*)$. To see this, suppose that player $i$'s inference is $\omega(i) = \tilde{\omega}$. Then from Lemma 1(iii), we must have $|f_i(h_i^T) - \pi_i^{\tilde{\omega}}(a^*)| < \varepsilon$. Then from the lemma above, player $i$ believes that the opponent's observation is not close to the ex-ante distribution. As will be explained, this result plays a crucial role in order to induce the truthful summary report.

*Proof.* Pick $h_i^T$ such that

$$|\pi_i^{\tilde{\omega}}(a^*) - f_i(h_i^T)| < \varepsilon.$$

Using (6), we have

$$|C_i^{\omega}(a^*)\pi_i^{\tilde{\omega}}(a^*) - C_i^{\omega}(a^*)f_i(h_i^T)| \leq \sqrt{\varepsilon}.$$

Combining this with (5),

$$|C_i^{\omega}(a^*)f_i(h_i^T) - \pi_{-i}^{\omega}(a^*)| \geq \sqrt{\varepsilon}.$$

Accordingly, in order to have $\left|\pi_{-i}^{\omega}(a^*) - f_{-i}\right| < \varepsilon$, the distance between $C_i^{\omega}(a^*) f_i(h_i^T)$ and $f_{-i}$ must be at least $\sqrt{\varepsilon} - \varepsilon$. However, Hoeffding's inequality implies that the probability of such an event is less than $\exp(-T^{\frac{1}{2}})$ for sufficiently large $T$. *Q.E.D.*

**Remark 2.** Allowing the null inference $\omega(i) = \emptyset$ is important. As noted in the introduction, given player $i$'s observation $f_i$, different states induce different beliefs about the opponent's observation $f_{-i}$. In particular, at the point $f_i = C$ in Figure 6, player $i$ has "conflicting beliefs" at different states; she believes that (i) conditional on the state $\omega_1$, the opponent's signal frequency $f_{-i}$ is typical of the state $\omega_1$ so that the opponent believes that player $i$'s inference is $\omega(i) = \omega_1$, but (ii) conditional on the state $\omega_2$, the opponent's signal frequency $f_{-i}$ is typical of the state $\omega_2$ so that the opponent believes that player $i$'s inference is $\omega(i) = \omega_2$. In this case, reporting $\omega(i) = \omega_1$ cannot be a best reply at the state $\omega_2$, because it contradicts with the opponent's expectation illustrated in (ii) above, and triggers a state-contingent punishment. (See the proof of Lemma 3 for the formal description of the punishment mechanism.) At the same time, reporting $\omega(i) = \omega_2$ cannot be a best reply at the state $\omega_1$, as it contradicts with the opponent's expectation described in (i). So reporting $\omega(i) = \omega_1$ and $\omega(i) = \omega_2$ cannot be ex-post incentive compatible when player $i$ has such conflicting beliefs. Instead, in our equilibrium, she makes the null inference $\omega(i) = \emptyset$ and reports it truthfully when she has such conflicting beliefs.

## 4.3 Transition Rule $\rho_i$ and Equilibrium Conditions

We have defined the block strategy $s_i^{x_i}$: Players learn the state in the learning round, report the summary inference $\omega(i)$ in the summary report round, coordinate the play in the main round, and then report the full information in the detailed report round. What remains is to find transition rules $\rho_i^{\omega_1}$ and $\rho_i^{\omega_2}$ so that the resulting automaton strategy is an equilibrium.

Formally, we choose the transition rules so that both the *promise-keeping condition* and the *incentive-compatibility condition* hold. The promise-keeping condition requires that the target payoffs be exactly achieved *state by state*; for example, if the opponent's current automaton state is $x_{-i} = GB$, player $i$'s payoff in

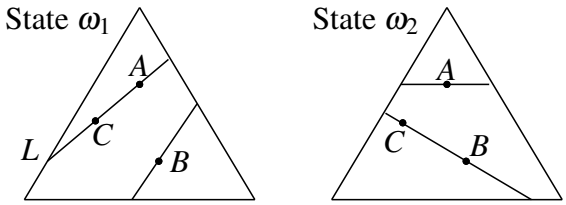


Figure 6: Each line in the left triangle is the set of signal frequencies $f_i$ which give the same expectation about the opponent's signal frequency at the state $\omega_1$. That is, $C_i^{\omega_1}(a^*)f_i = C_i^{\omega_1}(a^*)\tilde{f}_i$ for any $f_i$ and $\tilde{f}_i$ on the same line. At the point $f_i = A$, player $i$ believes that the opponent's observation is typical of the state $\omega_1$, in that $C_i^{\omega_1}(a^*)f_i = \pi_{-i}^{\omega_1}(a^*)$; so the same is true at the point $f_i = C$. Likewise, each line in the right triangle is the set of $f_i$ which induce the same expectation at the state $\omega_2$. At the point $f_i = B$, player $i$ believes that the opponent's observation is typical of the state $\omega_2$, and the same is true at the point $f_i = C$.

the continuation game must be $\overline{v}_i^{\omega_1}$ conditional on the state $\omega_1$ (since player $i$ is rewarded at $\omega_1$) and $\underline{v}_i^{\omega_2}$ conditional on the state $\omega_2$ (since player $i$ is punished at $\omega_2$). Formally, it requires

$$\overline{v}_i^{\omega} = (1-\delta^{T_b})\sum_{t=1}^{T_b}\delta^{t-1}E[g_i^{\omega}(a^t)|\omega,s^x] + \delta^{T_b}\left\{\overline{v}_i^{\omega} - E[\rho_{-i}^{\omega}(B|G,h_{-i}^{T_b})|\omega,s^x](\overline{v}_i^{\omega} - \underline{v}_i^{\omega})\right\} \tag{7}$$

for each $\omega$, $i$, and $x = (x_1,x_2)$ with $x_{-i}^{\omega} = G$, and

$$\underline{v}_i^{\omega} = (1-\delta^{T_b})\sum_{t=1}^{T_b}\delta^{t-1}E[g_i^{\omega}(a^t)|\omega,s^x] + \delta^{T_b}\left\{\underline{v}_i^{\omega} + E[\rho_{-i}^{\omega}(G|B,h_{-i}^{T_b})|\omega,s^x](\overline{v}_i^{\omega} - \underline{v}_i^{\omega})\right\} \tag{8}$$

for each $\omega$, $i$, and $x$ with $x_{-i}^{\omega} = B$. (7) asserts that if $x_{-i}^{\omega} = G$ so that the opponent plans to reward player $i$ for the state $\omega$, then player $i$'s payoff in the continuation game is exactly $\overline{v}_i^{\omega}$ conditional on the state $\omega$. Indeed, the first term in the right-hand side is player $i$'s payoff in the current block, and the second term is her continuation payoff. (The term $E[\rho_{-i}^{\omega}(B|G,h_{-i}^{T_b})|\omega,s^x]$ is the probability that the opponent switches to the punishment plan $x_{-i}^{\omega} = B$ after the block, in which case player $i$'s continuation payoff goes down from $\overline{v}_i^{\omega}$ to $\underline{v}_i^{\omega}$.) Similarly, (8) asserts

that if the opponent plans to punish player $i$ for the state $\omega$, player $i$'s payoff in the continuation game is $\underline{v}_i^\omega$ conditional on the state $\omega$. The above conditions imply that player $i$'s payoff is solely determined by the opponent's plan $x_{-i}$, and is independent of her own plan $x_i$. (While her current block payoff depends on the plan $x_i$, this effect is offset by the continuation payoffs, so the total payoff is indeed independent of $x_i$.) So in each block, player $i$ is indifferent over the four strategies, $s_i^{GG}$, $s_i^{GB}$, $s_i^{BG}$, and $s_i^{BB}$. This in turn implies that randomizing the automaton state $x_i$ at the beginning of the block is indeed a best reply for player $i$.

The incentive-compatibility condition requires that deviating to any other block strategy $s_i^{T_b} \neq s_i^{x_i}$ be not profitable, in each period of the block. That is,

$$\begin{aligned}
&(1-\delta^{T_b-t}) \sum_{\tau=t+1}^{T_b} \delta^{\tau-1} \left( E[g_i^\omega(a^\tau)|\omega, s_i^{T_b}, s_{-i}^{x_{-i}}, h_i^t] - E[g_i^\omega(a^\tau)|\omega, s^x, h_i^t] \right) \\
&\leq \delta^{T_b-t} \left( E[\rho_{-i}^\omega(B|G, h_{-i}^{T_b})|\omega, s_i^{T_b}, s_{-i}^{x_{-i}}, h_i^t] - E[\rho_{-i}^\omega(B|G, h_{-i}^{T_b})|\omega, s^x, h_i^t] \right) (\overline{v}_i^\omega - \underline{v}_i^\omega)
\end{aligned} \tag{9}$$

for each $\omega$, $i$, $s_i^{T_b}$, $t$, $h_i^t$, and $x$ with $x_{-i}^\omega = G$, and

$$\begin{aligned}
&(1-\delta^{T_b-t}) \sum_{\tau=t+1}^{T_b} \delta^{\tau-1} \left( E[g_i^\omega(a^\tau)|\omega, s_i^{T_b}, s_{-i}^{x_{-i}}, h_i^t] - E[g_i^\omega(a^\tau)|\omega, s^x, h_i^t] \right) \\
&\leq \delta^{T_b-t} \left( E[\rho_{-i}^\omega(B|B, h_{-i}^{T_b})|\omega, s_i^{T_b}, s_{-i}^{x_{-i}}, h_i^t] - E[\rho_{-i}^\omega(B|B, h_{-i}^{T_b})|\omega, s^x, h_i^t] \right) (\overline{v}_i^\omega - \underline{v}_i^\omega)
\end{aligned} \tag{10}$$

for each $\omega$, $i$, $s_i^{T_b}$, $t$, $h_i^t$, and $x$ with $x_{-i}^\omega = B$. Here the left-hand side measures how much the block payoff increases by deviating in period $t+1$ of the block, and the right-hand side measures how much it decreases the continuation payoff after the block. So these inequalities imply that in any period of the block, deviating from the prescribed strategy $s_i^{x_i}$ is not profitable, regardless of the true state. Accordingly, the resulting automaton strategy is an ex-post equilibrium.

## 4.4 Complete-Information Transfer Game

In what follows, we will explain how to find the transition rules which satisfy the above conditions (7) through (10). This completes our proof, because the resulting automaton strategy is indeed an equilibrium and any payoff in the set

$\times_{\omega\in\Omega}\times_{i\in I}[\underline{v}_i^\omega,\overline{v}_i^\omega]$ can be achieved by randomizing the initial automaton state. In particular, the payoff $v$ is exactly achievable.

It turns out that finding such transition rules is equivalent to finding appropriate "transfer rules." This is so because continuation payoffs after the block play a role like that of transfers in the mechanism design. A similar idea appears in various past work, e.g., Fudenberg and Levine (1994).

As such, we will focus on the following *complete-information transfer game*: Consider a repeated game with $T_b$ periods. Assume complete information, so that a state $\omega$ is given and common knowledge. After the game, player $i$ receives a transfer according to some transfer rule $U_i^\omega : H_{-i}^{T_b} \to \boldsymbol{R}$, so player $i$'s (unnormalized) payoff in this game is

$$\sum_{t=1}^{T_b} \delta^{t-1} g_i^\omega(a^t) + \delta^{T_b} U_i^\omega(h_{-i}^{T_b}).$$

Let $G_i^\omega(s^{T_b}, U_i^\omega)$ denote player $i$'s expected payoff in this game, when players play $s^{T_b}$. Also, for each history $h_i^t$ with $t \leq T_b$, let $G_i^\omega(s^{T_b}, U_i^\omega, h_i^t)$ denote player $i$'s payoff in the continuation game after history $h_i^t$.

A few remarks are in order. First, this is the complete-information game, so the state $\omega$ is given and common knowledge. The analysis of this complete-information game is useful, because our goal is to construct an equilibrium which satisfies the ex-post incentive compatibility conditions (9) and (10); these conditions require that player $i$'s deviation be not profitable even when the state $\omega$ is publicly revealed at the beginning of the game.

Second, the transfer $U_i^\omega$ is state-specific, that is, we use different transfer rules $U_i^\omega$ for different states $\omega$. This captures the idea that punishments are state-specific in our equilibrium in the infinite-horizon game. Specifically, once the block is over, the opponent chooses a state-specific punishment plan $x_{-i} = (x_{-i}^{\omega_1}, x_{-i}^{\omega_2})$ for the continuation game, and player $i$'s continuation payoff conditional on $\omega$ is solely determined by the punishment plan $x_{-i}^\omega$ for the state $\omega$ (see (7) and (8)). Since the opponent chooses these plans $x_{-i}^{\omega_1}$ and $x_{-i}^{\omega_2}$ independently, player $i$'s continuation payoffs for different states take quite different values. Hence the transfer rule $U_i^\omega$ should depend on $\omega$.

Third, the amount of the transfer depends on the opponent's history $h_{-i}^{T_b}$, but not on player $i$'s history $h_i^{T_b}$. Again this comes from the fact that player $i$'s con-

tinuation payoff is determined by the opponent's plan $x_{-i}$, which is influenced by the opponent's history $h_{-i}^{T_b}$ but not by $h_i^{T_b}$.

Our goal in this subsection is to prove the following two lemmas. The first lemma is:

**Lemma 3.** *There is $\overline{T}$ such that for any $T > \overline{T}$, there is $\overline{\delta} \in (0,1)$ such that for each $\delta \in (\overline{\delta}, 1)$, i, and $\omega$, there is a transfer rule $U_i^{\omega,G} : H_{-i}^{T_b} \to \boldsymbol{R}$ which satisfies the following properties.*

*(i)* $\frac{1-\delta}{1-\delta^{T_b}} G_i^{\omega}(s^x, U_i^{\omega,G}) = \overline{v}_i^{\omega}$ *for all x with* $x_{-i}^{\omega} = G$.

*(ii)* $G_i^{\omega}(s_i^{T_b}, s_{-i}^{x_{-i}}, U_i^{\omega,G}, h_i^t) \leq G_i^{\omega}(s^x, U_i^{\omega,G}, h_i^t)$ *for all* $s_i^{T_b}$, $h_i^t$, *and x with* $x_{-i}^{\omega} = G$.

*(iii)* $-(\overline{v}_i^{\omega} - \underline{v}_i^{\omega}) \leq (1-\delta) U_i^{\omega,G}(h_{-i}^{T_b}) \leq 0$ *for all* $h_{-i}^{T_b}$.

To interpret this lemma, consider the complete-information game with the state $\omega_1$. Suppose that the opponent plays the block strategy $s_{-i}^{GG}$ or $s_{-i}^{GB}$. That is, the opponent plans to reward player $i$ for the state $\omega_1$. Clauses (i) implies that if the transfer rule $U_i^{\omega_1,G}$ is appropriately chosen, then player $i$ becomes indifferent over the prescribed block strategies, $s_i^{GG}$, $s_i^{GB}$, $s_i^{BG}$, and $s_i^{BB}$, and these strategies yield the target payoff $\overline{v}_i^{\omega_1}$ exactly. Clause (ii) requires that with this transfer rule $U_i^{\omega_1,G}$, any deviation from the prescribed strategies should not be profitable. Clause (iii) requires that this transfer be non-positive (and bounded), that is, the transfer takes a form of welfare destruction. This last condition comes from the fact that player $i$'s continuation payoff at state $\omega$, which is represented by the second term in the right-hand side of (7) and (9), is in the interval $[\underline{v}_i^{\omega}, \overline{v}_i^{\omega}]$ and hence below the target payoff $\overline{v}_i^{\omega}$.

As noted in the introduction, a key step in the proof is to construct a transfer rule which induces the truthful summary report, while keeping the welfare destruction small. To do so, we consider a transfer rule with which the opponent makes player $i$ indifferent over reports in the summary report round after *some* (but not all) histories. In the next subsection, we will provide a sketch of the proof. The formal proof can be found in Appendix A. (In the complete-information transfer game, the state $\omega$ is common knowledge, but each player $i$ still makes an inference $\omega(i)$ and reports it, just as specified in the description of $s_i^{x_i}$. In particular, when the inference is $\omega(i) = \tilde{\omega}$, player $i$ reports it, even though she knows that it does

not coincide with the true state $\omega$. We need to find a transfer rule under which this report is indeed incentive compatible.)

Once we have this lemma, we can construct a transition rule $\rho_{-i}^{\omega}(\cdot|G,h_{-i}^{T_b})$ which satisfies the desired properties (7) and (9), by setting

$$\rho_{-i}^{\omega}(B|G,h_{-i}^{T_b}) = -\frac{(1-\delta)U_i^{\omega,G}(h_{-i}^{T_b})}{\overline{v}_i^{\omega} - \underline{v}_i^{\omega}}.$$

for each $h_{-i}^{T_b}$. Indeed, simple algebra shows that Lemma 3(i) implies (7), and Lemma 3(ii) implies (9). Lemma 3(iii) ensures that $\rho_{-i}^{\omega}(B|G,h_{-i}^{T_b})$ defined here is indeed a probability.

The second lemma is a counterpart to the above lemma. It considers the case in which the opponent plans to punish player $i$ (i.e., $x_{-i}^{\omega} = B$).

**Lemma 4.** *There is $\overline{T}$ such that for any $T > \overline{T}$, there is $\overline{\delta} \in (0,1)$ such that for each $\delta \in (\overline{\delta},1)$, $i$, and $\omega$, there is a transfer rule $U_i^{\omega,B} : H_{-i}^{T_b} \to \boldsymbol{R}$ which satisfies the following properties.*

*(i)* $\frac{1-\delta}{1-\delta^{T_b}} G_i^{\omega}(s^x, U_i^{\omega,B}) = \underline{v}_i^{\omega}$ *for all $x$ with $x_{-i}^{\omega} = B$.*

*(ii)* $G_i^{\omega}(s_i^{T_b}, s_{-i}^{x_{-i}}, U_i^{\omega,B}, h_i^t) \leq G_i^{\omega}(s^x, U_i^{\omega,B}, h_i^t)$ *for all $s_i^{T_b}$, $h_i^t$, and $x$ with $x_{-i}^{\omega} = B$.*

*(iii)* $0 \leq (1-\delta)U_i^{\omega,B}(h_{-i}^{T_b}) \leq \overline{v}_i^{\omega} - \underline{v}_i^{\omega}$ *for all $h_{-i}^{T_b}$.*

The last constraint requires the transfer to be non-negative. This comes from the fact that player $i$'s continuation payoff at state $\omega$ is chosen from the interval $[\underline{v}_i^{\omega}, \overline{v}_i^{\omega}]$ and always above the target payoff $\underline{v}_i^{\omega}$.

It turns out that the proof of this lemma is much simpler than that of the previous lemma. In particular, we can construct a transfer rule with which the opponent makes player $i$ indifferent over all reports in the summary report round after *every* history (just as in belief-free equilibria of Ely, Hörner, and Olszewski (2005)). This is analogous to Hörner and Olszewski (2006); their transfer rule for the punishment state makes a player indifferent over all actions each period of the block, while the transfer rule for the reward state has a much more complicated form. See Appendix A for the formal proof.

Again, once we have this lemma, we can construct a transition rule $\rho_{-i}^{\omega}(\cdot|G,h_{-i}^{T_b})$ which satisfies the desired properties (8) and (10), by setting

$$\rho_{-i}^{\omega}(G|B,h_{-i}^{T_b}) = \frac{(1-\delta)U_i^{\omega,B}(h_{-i}^{T_b})}{\overline{v}_i^{\omega} - \underline{v}_i^{\omega}}$$

So Proposition 1 immediately follows, once we prove the above two lemmas.

## 4.5 Proof Sketch of Lemma 3

As noted earlier, a key step in the proof is to show that the opponent can deter a misreport of the summary inference $\omega(i)$ using a transfer, subject to the constraint that the expected welfare destruction is small. In what follows, we will explain how to construct such a transfer rule. For simplicity, we will assume that players do not deviate from the prescribed strategy $s^x$ during the learning round and the main round. That is, we will focus on incentives in the two report rounds.

To begin with, it is useful to point out that player $i$'s deviation in the summary report round can be easily deterred by making her indifferent over all summary reports, but it requires a huge welfare destruction. Let $\overline{g}_i^{\omega} = \max_{a \in A} |g_i^{\omega}(a)|$. Pick a constant $C$, and for each block history $h_{-i}^{T_b}$, choose the transfer $\hat{U}_i^{\omega,G}(h_{-i}^{T_b})$ so that

$$\frac{1-\delta}{1-\delta^{T_b}} \left[ \sum_{t=1}^{T_b} \delta^{t-1} g_i^{\omega}(a^t) + \delta^{T_b} \hat{U}_i^{\omega,G}(h_{-i}^{T_b}) \right] = C. \tag{11}$$

That is, we choose the transfer so that player $i$'s total payoff is exactly $C$, regardless of the play during the block. Then obviously player $i$ is indifferent over all actions in each period of the block, so the truthful summary report is a best reply. Also, if we choose a small $C$ (say, $C = -2\overline{g}_i^{\omega}$), we can ensure that the transfer $\hat{U}_i^{\omega,G}(h_{-i}^{T_b})$ is negative for each $h_{-i}^{T_b}$ so that clause (iii) of the lemma holds. (From (11), the transfer $\hat{U}_i^{\omega,G}(h_{-i}^{T_b})$ is negative if the constant $C$ is less than the average block payoff, $\frac{1-\delta}{1-\delta^{T_b}} \sum_{t=1}^{T_b} \delta^{t-1} g_i^{\omega}(a^t)$.)

Unfortunately, this transfer rule $\hat{U}_i$ does not satisfy clause (i). Indeed, player $i$'s payoff in this transfer game is $C = -2\overline{g}_i^{\omega}$, which is much lower than the target payoff $\overline{v}_i^{\omega}$. This shows that making player $i$ indifferent requires a huge welfare destruction.

Intuitively, this inefficiency result can be understood as follows. Consider the infinite-horizon game, and suppose that the true state is $\omega_1$. Suppose that player $i$ is indifferent over all summary reports in each block. Then her equilibrium payoff must be equal to her payoff when she reports $\omega(i) = \omega_2$ in every block. But this payoff must be much lower than the target payoff $\overline{v}_i^{\omega}$ in general, because players never agree that the true state be $\omega_1$ and they always choose inefficient actions.

In what follows, we will show that by modifying the transfer rule above, the expected welfare destruction can be significantly reduced, without affecting player $i$'s incentive. We do so in two steps. As a first step, we will construct a transfer rule which "approximately" satisfies the desired properties; i.e., we will construct a transfer rule such that the expected welfare destruction is small and the truthful summary report is an approximate best reply (but not an exact best reply) for player $i$. As will be seen, in this transfer rule, the opponent makes player $i$ indifferent at some histories, but not in other cases; this helps to reduce the expected welfare destruction, without affecting player $i$'s incentives by much. Then as a second step, we will modify the transfer rule further so that the truthful summary report is an exact best reply for player $i$. In this second step, communication in the detailed report round plays a central role.

### 4.5.1 Step 1: Approximate Incentive Compatibility

In this step, we will construct a transfer rule such that the expected welfare destruction is small but yet the truthful summary report is an approximate best reply for player $i$. We will first describe how to choose the transfer rule, and then provide its interpretation.

- If the opponent could not make the correct inference (i.e., $\omega(-i) \neq \omega$), then choose the transfer $\tilde{U}_i^{\omega,G}(h_{-i}^{T_b})$ as in (11). This makes player $i$ indifferent over all reports in the summary report round.

- If the opponent's signal frequency $f_{-i}$ during player $i$'s learning round is not typical of $\omega$ (i.e., $|f_{-i} - \pi_{-i}^{\omega}(a^G)| > \varepsilon$), then choose the transfer $\tilde{U}_i^{\omega,G}(h_{-i}^{T_b})$ as in (11). Again, this makes player $i$ indifferent over all reports in the summary report round.

- If the opponent's inference is correct ($\omega(-i) = \omega$) and if the opponent's signal frequency $f_{-i}$ is typical of $\omega$ ($|f_{-i} - \pi_{-i}^{\omega}(a^G)| < \varepsilon$), then

  - If player $i$ reports the wrong inference $\omega(i) = \tilde{\omega}$, choose the transfer $\tilde{U}_i^{\omega,G}(h_{-i}^{T_b})$ as in (11).
  - If player $i$ reports $\omega(i) = \omega$ or $\omega(i) = \emptyset$, choose the transfer $\tilde{U}_i^{\omega,G}(h_{-i}^{T_b})$

so that

$$\frac{1-\delta}{1-\delta^{T_b}}\left[\sum_{t=1}^{T_b}\delta^{t-1}g_i^{\omega}(a^t)+\delta^{T_b}\tilde{U}_i^{\omega,G}(h_{-i}^{T_b})\right]=\overline{v}_i^{\omega}. \tag{12}$$

That is, we set the transfer so that player $i$'s total payoff is exactly $\overline{v}_i^{\omega}$. This transfer $\tilde{U}_i^{\omega,G}(h_{-i}^{T_b})$ is still negative and satisfies clause (iii) of the lemma, because in this case, players play $a^{\omega,x^{\omega}}$ with $x_{-i}^{\omega}=G$ during the main round, so that the average block payoff $\frac{1-\delta}{1-\delta^{T_b}}\sum_{t=1}^{T_b}\delta^{t-1}g_i^{\omega}(a^t)$ is greater than $\overline{v}_i^{\omega}$.

The first two bullet points consider the case in which the opponent's observation is "irregular." Indeed, in the complete-information game with the state $\omega$, the probability of the opponent not having the correct inference is close to zero (Lemma 1(i)), and the probability of the signal frequency $f_{-i}$ being not typical of $\omega$ is close to zero (the law of large numbers). After such irregular observations, the opponent makes player $i$ indifferent, using the huge welfare destruction (11).

The third bullet point considers the case in which the opponent's observation is "regular." In this case, (given that the opponent's signal frequency $f_{-i}$ is typical of $\omega$) the opponent believes that player $i$'s signal frequency is also typical of $\omega$, and hence the opponent believes that player $i$'s inference is $\omega(i)=\omega$ or $\omega(i)=\emptyset$. (See Figure 5.) So the opponent punishes player $i$ when her summary report is not consistent with this belief; that is, player $i$ receives the huge negative transfer (11) if she reports the wrong inference $\omega(i)=\tilde{\omega}$. Otherwise, the opponent sets the transfer as in (12), so that player $i$ enjoys a high payoff of $\overline{v}_i^{\omega}$.

The following table summarizes the discussions so far, and describes player $i$'s best reply when she knows the opponent's inference $\omega(-i)$ and signal inference $f_{-i}$.

| | If $\lvert f_{-i}-\pi_{-i}^{\omega}(a^*)\rvert<\varepsilon$ | If $\lvert f_{-i}-\pi_{-i}^{\omega}(a^*)\rvert\geq\varepsilon$ |
|---|---|---|
| If $\omega(-i)=\omega$ | Report $\omega(i)=\omega$ or $\omega(i)=\emptyset$ | All reports are indifferent |
| If $\omega(-i)\neq\omega$ | All reports are indifferent | All reports are indifferent |

Table 1: Player $i$'s best reply in the summary report round, given the state $\omega$.

A point of the transfer rule above is that the huge welfare destruction (11) occurs only when the opponent's observation is irregular, or player $i$'s summary

report is irregular (i.e., $\omega(i) = \tilde{\omega}$). In the complete-information game with the state $\omega$, these events do not occur almost surely, and hence the expected welfare destruction is small. Indeed, player $i$'s expected payoff in the transfer game is approximately $\bar{v}_i^{\omega}$, because on the equilibrium path, the transfer (12) will be used almost surely. Hence the above transfer rule approximately satisfies clause (i) of the lemma.

At the same time, with the transfer rule above, the truthful summary report is an approximate best reply for player $i$. To see this, suppose, hypothetically, that player $i$ knows the opponent's inference $\omega(-i)$ before it is reported in the summary report round. The following lemma shows that the truthful summary report is (at least) an approximate best reply, regardless of $\omega(-i)$. This result implies that the truthful summary report is an approximate best reply, even if player $i$ does not know $\omega(-i)$. A key in the proof is that when player $i$'s summary inference is $\omega(i) = \tilde{\omega}$ (which is not typical in the complete-information game with the state $\omega$), she believes that the opponent's observation $f_{-i}$ is not typical of $\omega$, in which case the opponent makes her indifferent over all summary reports using the transfer rule (11). This property ensures that player $i$ is almost indifferent over all summary reports, and hence the truthful report of $\omega(i) = \tilde{\omega}$ is an approximate best reply. Given player $i$'s signal frequency $f_i \in \triangle Z_i$ during her own learning round, let

$$p_i^{\omega}(f_i) = \sum_{f_{-i}:|\pi_{-i}^{\omega}(a^*) - f_{-i}|<\varepsilon} \Pr(f_{-i}|\omega, a^*, \cdots, a^*, f_i),$$

that is, $p_i^{\omega}(f_i)$ denotes the conditional probability of the opponent's signal frequency $f_{-i}$ being close to the ex-ante distribution $\pi_{-i}^{\omega}(a^*)$.

**Lemma 5.** *Suppose that no one has deviated from $a^*$ during the learning round. Suppose that player $i$ knows the opponent's inference $\omega(-i)$ before it is reported in the summary report round. If $\omega(-i) \neq \omega$, then player $i$ is indifferent over all actions in the summary report round, and hence the truthful summary report is a best reply for player $i$. If $\omega(-i) = \omega$, then the following properties hold;*

- *If player $i$'s inference is $\omega(i) = \omega$, the truthful summary report is a best reply.*
- *If player $i$'s inference is $\omega(i) = \emptyset$, the truthful summary report is a best reply.*

- *If player $i$'s inference is $\omega(i) = \tilde{\omega} \neq \omega$, the truthful summary report is not an exact best reply: A one-shot deviation by reporting $\omega(i) = \omega$ or $\omega(i) = \emptyset$ improves her payoff by $(\overline{v}_i^{\omega} + 2\overline{g}_i)p_i^{\omega}(f_i)$, where $f_i$ is player $i$'s signal frequency during her own learning round. However, we have $p_i^{\omega}(f_i) < \exp(-T^{\frac{1}{2}})$, so the truthful summary report is an approximate best reply when $T$ is large.*

*Proof.* From the last row of Table 1, it is clear that player $i$ is indifferent over all actions when $\omega(-i) \neq \omega$. So we will focus on the case in which $\omega(-i) = \omega$.

*Case 1: Player $i$'s inference is $\omega(i) = \omega$.* From Table 1, reporting $\omega(i) = \omega$ is a best reply regardless of $f_{-i}$. Hence, the truthful report of $\omega(i) = \omega$ is an exact best reply, regardless of player $i$'s belief about $f_{-i}$.

*Case 2: Player $i$'s inference is $\omega(i) = \emptyset$.* For the same reason, reporting $\omega(i) = \emptyset$ truthfully is a best reply for player $i$ regardless of her belief.

*Case 3: Player $i$'s inference is $\omega(i) = \tilde{\omega} \neq \omega$.* Note that player $i$ believes that $|f_{-i} - \pi_{-i}^{\omega}(a^*)| \geq \varepsilon$ with probability $1 - p_i^{\omega}(f_i)$, and $|f_{-i} - \pi_{-i}^{\omega}(a^*)| < \varepsilon$ with probability $p_i^{\omega}(f_i)$. From Table 1, player $i$ is indifferent over all summary reports in the former case. However, in the latter case, the truthful summary report is not a best reply; the truthful report of $\omega(i) = \tilde{\omega}$ leads to the huge negative transfer (11) and yields a payoff of $-2\overline{g}_i^{\omega}$, while reporting $\omega(i) = \omega$ or $\omega(i) = \emptyset$ leads to the transfer (12) and yields a payoff of $\overline{v}_i^{\omega}$. So the expected gain by reporting $\omega(i) = \omega$ or $\omega(i) = \emptyset$ is indeed $(\overline{v}_i^{\omega} + 2\overline{g}_i)p_i^{\omega}(f_i)$.

Now, recall that from Lemmas 1(iii) and 2, whenever player $i$'s inference is $\omega(i) = \tilde{\omega}$, we have $p_i^{\omega}(f_i) < \exp(-T^{\frac{1}{2}})$. Hence the expected gain above converges to zero as $T \to \infty$. *Q.E.D.*

A few comments are in order. First, under the transfer rule $U_i^{\omega}$ above, reporting the null inference $\omega(i) = \emptyset$ is "executed" in the sense that it always yields the same payoff as the one by reporting the correct inference $\omega(i) = \omega$, and hence always a best reply in the complete-information game with the state $\omega$. Since we choose such a transfer rule $U_i^{\omega}$ for each state $\omega$, reporting the null inference $\omega(i) = \emptyset$ is a best reply regardless of the state $\omega$, and of the opponent's inference $\omega(-i)$, and of the opponent's signal frequency $f_{-i}$. This property is useful to solve the problem raised in Remark 2, because even if player $i$ has conflicting beliefs about the opponent's beliefs at different states (recall the point $C$ in Figure 6 in

the introduction), reporting the null inference $\omega(i) = \emptyset$ is a best reply for player $i$ at both states.

Second, for the above argument to work, it is crucial that player $i$'s inference rule is chosen in such a way that the set of player $i$'s observations which induce the inference $\omega(i) = \omega$ is isolated with the one which induce the inference $\omega(i) = \tilde{\omega}$. That is, the two circles in Figure 5 are disjoint, and there is no "knife-edge" case in which player $i$'s inference switches from $\omega(i) = \omega_1$ to $\omega(i) = \omega_2$. This property, together with the correlated learning condition (Condition 3), ensures that the opponent can almost perfectly distinguish whether player $i$'s inference is $\omega(i) = \omega$ or $\omega(i) = \tilde{\omega}$. Indeed, conditional on the state $\omega$, the opponent's signal frequency $f_{-i}$ is typical of $\omega$ almost surely given that player $i$ has the correct inference $\omega(i) = \omega$, while $f_{-i}$ is not typical of $\omega$ almost surely given that player $i$ has the wrong inference $\omega(i) = \tilde{\omega}$. So if player $i$ deviates by reporting $\omega(i) = \omega$ when the true inference is $\omega(i) = \tilde{\omega}$, the opponent can detect this misreport almost surely. This property is useful in order to deter player $i$'s misreport, while maintaining the expected welfare destruction small.

### 4.5.2 Step 2: Exact Incentive Compatibility

The transfer rule $\tilde{U}_i^{\omega,G}$ in the previous step does not ensure that the truthful summary report be a best reply. Specifically, when player $i$ has the wrong inference $\omega(i) = \tilde{\omega}$, she can improve her payoff by misreporting. So in order to satisfy clause (ii) of the lemma, we need to modify the transfer rule further. The idea is to give a "bonus" to player $i$ when she reports the wrong inference $\omega(i) = \tilde{\omega}$. This gives an extra incentive to report $\omega(i) = \tilde{\omega}$ truthfully.

As in the previous step, we will first explain how to choose the transfer rule, and then provide its interpretation. Recall that in the detailed report round, player $i$ reports her full signal sequence $(z_i^t)_{t \in T(i)}$ during her own learning round. Let $(\hat{z}_i^t)_{t \in T(i)}$ denote the reported signal sequence, and let $\hat{f}_i \in \triangle Z_i$ denote the signal frequency computed from this sequence. That is, $\hat{f}_i(z_i) = \frac{|\{t \leq T | \hat{z}_i^t = z_i\}|}{T}$. Let $e(z_i)$ denote the $|Z_i|$-dimensional column vector where the component corresponding to $z_i$ is one and the remaining components are zero. Similarly, let $e(z_{-i})$ denote the $|Z_{-i}|$-dimensional column vector where the component corresponding to $z_{-i}$ is one and the remaining components are zero. We define the transfer rule $U_i^{\omega,G}$

as follows:

- If the opponent could not make the correct inference (i.e., $\omega(-i) \neq \omega$), then choose the transfer $U_i^{\omega,G}(h_{-i}^{T_b})$ as in (11). This makes player $i$ indifferent over all reports in the summary report round.

- If the opponent's inference is correct ($\omega(-i) = \omega$), then

  - If player $i$ reports $\omega(i) = \omega$ or $\omega(i) = \emptyset$, set

$$U_i^{\omega,G}(h_{-i}^{T_b}) = \tilde{U}_i^{\omega,G}(h_{-i}^{T_b}) - \frac{1-\delta^{T_b}}{\delta^{T_b}(1-\delta)} \frac{\varepsilon}{T} \sum_{t \in T(i)} \left| e(z_{-i}^t) - C_i^{\omega}(a^*) e(\hat{z}_i^t) \right|^2 .$$

  - If player $i$ reports $\omega(i) = \tilde{\omega}$, set

$$U_i^{\omega,G}(h_{-i}^{T_b}) = \tilde{U}_i^{\omega,G}(h_{-i}^{T_b}) + \frac{1-\delta^{T_b}}{\delta^{T_b}(1-\delta)} \left( b_i^{\omega}(\hat{f}_i) - \frac{\varepsilon}{T} \sum_{t \in T(i)} \left| e(z_{-i}^t) - C_i^{\omega}(a^*) e(\hat{z}_i^t) \right|^2 \right)$$

    where

$$b_i^{\omega}(\hat{f}_i) = \begin{cases} (\overline{v}_i^{\omega} + 2\overline{g}_i) p_i^{\omega}(f_i) & \text{if } |\hat{f}_i - \pi_i^{\tilde{\omega}}(a^G)| < \varepsilon \\ 0 & \text{otherwise} \end{cases} .$$

Compared to the transfer rule $\tilde{U}_i^{\omega,G}$ in the previous subsection, here we have two new terms, $b_i^{\omega}(\hat{f}_i)$ and $\frac{\varepsilon}{T} \sum_{t \in T(i)} \left| e(z_{-i}^t) - C_i^{\omega}(a^*) e(\hat{z}_i^t) \right|^2$. Very roughly speaking, the term $b_i^{\omega}(\hat{f}_i)$ helps to provide truthful incentives in the summary report round, while the term $\frac{\varepsilon}{T} \sum_{t \in T(i)} \left| e(z_{-i}^t) - C_i^{\omega}(a^*) e(\hat{z}_i^t) \right|^2$ helps to provide truthful incentives in the detailed report round. In what follows, we will explain this transfer rule in more detail.

The first bullet point considers the case in which the opponent does not have the correct inference. In this case, we choose the transfer rule just as in the previous step, that is, the transfer is chosen so that regardless of player $i$'s play, her payoff in the transfer game is $C = -2\overline{g}_i^{\omega}$. This implies that if player $i$ can observe the opponent's inference $\omega(-i)$ and if $\omega(-i) \neq \omega$, then she is indifferent over all summary reports, just as in Lemma 5.

The second bullet point considers the case in which the opponent has the correct inference $\omega(-i) = \omega$. In this case, if the transfer rule $\tilde{U}_i^{\omega,G}$ in the previous step is used, the truthful report of $\omega(i) = \tilde{\omega}$ is suboptimal; indeed, as

shown in Lemma 5, reporting $\omega(i) = \omega$ or $\omega(i) = \emptyset$ improves her payoff by $(\overline{v}_i^{\omega} + 2\overline{g}_i)p_i^{\omega}(f_i)$. To fix this problem, we give a bonus payment $b_i^{\omega}(\hat{f}_i)$ to player $i$ when she reports $\omega(-i) = \tilde{\omega}$. For simplicity, assume for now that player $i$ is truthful in the detailed report round so that $\hat{f}_i = f_i$. When $|f_i - \pi_i^{\tilde{\omega}}(a^G)| < \varepsilon$, we set the amount of the bonus equal to the expected gain by misreporting in the summary report round, $(\overline{v}_i^{\omega} + 2\overline{g}_i)p_i^{\omega}(f_i)$. This makes player $i$ indifferent over all reports in the summary report round, so the truthful summary report becomes an exact best reply. See the shaded area in Figure 7.

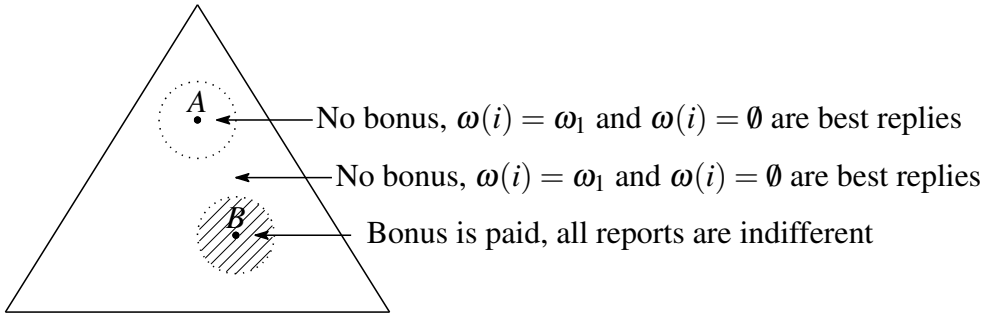


Figure 7: Player $i$'s incentive in the complete-information game with $\omega_1$, assuming that $f_i = \hat{f}_i$.

On the other hand, when $|f_i - \pi_i^{\tilde{\omega}}(a^G)| \geq \varepsilon$, we set $b_i^{\omega}(f_i) = 0$. That is, we do not pay a bonus payment even if player $i$ reports $\omega(i) = \tilde{\omega}$. This is so because in this case, Lemma 1(iii) ensures that player $i$'s true inference must be either $\omega(i) = \omega$ or $\omega(i) = \emptyset$; so if player $i$ reports $\omega(i) = \tilde{\omega}$, it should be regarded as a misreport, and we do not pay a bonus payment.

Thanks to the bonus payment above, the truthful summary report becomes an exact best reply, provided that player $i$ is truthful in the detailed report round. However, given the specification of the bonus function $b_i^{\omega}$ above, player $i$ may want to misreport in the detailed report round. Indeed, since the bonus payment $b_i^{\omega}(\hat{f}_i)$ depends on player $i$'s detailed report $\hat{f}_i$, she may want to manipulate $\hat{f}_i$ in order to maximize this bonus payment $b_i^{\omega}(\hat{f}_i)$.

To deter such a misreport in the detailed report round, we have the additional term, $\frac{\varepsilon}{T}\sum_{t\in T(i)}\left|e(z_{-i}^t) - C_i^{\omega}(a^*)e(\hat{z}_i^t)\right|^2$, in the transfer $U_i^{\omega,G}$. To better understand

this term, note that $C_i^\omega(a^*)e(\hat{z}_i^t)$ is player $i$'s *forecast* about the opponent's signal distribution in period $t$ when she observed $\hat{z}_i^t$ in that period. On the other hand, the term $e(z_{-i}^t)$ is the *actual realization* of the opponent's signal. It turns out that if player $i$ misreports $\hat{z}_i^t$, then the difference $\left|e(z_{-i}^t) - C_i^\omega(a^*)e(\hat{z}_i^t)\right|^2$ between the forecast and the realization becomes larger, which decreases the amount of the transfer. This provides an extra incentive to report $z_i^t$ truthfully in the detailed report round, and this effect is of order $\frac{1}{T}$, as the coefficient on the this term is $\frac{\varepsilon}{T}$. On the other hand, the gain by misreporting $z_i^t$ is at most of order $O(\exp(-T^{\frac{1}{2}}))$, because Lemma 2 ensures that the bonus payment is of order $O(\exp(-T^{\frac{1}{2}}))$. Since the former effect is larger than the latter, player $i$ indeed reports truthfully in the detailed report round. See Lemma 10 in the formal proof for more details.

So far we have explained that the transfer rule above induces right incentives in the two (both summary and detailed) report rounds. Note also that we have made only a small change in the transfer rule, relative to the one in the previous step; indeed, the two new terms, $b_i^\omega(\hat{f}_i)$ and $\frac{\varepsilon}{T}\sum_{t\in T(i)}\left|e(z_{-i}^t) - C_i^\omega(a^*)e(\hat{z}_i^t)\right|^2$, are quite small. Accordingly, player $i$'s payoff in the transfer game is still approximately the target payoff $\bar{v}_i^\omega$, so that clause (i) of the lemma is approximately satisfied. So by adding a small constant term to the transfer, we can satisfy clause (i) of the lemma exactly. More details are given in the formal proof.

# References


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

Aumann, R., and M. Maschler (1995): *Repeated Games with Incomplete Information*. MIT Press, Cambridge, MA. With the collaboration of R.E. Stearns.

Basu, P., K. Chatterjee, T. Hoshino, and O. Tamuz (2017): "Repeated Coordination with Private Learning," working paper.

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

Chen, B. (2010): "A Belief-Based Approach to the Repeated Prisoners' Dilemma with Asymmetric Private Monitoring," *Journal of Economic Theory* 145, 402-420.

Crémer, J., and R.P. McLean (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica* 56, 1247-1257.

Cripps, M., J. Ely, G.J. Mailath, and L. Samuelson (2008): "Common Learning," *Econometrica* 76, 909-933.

Cripps, M., J. Ely, G.J. Mailath, and L. Samuelson (2013): "Common Learning with Intertemporal Dependence," *International Journal of Game Theory* 42, 55-98.

Cripps, M., and J. Thomas (2003): "Some Asymptotic Results in Discounted Repeated Games of One-Side Incomplete Information," *Mathematics of Operations Research* 28, 433-462.

Dekel, E., D. Fudenberg, and D.K. Levine (2004): "Learning to Play Bayesian Games," *Games and Economic Behavior* 46, 282-303.

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

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

Fong, K., O. Gossner, J. Hörner, and Y. Sannikov (2011): "Efficiency in a Repeated Prisoner's Dilemma with Imperfect Private Monitoring," mimeo.

Forges, F. (1984): "Note on Nash Equilibria in Infinitely Repeated Games with Incomplete Information," *International Journal of Game Theory* 13, 179-187.

Fudenberg, D., and D.K. Levine (1991): "Approximate Equilibria in Repeated Games with Imperfect Private Information," *Journal of Economic Theory* 54, 26-47.

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

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

Fudenberg, D., and Y. Yamamoto (2010): "Repeated Games where the Payoffs and Monitoring Structure are Unknown," *Econometrica* 78, 1673-1710.

Fudenberg, D., and Y. Yamamoto (2011a): "Learning from Private Information in Noisy Repeated Games," *Journal of Economic Theory* 146, 1733-1769.

Hart, S. (1985): "Nonzero-Sum Two-Person Repeated Games with Incomplete Information," *Mathematics of Operations Research* 10, 117-153.

Hörner, J., and S. Lovo (2009): "Belief-Free Equilibria in Games with Incomplete Information," *Econometrica* 77, 453-487.

Hörner, J., S. Lovo, and T. Tomala (2011): "Belief-Free Equilibria in Games with Incomplete Information: Characterization and Existence," *Journal of Economic Theory* 146, 1770-1795.

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

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

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

Kandori, M. (2011): "Weakly Belief-Free Equilibria in Repeated Games with Private Monitoring," *Econometrica* 79, 877-892.

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

Koren, G. (1992): "Two-Person Repeated Games where Players Know Their Own Payoffs," mimeo.

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

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

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

Mailath, G.J., and W. Olszewski (2011): "Folk Theorems with Bounded Recall and (Almost) Perfect Monitoring," *Games and Economic Behavior* 71, 174-192.

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

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

Miller, D. (2012): "Robust collusion with private information," *Review of Economic Studies* 79, 778-811.

Monderer, D., and D. Samet (1989) "Approximating Common Knowledge with Common Beliefs," *Games and Economic Behavior* 1, 170-190.

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

Radner, R., R. Myerson, and E. Maskin (1986): "An Example of a Repeated Partnership Game with Discounting and with Uniformly Inefficient Equilibria," *Review of Economic Studies* 53, 59-70.

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

Shalev, J. (1994): "Nonzero-Sum Two-Person Repeated Games with Incomplete Information and Known-Own Payoffs," *Games and Economic Behavior* 7, 246-259.

Sorin, S. (1984): "Big Match with Lack of Information on One Side (Part I)," *International Journal of Game Theory* 13, 201-255.

Sorin, S. (1985): "Big Match with Lack of Information on One Side (Part II)," *International Journal of Game Theory* 14, 173-204.

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

Sugaya, T. (2012): "Belief-Free Review-Strategy Equilibrium without Conditional Independence," mimeo.

Sugaya, T. (2019): "Folk Theorem in Repeated Games with Private Monitoring," mimeo.

Sugaya, T., and Y. Yamamoto (2019): "Common Learning and Cooperation in Repeated Games," PIER Working Paper 19-008.

Wiseman, T. (2005): "A Partial Folk Theorem for Games with Unknown Payoff Distributions," *Econometrica* 73, 629-645.

Wiseman, T. (2012) "A Partial Folk Theorem for Games with Private Learning," *Theoretical Economics* 7, 217-239.

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

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

Yamamoto, Y. (2012): "Characterizing Belief-Free Review-Strategy Equilibrium Payoffs under Conditional Independence," *Journal of Economic Theory* 147, 1998-2027.

Yamamoto, Y. (2014): "Individual Learning and Cooperation in Noisy Repeated Games," *Review of Economic Studies* 81, 473-500.

# Appendix A: Proofs of Lemmas

## A.1 Proof of Lemma 1

We will formally explain how each player $i$ forms the inference $\omega(i)$ from her history $h_i^T$ in the learning round. We will introduce three different scoring rules, a *base score*, a *random score*, and a *final score*. Then we will explain how these scores are converted to the inference $\omega(i)$ and show that the resulting inference rule satisfies all the desired conditions.

*Step 1: Base Score*

For simplicity, we first consider the case in which no one deviates from $a^*$ during player $i$'s learning round. Let $f_i(a^*) = (f_i(a^*)[z_i])_{z_i \in Z_i} \in \triangle Z_i$ denote player $i$'s signal frequency during this round. Given a signal frequency $f_i(a^*)$, we compute a *base score* $q_i^{\text{base}} \in \mathbb{R}^{|Z_i|}$ using the following formula:

$$q_i^{\text{base}} = Q_i(a^*) f_i(a^*).$$

Here, $Q_i(a^*)$ is a $|Z_i| \times |Z_i|$ matrix, so it is a linear operator which maps a signal frequency $f_i(a^*) \in \triangle Z_i$ to a score vector $q_i^{\text{base}} \in \mathbb{R}^{|Z_i|}$. (Here, both $f_i(a^*)$ and $q_i^{\text{base}}$ are column vectors.) The specification of the matrix $Q_i(a^*)$ will be given later. From the law of large numbers, if the true state were $\omega$, the score $q_i^{\text{base}}$ should be close to the expected score $Q_i(a^*)\pi_i^{\omega}(a^*)$ almost surely. So if we choose a matrix such that $Q_i(a^*)\pi_i^{\omega_1}(a^*) \neq Q_i(a^*)\pi_i^{\omega_2}(a^*)$, then player $i$ can distinguish $\omega_1$ from $\omega_2$ using the base score.

If someone deviates from $a^*$ during the learning round, the base score will be computed by a slightly different formula. Given a history $h_i^T = (a^t, z_i^t)_{t=1}^T$ in player $i$'s learning round, let $\beta(a)$ denote the frequency of an action profile $a$ during the round, that is, let $\beta(a) = \frac{|\{t \in \{1,\cdots,T\} | a^t = a\}|}{T}$ for each $a$. Also, let $f_i(a) \in \triangle Z_i$ denote the signal frequency for periods in which the profile $a$ was played, that is, $f_i(a) = (f_i(a)[z_i])_{z_i \in Z_i}$ where $f_i(a)[z_i] = \frac{|\{t \in \{1,\cdots,T\} | (a^t, z_i^t) = (a, z_i)\}|}{T\beta(a)}$. For $a$ which was not played during the $T$ periods, we set $f_i(a) = 0$. We define the base score as:

$$q_i^{\text{base}} = \sum_{a \in A} \beta(a) Q_i(a) f_i(a)$$

where for each $a$, $Q_i(a)$ is a $|Z_i| \times |Z_i|$ matrix which will be specified later. In words, player $i$ computes the score vector $q_i^{\text{base}}(a) = Q_i(a) f_i(a)$ for each action

profile $a$, and takes a weighted average of these scores over all $a$. Note that this formula reduces to the previous one when no one deviates from $a^*$.

We choose the matrices $Q_i(a)$ as in the following lemma: (This lemma specifies the matrix $Q_i(a)$ only for $a$ with $a_{-j} = a^*_{-j}$. For other $a$, let $Q_i(a)$ be the normal matrix.)

**Lemma 6.** *Suppose that Conditions 2 holds. Then for each i, there are $|Z_i|$-dimensional column vectors $q_i^{\omega_1}$ and $q_i^{\omega_2}$ with $q_i^{\omega_1} \neq q_i^{\omega_2}$ such that for each $j \neq i$ and $a_j$, there is a full-rank matrix $Q_i(a_j, a^*_{-j})$ such that*

$$Q_i(a_j, a^*_{-j})\pi_i^{\omega}(a_j, a^*_{-j}) = \begin{cases} q_i^{\omega_1} & \text{if } \omega = \omega_1 \\ q_i^{\omega_2} & \text{if } \omega = \omega_2 \end{cases}.$$

*Proof.* Directly follows from Condition 2. *Q.E.D.*

That is, we choose the matrices $Q_i(a)$ so that if the true state is $\omega$, the expected base score is $q_i^{\omega}$ regardless of the opponent's actions during the learning round. Since $q_i^{\omega_1} \neq q_i^{\omega_2}$, player $i$ can indeed distinguish the true state using the base score.

While the opponent's action cannot influence the expected value of the base score, it may still influence the *distribution* of player $i$'s base score. Thus, if player $i$ uses the base score to distinguish the true state, player $j$ may be able to manipulate player $i$'s inference by deviating from $a^*$, so that clause (ii) of the lemma fails. In the next step, we will modify the scoring rule to avoid this problem.

*Step 2: Random Score*

Let $Q_i(a)$ be as in Lemma 6, and for each $z_i$, let $q_i(a, z_i)$ be the column of the matrix $Q_i(a)$ corresponding to signal $z_i$. Note that $q_i(a, z_i)$ is a $|Z_i|$-dimensional column vector, so let $q_{i,k}(a, z_i)$ denote its $k$th component. Without loss of generality, we assume that each entry of the matrix $Q_i(a)$ be in the interval $[0, 1]$, i.e., we assume that $q_{i,k}(a, z_i) \in [0, 1]$.[18]

For each $(a, z_i)$, let $\kappa_i(a, z_i) \in \{0, 1\}^{|Z_i|}$ be a random variable such that each component is randomly and independently drawn from $\{0, 1\}$ and such that for each $k$, the probability of the $k$th component being 1 is $q_{i,k}(a, z_i)$. Note that

[18] If some entry of $Q_i(a)$ is not in $[0, 1]$, we consider the affine transformation of $q_i(a, z_i)$, $q_i^{\omega_1}$, and $q_i^{\omega_2}$ so that each entry is in $[0, 1]$.

given $(a, z_i)$, the expected value of this random variable $\kappa_i(a, z_i)$ is exactly equal to $q_i(a, z_i)$.

Let $h_i^T = (a^t, z_i^t)_{t=1}^T$ denote player $i$'s history during her learning round. Given such a history $h_i^T$, define the *random score* $q_i^{\text{random}} \in \mathbb{R}^{|Z_i|}$ as

$$q_i^{\text{random}} = \frac{1}{T} \sum_{t=1}^{T} \kappa_i(a^t, z_i^t).$$

That is, we generates independent random variables $(\kappa_i(a^t, z_i^t))_{t=1}^T$ for each period-$t$ outcome $(a^t, z_i^t)$, and define the random score as its average.

Note that for a given history $h_i^T$ during the learning round, the expected value of the random score is exactly equal to the base score. This, together with the law of large numbers, implies that if the true state is $\omega$, the random score is close to $q_i^\omega$ almost surely; hence player $i$ can distinguish the state using the random score. Also, by the construction, the opponent's action cannot influence the distribution of player $i$'s random score. (Here we use Lemma 6, which ensures that the expected value of the base score does not depend on the opponent's actions.) This implies that if player $i$ uses the random score to distinguish the true state, then player $j$ cannot manipulate player $i$'s inference at all.

However, the random score is not a sufficient statistic of player $i$'s signal frequency $f_i$. For example, even when the base score is close to $q_i^\omega$ so that the signals indicate that $\omega$ is likely to be the true state, if there are too many unlucky draws of the random variables $\kappa_i(a^t, z_i^t)$, the random scores can be far away from $q_i^\omega$. Accordingly clause (iii) does not hold if player $i$ uses the random score to make the inference. In the next step, we will introduce the notion of the *final score* in order to fix this problem.

*Step 3: Final Score*

Now we introduce the concept of a *final score*, which combines the advantages of the base and random scores. Let $\tilde{\varepsilon} > 0$ be a small number. Player $i$'s final score $q_i^{\text{final}}$ is defined as

$$q_i^{\text{final}} = \begin{cases} q_i^{\text{random}} & \text{if } |q_i^{\text{random}} - q_i^{\text{base}}| < \tilde{\varepsilon} \\ q_i^{\text{base}} & \text{otherwise} \end{cases}.$$

In words, if the random score is close to the base score, it is used as the final score Otherwise, the base score is used as the final score.

By the definition, the final score is always close to the base score. This means that player $i$'s final score is an "almost sufficient" statistic for her $T$-period private history.

Another important property of the final score is that a player's action cannot influence the opponent's score almost surely. To see this, note that conditional on the $T$-period history $(a^t, z_i^t)_{t=1}^T$, the expected value of the random score $q_i^{\text{random}}$ is equal to the base score $q_i^{\text{base}}$. This implies that with probability close to one, the random score is close to the base score and hence the final score is equal to the random score, which does not depend on the opponent's deviation. Formally, for any $\tilde{\varepsilon} > 0$, there is $\overline{T}$ such that for any $T > \overline{T}$, in any period of the learning round, the probability that the opponent's action can influence player $i$'s final score is less than $\exp(-T^{\frac{1}{2}})$.

*Step 4: From the Final Score to the Inference*

Now we will describe how each player $i$ makes the inference $\omega(i)$. Recall that $\tilde{\varepsilon} > 0$ is a small number. We set $\omega(i) = \omega_1$ if

$$\left| q_i^{\omega_1} - q_i^{\text{final}} \right| < 2\tilde{\varepsilon}, \tag{13}$$

and we set $\omega(i) = \omega_2$ if

$$\left| q_i^{\omega_2} - q_i^{\text{final}} \right| < 2\tilde{\varepsilon}. \tag{14}$$

If neither (13) nor (14) holds, then we set $\omega(i) = \emptyset$. In words, if the score is in the $2\tilde{\varepsilon}$-neighborhood of the expected score at $\omega$, then we set $\omega(i) = \omega$. Note that the inference $\omega(i)$ is indeed well-defined if $\tilde{\varepsilon}$ is sufficiently small.

Now we show that this inference rule satisfies all the desired properties. Clause (i) is simply a consequence of the law of large numbers. Clause (ii) follows from the fact that the opponent's deviation cannot influence player $i$'s final score almost surely.

To prove clause (iii), suppose that no one deviates from $a^*$, and pick a signal frequency $f_i$ such that player $i$ will choose $\omega(i) = \omega$ with positive probability. By the definition of the final score, given this signal frequency $f_i$, the resulting final score is always within $\tilde{\varepsilon}$ of the base score $q_i^{\text{base}}$, which is equal to $Q_i(a^*)f_i$. Hence, from (13) and (14), we must have

$$|q_i^{\omega} - Q_i(a^*)f_i| < 3\tilde{\varepsilon}. \tag{15}$$

Since $Q_i(a^*)$ has a full rank, this implies

$$|\pi_i^{\omega}(a^*) - f_i| < K\tilde{\varepsilon} \tag{16}$$

for some constant $K > 0$. Hence clause (iii) follows.

## A.2 Proof of Lemma 3

As in Section 4.5, we first construct a transfer rule $\tilde{U}_i^{\omega,G}$ which "approximately" satisfies clause (ii) of the lemma. That is, we construct $\tilde{U}_i^{\omega,G}$ such that playing the prescribed strategy $s_i^{x_i}$ is a best reply for player $i$ except the summary report round, and it is an approximate best reply in the summary report round. Then we modify this transfer rule $\tilde{U}_i^{\omega,G}$ and construct a new transfer rule $U_i^{\omega,G}$ which satisfies clause (ii) exactly. Then we show that the modified transfer rule $U_i^{\omega,G}$ satisfies clauses (i) and (iii) as well.

We begin with introducing the notion of *regular histories*. We first give the definition and then give its interpretation. A block history $h_{-i}^{T_b}$ is *regular given* $(\omega, G)$ if it satisfies all the following conditions:

(G1) Players choose $a^*$ in the learning round.

(G2) In the summary report round, the opponent reports $\omega(-i) = \omega$, and player $i$ reports $\omega(i) = \omega$ or $\omega(i) = \emptyset$.

(G3) The opponent reports $x_{-i}^{\omega} = G$ in the first period of the main round,

(G4) Players follow the prescribed strategy in the second or later periods of the main round.

(G5) The opponent's signal frequency $f_{-i}$ during player $i$'s learning round is close to the ex-ante distribution $\pi_{-i}^{\omega}(a^*)$, i.e., $|f_{-i} - \pi_{-i}^{\omega}(a^*)| < \varepsilon$.

A history $h_{-i}^{T_b}$ is *irregular given* $(\omega, G)$ if it is not regular.

Roughly, a history is regular if (i) no one makes an observable deviation from the prescribed strategy $s^x$, and (ii) no one reports a wrong inference, and (iii) the opponent's signal frequency $f_{-i}$ is typical of $\omega$. Note that this concept is an extension of "regular observations" briefly discussed in Section 4.5; now we allow players' deviations in the learning and the main round, and we call the history irregular if such a deviation occurs.

### A.2.1 Step 1: Construction of $\tilde{U}_i^{\omega,G}$

Choose a transfer rule $\tilde{U}_i^{\omega,G} : H_{-i}^{T_b} \to \boldsymbol{R}$ such that

- If the history $h_{-i}^{T_b}$ is regular given $(\omega, G)$, choose $\tilde{U}_i^{\omega,G}(h_{-i}^{T_b})$ so that it solves

$$\frac{1-\delta}{1-\delta^{T_b}}\left[\sum_{t=1}^{T_b}\delta^{t-1}g_i^{\omega}(a^t)+\delta^{T_b}\tilde{U}_i^{\omega,G}(h_{-i}^{T_b})\right]=\overline{v}_i^{\omega}. \tag{17}$$

- If the history $h_{-i}^{T_b}$ is irregular, choose $\tilde{U}_i^{\omega,G}(h_{-i}^{T_b})$ so that

$$\frac{1-\delta}{1-\delta^{T_b}}\left[\sum_{t=1}^{T_b}\delta^{t-1}g_i^{\omega}(a^t)+\delta^{T_b}\tilde{U}_i^{\omega,G}(h_{-i}^{T_b})\right]=-2\overline{g}_i^{\omega}. \tag{18}$$

In words, if (i) no one makes an observable deviation, and (ii) no one reports a wrong inference, and (iii) the opponent's observation $f_{-i}$ is typical of $\omega$, then the transfer $U_i^{\omega,G}$ is chosen in such a way that player $i$'s payoff in the complete-information transfer game is exactly the target payoff $\overline{v}_i^{\omega}$. On the other hand, if player $i$ makes an observable deviation or reports a wrong inference, or if the opponent's observation is not typical of $\omega$, then we give a huge negative transfer to player $i$ so that the payoff goes down to $-2\overline{g}_i^{\omega}$. Note that this transfer rule is very similar to the one in Section 4.5; the only difference is that player $i$ receives a huge negative transfer when there is a deviation in the learning round or in the main round. So (assuming that no one has deviated in the learning round) player $i$'s best reply in the summary report round is still as in Table 1 in Section 4.5.

### A.2.2 Step 2: $\tilde{U}_i^{\omega,G}$ approximately satisfies clause (ii)

Consider the complete-information transfer game with the state $\omega$ and the transfer rule $\tilde{U}_i^{\omega,G}$ above. Suppose that the opponent's current plan is $x_{-i}$ with $x_{-i}^{\omega} = G$. We will show that the prescribed strategies $s_i^{GG}$, $s_i^{GB}$, $s_i^{BG}$, and $s_i^{BB}$ are approximate best replies for player $i$. We will first show that the strategies $s_i^{GG}$, $s_i^{GB}$, $s_i^{BG}$, and $s_i^{BB}$ are exact best replies except the summary report round.

**Lemma 7.** *In the learning round, the main round, and the detailed report round, the strategies $s_i^{GG}$, $s_i^{GB}$, $s_i^{BG}$, and $s_i^{BB}$ are best replies for player i, regardless of the past history.*

*Proof.* Actions in the detailed report round and in the first period of the main round do not influence whether the resulting history is regular or not. Hence player $i$ is indifferent over all actions in these periods.

In the learning round and the second or later periods of the main round, player $i$ prefers not to deviate from the prescribed strategy $s_i^{x_i}$. This is so because such deviations are observable and make the history irregular for sure, which yields the worst payoff payoff $-2\overline{g}_i^{\omega}$. *Q.E.D.*

In what follows, we will focus on the incentive problem in the summary report round. The next lemma shows that if someone has deviated during the learning round, then the truthful summary report is an exact best reply.

**Lemma 8.** *Suppose that someone has deviated from $a^*$ during the learning round. Then player $i$ is indifferent over all actions in the summary report round, and hence the truthful summary report is a best reply.*

*Proof.* If someone has deviated from $a^*$ in the learning round, then the opponent's history $h_{-i}^{T_b}$ becomes irregular, regardless of player $i$'s summary report. Hence player $i$ is indifferent over all summary reports. *Q.E.D.*

Now, consider the case in which no one has deviated during the learning round. In this case, Lemma 5 still holds, because the transfer rule constructed above is exactly the same as the one in Section 4.5. So the truthful summary report is indeed an approximate best reply.

### A.2.3 Step 3: Construction of $U_i^{\omega,G}$ and Clause (ii)

As explained, the transfer rule $\tilde{U}_i^{\omega,G}$ approximately satisfies clause (ii) of Lemma 3, but not exactly. Indeed, as shown in Lemma 5, the truthful report of $\omega(i) = \tilde{\omega}$ in the summary report round is not an exact best reply. So we will modify the transfer rule $\tilde{U}_i^{\omega,G}$ in such a way that (ii) holds exactly. The idea here is very similar to the one presented in Step 2 in Section 4.5; we give a "bonus" to player $i$ when she reports the incorrect inference $\omega(i) = \tilde{\omega}$, which gives her an extra incentive to report $\omega(i) = \tilde{\omega}$ truthfully.

Define a *bonus function* $b_i^{\omega} : H_{-i}^{T_b} \to \boldsymbol{R}$ as

$$b_i^{\omega}(h_{-i}^{T_b}) = \begin{cases} 0 & \text{if player } i \text{ reports } \omega(i) = \omega \text{ or } \omega(i) = \emptyset \\ 0 & \text{if someone deviates in the learning round} \\ 0 & \text{if } \omega(-i) \neq \omega \\ 0 & \text{if } |\hat{f}_i - \pi_i^{\tilde{\omega}}(a^*)| \geq \varepsilon \\ (\overline{v}_i^{\omega} + 2\overline{g}_i^{\omega}) p_i^{\omega}(\hat{f}_i) & \text{otherwise} \end{cases}.$$

This bonus function is the same as the one in Section 4.5, except that we specify values for the case in which someone makes observable deviations. Recall that the amount of the bonus by reporting $\omega(i) = \tilde{\omega}$ is $(\overline{v}_i^{\omega} + 2\overline{g}_i^{\omega}) p_i^{\omega}(\hat{f}_i)$, which is exactly equal to the expected gain by misreporting in the summary report round (Lemma 5). This makes player $i$ indifferent over all reports in the summary report round, and thus the truthful report of $\omega(i) = \tilde{\omega}$ becomes a best reply.

The following lemma shows that the amount of the bonus, $b_i^{\omega}(h_{-i}^{T_b})$, is very small regardless of the opponent's history $h_{-i}^{T_b}$. In order to obtain this lemma, it is crucial that we pay a bonus only if $|\hat{f}_i - \pi_i^{\tilde{\omega}}(a^*)| < \varepsilon$; this condition ensures that $p_i^{\omega}(\hat{f}_i)$ is small and so is the bonus.

**Lemma 9.** *There is $\overline{T}$ such that for any $T > \overline{T}$ and $h_{-i}^{T_b}$, we have $b_i^{\omega}(h_{-i}^{T_b}) < 3\overline{g}_i^{\omega} \exp(-T^{\frac{1}{2}})$.*

*Proof.* Lemma 2 implies that whenever $|\hat{f}_i - \pi_i^{\tilde{\omega}}(a^*)| < \varepsilon$, we have $p_i^{\omega}(\hat{f}_i) < \exp(-T^{\frac{1}{2}})$. Then by the definition of $b_i^{\omega}$, we obtain the lemma. *Q.E.D.*

Now we define the new transfer rule $U_i^{\omega,G}$ as

$$U_i^{\omega,G}(h_{-i}^{T_b}) = \tilde{U}_i^{\omega,G}(h_{-i}^{T_b}) + \frac{1-\delta^{T_b}}{\delta^{T_b}(1-\delta)} \left( c^G + b_i^{\omega}(h_{-i}^{T_b}) - \frac{\varepsilon}{T} \sum_{t \in T(i)} \left| e(z_{-i}^t) - C_i^{\omega}(a^*) e(\hat{z}_i^t) \right|^2 \right).$$

where $c^G$ is a constant term which will be specified later. Again the specification of the transfer rule is very similar to the one in Section 4.5; a key is that we add the terms $b_i^{\omega}(h_{-i}^{T_b})$ and $\frac{\varepsilon}{T} \sum_{t \in T(i)} \left| e(z_{-i}^t) - C_i^{\omega}(a^*) e(\hat{z}_i^t) \right|^2$ in order to provide right incentives in the two report rounds.

In what follows, we will verify that this transfer rule indeed satisfies clause (ii) of the lemma. That is, the prescribed strategy $s_i^{x_i}$ is a best reply in the transfer game. The following lemma considers incentives in the detailed report round:

**Lemma 10.** *There is $\overline{T} > 0$ such that for any $T > \overline{T}$, the truthful report in the detailed report round is a best reply for player i regardless of the past history. In particular, the truthful report is a best reply even if player i has misreported in the summary report round.*

*Proof.* Recall that under the transfer rule $\tilde{U}_i^{\omega,G}$, player $i$ is indifferent over all actions in the detailed report round. (This is so because her actions in the detailed report round cannot influence whether the opponent's history is regular or not.) Thus, it is sufficient to check how player $i$'s deviation in the detailed report round influences the additional terms, $b_i^{\omega}(h_{-i}^{T_b}) - \frac{\varepsilon}{T}\sum_{t\in T(i)} \left|e(z_{-i}^t) - C_i^{\omega}(a^*)e(\hat{z}_i^t)\right|^2$.

In the detailed report round, player $i$ reports the signals $(z_i^t)_{t\in T(i)}$ during her own learning round, and the ones $(z_i^t)_{t\in T(-i)}$ during the opponent's learning round. It is easy to see that the truthful report of $(z_i^t)_{t\in T(-i)}$ is a best reply for player $i$, because this report does not influence the additional terms above. So what remains is to show that the truthful report of the signals $(z_i^t)_{t\in T(i)}$ during her own learning round is a best reply for player $i$.

Pick some $t \in T(i)$, and suppose that player $i$ deviates by reporting a signal $\tilde{z}_i \neq z_i^t$ such that $C_i^{\omega}(a^*)e(z_i^t) \neq C_i^{\omega}(a^*)e(\tilde{z}_i)$; that is, consider a misreport $\tilde{z}_i$ such that the corresponding posterior distribution of $z_{-i}$ differs from the true posterior distribution $C_i^{\omega}(a^*)e(z_i^t)$. This misreport increases the expected value of $\left|e(z_{-i}^t) - C_i^{\omega}(a^*)e(\hat{z}_i^t)\right|^2$, and hence reduces the expected transfer.[19] This effect is of order $\frac{1}{T}$, as we have the coefficient $\frac{\varepsilon}{T}$. This implies that this misreport is not profitable, as the gain is at most of order $\exp(-T^{\frac{1}{2}})$ from Lemma 9.

Next, suppose that player $i$ deviates by reporting a signal $\tilde{z}_i \neq z_i^t$ such that $C_i^{\omega}(a^*)e(z_i^t) = C_i^{\omega}(a^*)e(\tilde{z}_i)$. In this case, player $i$'s payoff is the same as the one when she does not deviate; indeed, this misreport does not change $b_i^{\omega}(h_{-i}^{T_b})$ or $\left|e(\hat{z}_{-i}^t) - C_i^{\omega}(a^*)e(\hat{z}_i^t)\right|^2$. Hence this misreport is not profitable. *Q.E.D.*

The next lemma shows that thanks to the bonus function $b_i^{\omega}$, the truthful report

[19] Indeed, as explained in Section 4.2 of Kandori and Matsushima (1998), we have

$$\sum_{z_{-i}\in Z_{-i}} C_i^{\omega}(a^*)e(z_i^t)[z_{-i}]\left|e(z_{-i}) - C_i^{\omega}(a^*)e(z_i^t)\right|^2 < \sum_{z_{-i}\in Z_{-i}} C_i^{\omega}(a^*)e(z_i^t)[z_{-i}]\left|e(z_{-i}) - C_i^{\omega}(a^*)e(\tilde{z}_i^t)\right|^2$$

for this misreport $\tilde{z}_i^t$, so the expected transfer indeed decreases. Note that the opponent's block strategy does not depend on the signal $z_{-i}^t$, so regardless of the opponent's past actions, player $i$'s posterior belief about $z_{-i}^t$ is indeed $C_i^{\omega}(a^*)e(z_i^t)$.

in the summary report round is an exact best reply. This implies that the modified transfer $U_i^{\omega,G}$ satisfies Lemma 3(ii).

**Lemma 11.** *The truthful report in the summary report round is a best reply for player i, regardless of the past history.*

*Proof.* Throughout the proof, we assume that player $i$ will be truthful in the detailed report round, since we have Lemma 10. Suppose, hypothetically, that player $i$ knows the opponent's inference $\omega(-i)$ before it is revealed in the summary report round. We will show that the truthful report of $\omega(i)$ is a best reply for player $i$ regardless of $\omega(-i)$. This implies that the truthful report is a best reply even if player $i$ does not know $\omega(-i)$, and hence the result.

First, suppose that someone deviated from $a^*$ in the learning round or the opponent's inference is $\omega(-i) \neq \omega$. In these cases, the bonus payment is zero regardless of player $i$'s summary report. Also, from Lemmas 8 and 5, player $i$ is indifferent over all actions in the summary report round with the transfer $\tilde{U}_i^{\omega,G}$. Hence player $i$ is indifferent over all actions in the summary report round even with the new transfer rule, and the truthful report is a best reply.

Next, suppose that no one has deviated in the learning round, and that the opponent's inference is $\omega(-i) = \omega$. There are two cases to be considered.

*Case 1: Player i's signal frequency $f_i$ during her own learning round is such that $|\pi_i^{\tilde{\omega}}(a^*) - f_i| \geq \varepsilon$.* In this case, from Lemma 1(iii), player $i$'s inference must be either $\omega(i) = \omega$ or $\omega(i) = \emptyset$. Then from Lemma 5, the truthful report of $\omega(i)$ in the summary report round is a best reply under the transfer rule $\tilde{U}_i^{\omega,G}$. The same result holds even under the new transfer $U_i^{\omega,G}$, because given that $|\pi_i^{\tilde{\omega}}(a^*) - f_i| \geq \varepsilon$, the bonus payment $b_i^{\omega}$ is zero regardless of player $i$'s summary report.

*Case 2: Player i's signal frequency $f_i$ during her own learning round is such that $|\pi_i^{\tilde{\omega}}(a^*) - f_i| < \varepsilon$.* We claim that in this case, player $i$ is indifferent over all summary reports (and hence the truthful report of $\omega(i)$ is a best reply). Under the transfer rule $\tilde{U}_i^{\omega,G}$, reporting $\omega(i) = \omega$ yields an expected payoff of $p_i^{\omega}(f_i)\overline{v}_i^{\omega} + (1 - p_i^{\omega}(f_i))(-2\overline{g}_i^{\omega})$, since the probability of the block history being regular is $p_i^{\omega}(f_i)$. The same is true when player $i$ reports $\omega(i) = \emptyset$. On the other hand, when player $i$ reports $\omega(i) = \tilde{\omega}$, the block history is always irregular, and hence the expected payoff is $-2\overline{g}_i^{\omega}$. Obviously this payoff is worse than the one by

reporting $\omega(i) = \omega$, and the payoff difference is

$$(p_i^\omega(f_i)\overline{v}_i^\omega + (1 - p_i^\omega(f_i))(-2\overline{g}_i^\omega)) - 2\overline{g}_i^\omega = (\overline{v}_i^\omega + 2\overline{g}_i^\omega)p_i^\omega(f_i).$$

Now, consider the modified transfer $U_i^{\omega,G}$, with which player $i$ can obtain the bonus $b_i^\omega$ by reporting $\tilde{\omega}$ in the summary report round. Since the amount of the bonus is precisely equal to the payoff difference above, player $i$ is indifferent over all summary reports, as desired. *Q.E.D.*

### A.2.4 Step 4: Proof of Clause (i)

In what follows, we will show that the transfer rule $U_i^{\omega,G}$ satisfies clauses (i) and (iii) of Lemma 3, if we choose the constant term $c^G$ appropriately.

Let $p_{-i}^\omega$ denote the probability of the opponent's block history $h_{-i}^{T_b}$ being regular given $(\omega, G)$, conditional on that the state is $\omega$ and players play $s^x$ with $x_{-i}^\omega = G$. Note that this probability does not depend on the choice of $x$ as long as $x_{-i}^\omega = G$, so it is well-defined. Then let

$$c^G = (1 - p_{-i}^\omega)(\overline{v}_i^\omega + 2\overline{g}_i^\omega) + E\left[\frac{\varepsilon}{T}\sum_{t\in T(i)}\left|e(z_{-i}^t) - C_i^\omega(a^*)e(\hat{z}_i^t)\right|^2 - b_i^\omega(h_{-i}^{T_b})\middle|\omega, s^x\right]. \tag{19}$$

Again, the expected value of $|e(z_{-i}^t) - C_i^\omega(a^*)e(\hat{z}_i^t)|^2$ and $b_i^\omega(h_{-i}^{T_b})$ does not depend on the choice of $x$, and thus $c^G$ is well-defined.

Given this constant term $c^G$, the resulting transfer rule $U_i^{\omega,G}$ satisfies Lemma 3(i). To see why, suppose that players play $s^x$ with $x_{-i}^\omega = G$. It follows from (17) and (18) that if the transfer rule $\tilde{U}_i^{G,\omega}$ is used, player $i$'s expected payoff in the complete-information transfer game is

$$p_{-i}^\omega \overline{v}_i^\omega - (1 - p_{-i}^\omega)2\overline{g}_i^\omega,$$

where $p_{-i}^\omega$ is the probability of the opponent's history being regular. Hence, if the modified transfer rule $U_i^{G,\omega}$ is used, player $i$'s payoff in the complete-information transfer game is

$$\begin{aligned}\frac{1-\delta}{1-\delta^{T_b}}G_i^\omega(s^x, U_i^{\omega,G}) =& p_{-i}^\omega \overline{v}_i^\omega - (1 - p_{-i}^\omega)2\overline{g}_i^\omega + c^G \\ &+ E\left[b_i^\omega(h_{-i}^{T_b}) - \frac{\varepsilon}{T}\sum_{t\in T(i)}\left|e(z_{-i}^t) - C_i^\omega(a^*)e(\hat{z}_i^t)\right|^2\middle|\omega, s^x\right].\end{aligned}$$

Plugging (19) into this equation, we obtain clause (i) of Lemma 3.

### A.2.5 Step 5: Proof of Clause (iii)

What remains is to prove Lemma 3(iii). That is, we need to show $-(\overline{v}_i^{\omega} - \underline{v}_i^{\omega}) < (1-\delta)U_i^{\omega,G}(h_{-i}^{T_b}) < 0$ for all $h_{-i}^{T_b}$.

We begin with showing the first inequality, $-(\overline{v}_i^{\omega} - \underline{v}_i^{\omega}) < (1-\delta)U_i^{\omega,G}(h_{-i}^{T_b})$. By the definition of $\overline{g}_i^{\omega}$, we have $\frac{1-\delta}{1-\delta^{T_b}}\sum_{t=1}^{T_b}\delta^{t-1}g_i^{\omega}(a^t) \geq \overline{g}_i^{\omega}$ regardless of the action sequence $(a^1,\cdots,a^{T_b})$. Plugging this into (17) and (18), we obtain

$$\frac{\delta^{T_b}(1-\delta)}{1-\delta^{T_b}}\tilde{U}_i^{\omega,G}(h_{-i}^{T_b}) \geq -3\overline{g}_i^{\omega},$$

and hence

$$\frac{\delta^{T_b}(1-\delta)}{1-\delta^{T_b}}U_i^{\omega,G}(h_{-i}^{T_b}) \geq -3\overline{g}_i^{\omega} + c^G + b_i^{\omega}(h_{-i}^{T_b}) - \frac{\varepsilon}{T}\sum_{t\in T(i)}\left|e(z_{-i}^t) - C_i^{\omega}(a^*)e(\hat{z}_i^t)\right|^2$$

for each $h_{-i}^{T_b}$. Equivalently,

$$(1-\delta)U_i^{\omega,G}(h_{-i}^{T_b}) \geq \frac{1-\delta^{T_b}}{\delta^{T_b}}\left(-3\overline{g}_i^{\omega} + c^G + b_i^{\omega}(h_{-i}^{T_b}) - \frac{\varepsilon}{T}\sum_{t\in T(i)}\left|e(z_{-i}^t) - C_i^{\omega}(a^*)e(\hat{z}_i^t)\right|^2\right).$$

For a fixed $T$, if we take $\delta$ close to one, $\frac{1-\delta^{T_b}}{\delta^{T_b}}$ becomes arbitrarily close to zero, so that the right-hand side is greater than $-(\overline{v}_i^{\omega} - \underline{v}_i^{\omega})$. This implies the desired inequality, $-(\overline{v}_i^{\omega} - \underline{v}_i^{\omega}) < (1-\delta)U_i^{\omega,G}(h_{-i}^{T_b})$.

Now we prove the remaining inequality, $(1-\delta)U_i^{\omega,G}(h_{-i}^{T_b}) < 0$. We consider the following two cases.

*Case 1: $h_{-i}^{T_b}$ is regular given $(\omega,G)$.* In this case, in all but one period of the main round, players play $a^{\omega,x^{\omega}}$ with $x_{-i}^{\omega} = G$, which yields more than $\overline{v}_i^{\omega} + 2\varepsilon$ to player $i$, according to (3) and (4). So for sufficiently large $T$ and $\delta$ close to one, we have $\frac{1-\delta}{1-\delta^{T_b}}\sum_{t=1}^{T_b}\delta^{t-1}g_i^{\omega}(a^t) > \overline{v}_i^{\omega} + 2\varepsilon$. Plugging this into (17),

$$\frac{(1-\delta)\delta^{T_b}}{1-\delta^{T_b}}\tilde{U}_i^{\omega,G}(h_{-i}^{T_b}) < -2\varepsilon.$$

Hence

$$\begin{aligned}\frac{(1-\delta)\delta^{T_b}}{1-\delta^{T_b}}U_i^{\omega,G}(h_{-i}^{T_b}) &< -2\varepsilon + c^G + b_i^{\omega}(h_{-i}^{T_b}) - \frac{\varepsilon}{T}\sum_{t\in T(i)}\left|e(z_{-i}^t) - C_i^{\omega}(a^*)e(\hat{z}_i^t)\right|^2 \\ &\leq -2\varepsilon + c^G + b_i^{\omega}(h_{-i}^{T_b}).\end{aligned}$$

Note that

$$c^G \leq (1-p_{-i}^{\omega})(\overline{v}_i^{\omega}+2\overline{g}_i^{\omega})+\sqrt{2}\varepsilon - E\left[b_i^{\omega}(h_{-i}^{T_b})|\omega,s^x\right],$$

since $|e(z_{-i}^t)-C_i^{\omega}(a^*)e(\hat{z}_i^t)|^2 \leq \sqrt{2}$. Plugging this into the above inequality, we have

$$\begin{aligned}&\frac{(1-\delta)\delta^{T_b}}{1-\delta^{T_b}}U_i^{\omega,G}(h_{-i}^{T_b})\\ &<-(2-\sqrt{2})\varepsilon+(1-p_{-i}^{\omega})(\overline{v}_i^{\omega}+2\overline{g}_i^{\omega})-E\left[b_i^{\omega}(h_{-i}^{T_b})|\omega,s^x\right]+b_i^{\omega}(h_{-i}^{T_b}).\end{aligned}$$

Note that when $T$ is large, $p_{-i}^{\omega}$ approximates 1 and $b_i^{\omega}(h_{-i}^{T_b})$ approximates 0 for all $h_{-i}^{T_b}$. (This follows from Lemma 9.) Hence for sufficiently large $T$,

$$\frac{(1-\delta)\delta^{T_b}}{1-\delta^{T_b}}U_i^{\omega,G}(h_{-i}^{T_b})<-(2-\sqrt{2})\varepsilon<0$$

as desired.

*Case 2: $h_{-i}^{T_b}$ is irregular given $(\omega,G)$.* The proof is very similar to the one for Case 1, and hence omitted.

## A.3 Proof of Lemma 4

Fix $i$ and $\omega$ arbitrarily. In what follows, we will construct a transfer rule $U_i^{\omega,B}$ which satisfies clauses (i) through (iii) in Lemma 4.

We begin with introducing the notion of *regular histories*. The definition here is slightly different from the one in the proof of Lemma 3. The opponent's history is regular if she does not deviate from the prescribed strategy $s_{-i}^{x_{-i}}$ and she makes the correct inference $\omega(-i)=\omega$. Formally, the opponent's block history $h_{-i}^{T_b}$ is *regular given* $(\omega,B)$ if it satisfies all the following conditions:

(B1) Player $-i$ chooses $a_{-i}^*$ in the learning round.

(B2) Player $-i$ reports $\omega(-i)=\omega$.

(B3) Player $-i$ reports $x_{-i}^{\omega}=B$ in the first period of the main round.

(B4) Player $-i$ followed the prescribed strategy $s_{-i}^{x_{-i}}$ in the second or later periods of the main round.

A history $h_{-i}^{T_b}$ is *irregular given* $(\omega,B)$ if it is not regular.

### A.3.1 Step 1: Construction of $U_i^{\omega,B}$

Let $c^B > 0$ be a constant which will be specified later. Then choose a transfer rule $U_i^{\omega,B} : H_{-i}^{T_b} \to \boldsymbol{R}$ so that

- For each history $h_{-i}^{T_b} = (a^t, z_{-i}^t)_{t=1}^{T_b}$ which is regular given $(\omega, B)$, choose $U_i^{\omega,B}(h_{-i}^{T_b})$ so that it solves

$$\frac{1-\delta}{1-\delta^{T_b}}\left[\sum_{t=1}^{T_b}\delta^{t-1}g_i^{\omega}(a^t) + \delta^{T_b}U_i^{\omega,B}(h_{-i}^{T_b})\right] = \underline{v}_i^{\omega} - \frac{\tau\varepsilon}{T} - c^B \tag{20}$$

  where $\tau$ is the number of periods such that player $i$ deviated from $a^*$ during the opponent's learning round.

- For each irregular $h_{-i}^{T_b}$, choose $U_i^{\omega,B}(h_{-i}^{T_b})$ so that

$$\frac{1-\delta}{1-\delta^{T_b}}\left[\sum_{t=1}^{T_b}\delta^{t-1}g_i^{\omega}(a^t) + \delta^{T_b}U_i^{\omega,B}(h_{-i}^{T_b})\right] = 2\overline{g}_i^{\omega} - \frac{\tau\varepsilon}{T} - c^B. \tag{21}$$

In words, if the opponent plays the prescribed strategy and reports the correct inference $\omega(-i) = \omega$ (so that the history $h_{-i}^{T_b}$ is regular), we adjust the transfer $U_i^{\omega,B}(h_{-i}^{T_b})$ in such a way that player $i$'s total payoff in the complete-information transfer game is $\underline{v}_i^{\omega} - c^B$. As will be explained, $c^B$ is a constant number close to zero; so this payoff is approximately the target payoff $\underline{v}_i^{\omega}$. On the other hand, if the opponent deviates or reports something else, we give a huge positive transfer to player $i$, and her total payoff goes up to $2\overline{g}_i^{\omega} - c^B$. If player $i$ deviates in the opponent's learning round, it decreases the transfer a bit, due to the term $\frac{\tau\varepsilon}{T}$.

### A.3.2 Step 2: Proof of Clause (ii)

We claim that the transfer rule above satisfies clause (ii) of Lemma 4. That is, we will show that the prescribed strategies $s_i^{GG}$, $s_i^{GB}$, $s_i^{BG}$, and $s_i^{BB}$ are all best replies in the complete-information transfer game with $(\omega, U_i^{\omega,B})$, if the opponent's current plan is $x_{-i}$ with $x_{-i}^{\omega} = B$. The result follows from the following two lemmas.

**Lemma 12.** *Player i is indifferent over all actions in player i's learning round, the summary report round, the main round, and the detailed report round, regardless of the past history. Hence, deviating from $s_i^{x_i}$ during these rounds is not profitable.*

*Proof.* By the construction of $U_i^{\omega,B}$, player $i$'s payoff in the complete-information transfer game depends only on whether the opponent's block history $h_{-i}^{T_b}$ is regular or not, and on the number of periods such that player $i$ deviated from $a^*$ during the opponent's learning round. The result follows because player $i$'s play cannot influence whether the resulting history is regular or not. *Q.E.D.*

**Lemma 13.** *When $T$ is large enough, $a_i^*$ is the unique best reply in each period of the opponent's learning round, regardless of the past history. Hence, deviating from $s_i^{x_i}$ during the opponent's learning round is not profitable.*

*Proof.* During the opponent's learning round, deviating from $a_i^*$ has two effects: First, it affects the distribution of the opponent's inference $\omega(-i)$, and hence the probability of the opponent's history being regular. Second, it decreases the transfer $U_i^{\omega,B}$ due to the term $\frac{\tau\varepsilon}{T}$. From Lemma 1(ii) and the law of large numbers (more precisely, Hoeffding's inequality), the first effect is at most of order $O(\exp(-T^{\frac{1}{2}}))$. On the other hand, the second effect is proportional to $\frac{1}{T}$. Thus for large $T$, the second effect dominates, so that playing $a_i^*$ is optimal. This shows that clause (ii) of Lemma 4 holds. *Q.E.D.*

### A.3.3 Step 3: Proof of Clause (i)

Now we choose the constant term $c^B$ in such a way that the resulting transfer rule $U_i^{\omega,B}$ satisfies clause (i) of Lemma 4.

Let $p_{-i}^{\omega}$ denote the probability of the opponent making the correct inference $\omega(-i)=\omega$, given that the true state is $\omega$ and players play $a^*$ in the learning round. Then let

$$c^B = (1-p_{-i}^{\omega})(2\overline{g}_i^{\omega} - \underline{v}_i^{\omega}) > 0. \tag{22}$$

Given this constant term $c^B$, the resulting transfer rule $U_i^{\omega,B}$ satisfies clause (i) of Lemma 4. To see why, suppose that players play $s^x$ with $x_{-i}^{\omega}=B$. It follows from (20) and (21) that player $i$'s expected payoff in the complete-information transfer game is

$$\frac{1-\delta}{1-\delta^{T_b}} G_i^{\omega}(s^x, U_i^{\omega,B}) = p_{-i}^{\omega}(\underline{v}_i^{\omega} - c^B) + (1-p_{-i}^{\omega})(2\overline{g}_i^{\omega} - c^B),$$

where $p_{-i}^{\omega}$ is the probability of the opponent's history being regular. Plugging (22) into this equation, we obtain clause (i) of Lemma 4.

### A.3.4 Step 4: Proof of Clause (iii)

To complete the proof of Lemma 4, we need to show that the constructed transfer rule $U_i^{\omega,B}$ satisfies clause (iii) of Lemma 4.

We first show that $(1-\delta)U_i^{\omega,B}(h_{-i}^{T_b}) < \overline{v}_i^{\omega} - \underline{v}_i^{\omega}$ for each $h_{-i}^{T_b}$. By the definition of $\overline{g}_i^{\omega}$, player $i$'s average payoff in the block, $\frac{1-\delta}{1-\delta^{T_b}}\left[\sum_{t=1}^{T_b}\delta^{t-1}g_i^{\omega}(a^t)\right]$, is at least $-\overline{g}_i^{\omega}$. Then from (20), (21), and $c^B > 0$, we have $\frac{\delta^{T_b}(1-\delta)}{1-\delta^{T_b}}U_i^{\omega,B}(h_{-i}^{T_b}) < 3\overline{g}_i^{\omega}$, equivalently, $(1-\delta)U_i^{\omega,B}(h_{-i}^{T_b}) < \frac{(1-\delta^{T_b})3\overline{g}_i^{\omega}}{\delta^{T_b}}$. For a fixed $T$, by taking sufficiently large $\delta$, the right-hand side becomes arbitrarily small. Hence we have $(1-\delta)U_i^{\omega,B}(h_{-i}^{T_b}) < \overline{v}_i^{\omega} - \underline{v}_i^{\omega}$.

Next, we show that $U_i^{\omega,B}(h_{-i}^{T_b}) > 0$ for each $h_{-i}^{T_b}$. We consider the following two cases.

*Case 1: $h_{-i}^{T_b}$ is regular given $(\omega,B)$.* In this case, in most periods of the main round, players played the action profile $a^{\omega,x^{\omega}}$ with $x_{-i}^{\omega} = B$ or the opponent played the minimax action $\underline{\alpha}_{-i}^{\omega}(i)$. Both these actions yield payoffs lower than $\underline{v}_i^{\omega} - \varepsilon$ to player $i$, according to (1) and (2). Hence, when $T$ is sufficiently large and $\delta$ is close to one, we have

$$\frac{1-\delta}{1-\delta^{T_b}}\left[\sum_{t=1}^{T_b}\delta^{t-1}g_i^{\omega}(a^t)\right] < \underline{v}_i^{\omega} - \varepsilon.$$

Then since $c^B \to 0$ as $T \to \infty$ (this follows from the fact that Lemma 1 ensures $p_{-i}^{\omega} \to 1$), we obtain

$$\frac{1-\delta}{1-\delta^{T_b}}\left[\sum_{t=1}^{T_b}\delta^{t-1}g_i^{\omega}(a^t)\right] < \underline{v}_i^{\omega} - \varepsilon - c^B.$$

Plugging this into (20), we obtain $U_i^{\omega,B}(h_{-i}^{T_b}) > 0$.

*Case 2: $h_{-i}^{T_b}$ is irregular given $(\omega,B)$.* Since the value $\overline{g}_i^{\omega}$ is greater than player $i$'s stage-game payoff for any action profile $a$, we have

$$\frac{1-\delta}{1-\delta^{T_b}}\left[\sum_{t=1}^{T_b}\delta^{t-1}g_i^{\omega}(a^t)\right] < 2\overline{g}_i^{\omega} - \varepsilon - c^B.$$

Plugging this into (21), we obtain $U_i^{\omega,B}(h_{-i}^{T_b}) > 0$.