PENN INSTITUTE for ECONOMIC RESEARCH
UNIVERSITY of PENNSYLVANIA

The Ronald O. Perelman Center for Political
Science and Economics (PCPSE)
133 South 36th Street
Philadelphia, PA 19104-6297

pier@econ.upenn.edu
http://economics.sas.upenn.edu/pier

PIER Working Paper

19-008

Common Learning and Cooperation in Repeated Games

TAKUO SUGAYA
Stanford

YUICHI YAMAMOTO
University of Pennsylvania

April 26, 2019

Common Learning and Cooperation in Repeated Games[*]

Takuo Sugaya[†]and Yuichi Yamamoto[‡]

First Draft: December 10, 2012
This Version: April 26, 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 ω (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 an equilibrium even if the initial prior changes.[1] For a fixed discount factor δ, 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 expost 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 Appendix D 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, ω_1 and ω_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 ω_1 but punish at ω_2," or "reward at state ω_2 but punish at ω_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 ω. 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 ω_1 and plan to reward each other at ω_1, they will choose an action profile which yields high payoffs to both players at ω_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 ω 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 ω_1, but is correlated with a different signal $\tilde{z}_{-i}$ conditional on the state ω_2.

To deter player i's misreport using such ambiguous signals, state-contingent

[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.

punishments are helpful. A rough idea is that the opponent interprets her signal frequency f_{-i} *taking a state* ω *as given*, and decides whether to punish player i or not for that state ω. For example, suppose that the opponent's signal frequency f_{-i} is typical of the state ω_1, i.e., it is close to the true signal distribution at ω_1. Then *conditional on the state* ω_1, the opponent believes that player i's observation is also typical of the state ω_1 and hence i's inference is $\omega(i) = \omega_1$. On the other hand, *conditional on the state* ω_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 ω_1, but not conditional on ω_2. This suggests that the opponent should punish player i *only at the state* ω_2, by playing a continuation strategy which yields a low payoff to player i conditional on the state ω_2 but a high payoff conditional on ω_1. That is, the opponent should choose the plan "reward player i at ω_1 but punish at ω_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 ω 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.[3] The second property implies that any misreport is not profitable, regardless of player i's belief about the state ω. This in particular implies that player i's history in the previous blocks, which influences her belief about ω, is irrelevant to her incentive in the current block; her incentive is solely determined by her history within the current block. This

[3] 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.

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 ω. For example, suppose that player i's signal frequency f_i during the T period interval is typical of the state ω_1, so that her inference is $\omega(i) = \omega_1$. With such an observation f_i, *conditional on the state* ω_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* ω_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 ω_2, that is, reporting $\omega(i) = \omega_1$ must be a best reply for player i at ω_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.[4] 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 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

[4] 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.

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.

In Section 5, 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 ω. 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[5] (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 learning with strategic players.[6] Cripps, Ely, Mailath, and Samuelson (2008) shows that common learning occurs when players are not strategic, i.e., players observe private signals about the state each period, without taking actions. In the

[5] 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.

[6] 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.

follow-up paper (Cripps, Ely, Mailath, and Samuelson (2013)), they extend the analysis to the case in which signals are not i.i.d.. They argue:

> We are motivated by a desire to better understand the structure of equilibria in repeated games of incomplete information. [...] An understanding of common learning in this setting requires extending the setting of Cripps, Ely, Mailath, and Samuelson (2008) in two challenging directions: The signal distributions are intertemporally dependent and endogenous (being affected by the actions of the agents). [...] While we are ultimately interested in the signals that both exhibit intertemporal dependence and endogenously determined distributions, this paper focuses on intertemporal dependence, maintaining the assumption that the distributions are exogenously determined.

Our work addresses their concern above. Indeed, in our model, signal distributions are endogenously determined in equilibrium and intertemporally dependent. We find that players' strategic behavior has a substantial impact on the result: With non-strategic players, Cripps, Ely, Mailath, and Samuelson (2013) show that common learning occurs only when the signal distribution satisfies some restrictive condition. In contrast, we find that with strategic players, 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.

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 ω from a finite set Ω. Assume that players cannot observe the true state ω, and let $\mu \in \triangle\Omega$ denote their common prior over ω.[7] Throughout the paper, we assume that the game begins with symmetric information: Each player's initial belief about ω is equal to the prior μ. But it is straightforward to extend our analysis to the asymmetric-information case as in Fudenberg and Yamamoto (2011a).[8]

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 ω 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 ω 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 ω 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 ω and the action profile a is $g_i^\omega(a) = \sum_{z_i \in Z_i} \pi_i^\omega(z_i|a) u_i^\omega(a, z_i)$.[9] Let $g^\omega(a) = (g_i^\omega(a))_{i\in I}$ be the payoff vector given ω and a. As usual, we write $\pi^\omega(\alpha)$ and $g_i^\omega(\alpha)$ for the signal distribution

[7] 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.

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

[9] 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.

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 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 ω, 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 ω, 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 τ, 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 ω 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 ω. 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 ω. 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 ω 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 ω. 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 ω.[10] Throughout the paper, we assume that the set V^* is full dimensional:

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

[10] If there are only two players and our Condition 2 holds, the minimax payoff m_i^ω 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 μ) is at least v_i for all Nash equilibria. For simplicity, suppose that there are only two states, ω and $\tilde{\omega}$. (It is not difficult to extend the argument to the case with more than two states.) Fix an arbitrary Nash equilibrium σ. Let a^* be as in Condition 2, and let σ_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 ω^*.

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 σ_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.

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:

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

In an ex-post equilibrium, after every history h^t, player i's continuation play is a best reply regardless of the true state ω. 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 ω 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, ω, and $\tilde{\omega} \neq \omega$.

Intuitively, the statewise full rank implies that player i can statistically distinguish ω from $\tilde{\omega}$ through her private signal z_i, even if someone else unilaterally deviates from a^*.[11] 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

[11] 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.

of z_{-i} given that the true state is ω, 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' observations *conditional on the true state being* ω. 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 ω), 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 ω is given.

We impose the following condition:

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, 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 equal to the true 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 ω, 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 $C_i^{\omega}(a^*)$.[12]

The following is the main result of this paper:

[12] 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 ω and a. In Appendix D, we present an example with conditionally independent signals in which ex-post equilibria cannot approximate the Pareto-efficient frontier.

Proposition 1. *Under Conditions 1 through 3, 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.*

This proposition 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. The next proposition shows that in these equilibria, the state indeed becomes (approximate) common knowledge in the long run.

Proposition 2. *Suppose that players play an equilibrium constructed in the proof of Proposition 1. Then common learning occurs, i.e., the state becomes approximate common knowledge in the sense of Monderer and Samet (1989).*

For completeness, the formal definition of common learning is given in Appendix C. Cripps, Ely, Mailath, and Samuelson (2008) argue that common learning helps to facilitate players' coordination, and show that it occurs if player are not strategic (so there is no signaling effect) and their signals are i.i.d. over time. In the follow-up paper (Cripps, Ely, Mailath, and Samuelson (2013)), they also show that this result relies on the i.i.d. assumption; that is, they show that common learning does not occur in general, if players are not strategic and signals are not i.i.d. over time. Our Proposition 2 shows that this negative result overturns if players are strategic. That is, for generic signal distributions, there are equilibria in which players commonly learn the state and coordinate the play to approximate the Pareto-efficient outcome.

In the next section, we provide the proof of Proposition 1 for games with two players and two states. The proof for the general case can be found in Appendix B. The proof of Proposition 2 can be found in Appendix C.

4 Proof of Proposition 1 with $|I| = |\Omega| = 2$ and $|A_i| \geq |Z_i|$

In this section, we will prove Proposition 1 for two-player games with two states (so $I = \{1,2\}$ and $\Omega = \{\omega_1, \omega_2\}$). Our proof technique is valid even in the case with more players and more states, but it significantly complicates the notation. See Appendix B for more details. We also assume $|A_i| \geq |Z_i|$ for each i. This assumption greatly simplifies the structure of the "detailed report round" which will appear in our proof. Again, we will explain how to drop this assumption in Appendix B.

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 ω 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 ω, 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 ω, 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 ω is $\underline{v}_i^\omega$ if the opponent plans to punish her, and $\overline{v}_i^\omega$ if the opponent plans to reward her.

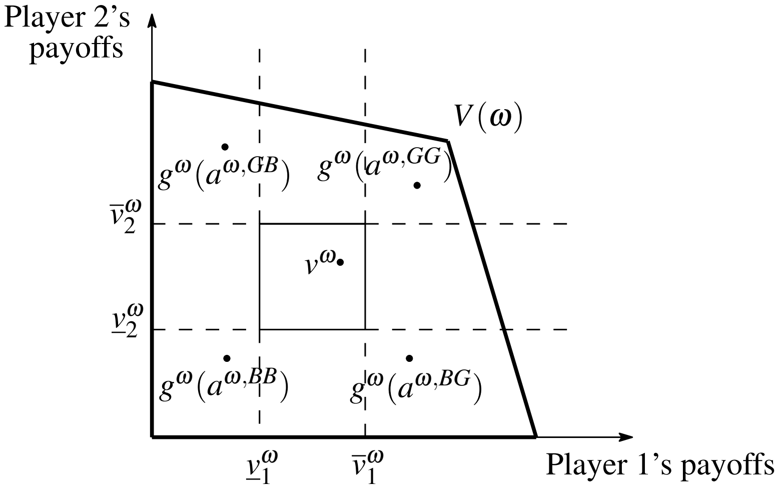

Figure 1: Actions $a^{\omega,GG}$, $a^{\omega,GB}$, $a^{\omega,BG}$, and $a^{\omega,BB}$

For each state ω, 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[13]

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

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 ith 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 ω,

$$\begin{aligned}
\max\{g_1^\omega(a^{\omega,GB}),g_1^\omega(a^{\omega,BB}),m_1^{\omega_1}\}&<\underline{v}_1^\omega-\varepsilon, && (1)\\
\max\{g_2^\omega(a^{\omega,BG}),g_2^\omega(a^{\omega,BB}),m_2^{\omega_2}\}&<\underline{v}_2^\omega-\varepsilon, && (2)\\
\min\{g_1^\omega(a^{\omega,GG}),g_1^\omega(a^{\omega,BG})\}&>\overline{v}_1^\omega+2\varepsilon, && (3)\\
\min\{g_2^\omega(a^{\omega,GG}),g_2^\omega(a^{\omega,GB})\}&>\overline{v}_2^\omega+2\varepsilon. && (4)
\end{aligned}$$

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

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

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

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

Note that (5) indeed holds for small ε, 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^ω. In the rest of the proof, we fix this parameter ε.

[13] For some payoff function, such action profiles a^{ω,x^ω} 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.

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.

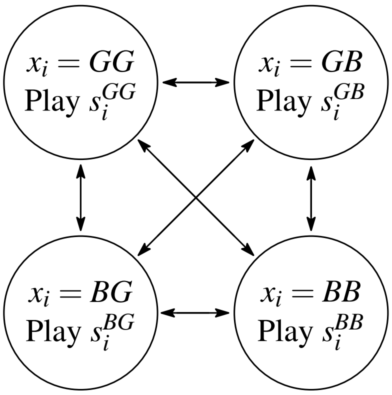

Figure 2: Automaton

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* ω_1. Similarly, the second component $x_i^{\omega_2}$ represents her plan *if the true state were* ω_2. For example, if player i's automaton state is $x_i = GB$, then during the current block, she rewards the opponent at state ω_1 and punishes the opponent at state ω_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 ω_1 but a low payoff conditional on ω_2.) Likewise, If $x_i = BG$, she punishes the opponent at state ω_1 but rewards at state ω_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 ω, the new plan for the state ω 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^{ω} 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 ρ_i^{ω}. 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 $\omega \neq \tilde{\omega}$.

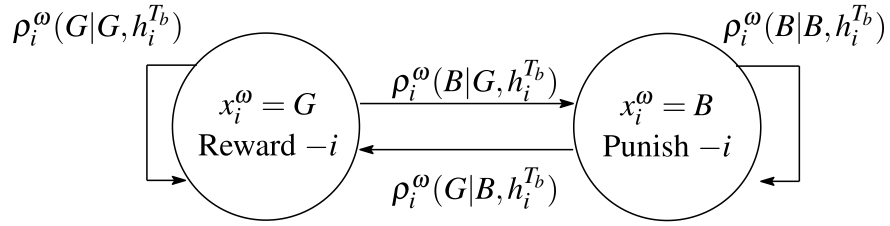
$\rho_i^{\omega}(G|G,h_i^{T_b})$
$\rho_i^{\omega}(B|B,h_i^{T_b})$
$x_i^{\omega} = G$
Reward $-i$
$\rho_i^{\omega}(B|G,h_i^{T_b})$
$x_i^{\omega} = B$
Punish $-i$
$\rho_i^{\omega}(G|B,h_i^{T_b})$

Figure 3: Transition of x_i^{ω}

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.

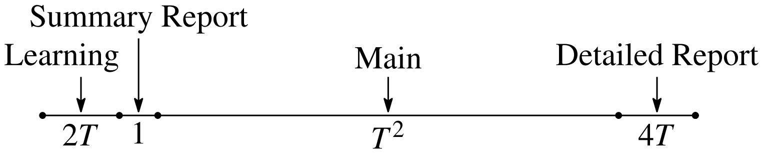

Figure 4: Structure of the block. Time goes from left to right.

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 ω. 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: ω_1, ω_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 ω_1, and $\omega(i) = \omega_2$ if it is close to the true distribution $\pi_i^{\omega_2}(a^*)$ at ω_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.[14] This "communication" allows 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."

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

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 ω 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 ω 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 ω. After that, players choose the action profile a^{ω,x^ω} until the main round ends, where $x^\omega = (x_1^\omega x_2^\omega)$ is the reported plan. (Recall that this action profile a^{ω,x^ω} is chosen as in Figure 1.) If someone (say player i) deviates from this action profile a^{ω,x^ω}, she will be minimaxed by $\underline{\alpha}^\omega(i)$. That is, players minimax the deviation, assuming that the summary report ω is the true state.

So if players report the same state ω 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^{ω,x^ω}, 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 ω 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 ω_1.

- If one player reports ω_1 while the other reports ω_2, then each player i reveals $x_i^{\omega(i)}$ in the first period of the main round, and then chooses the minimax 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. Since player 1 makes her inference $\omega(1)$ in the first T periods of the block and player 2 makes her inference $\omega(2)$ in the next T periods, player 1's belief about the opponent's inference $\omega(2)$ and player 1's belief about the opponent's belief about her inference $\omega(1)$ are not correlated. Indeed, the latter belief depends only on the history during the first T periods of the block, while the former depends on the history during the next T periods. This property is crucial in order to prove Lemma 3.

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 ω 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 ω 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 ω, the opponent's history up to the tth 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.

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 ω,

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

(ii) Regardless of the past history, the opponent's deviation cannot manipulate player i's inference with high probability: For each ω, $t \in \{0, \cdots, T-1\}$, 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 ω:

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. 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 ω. 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.)

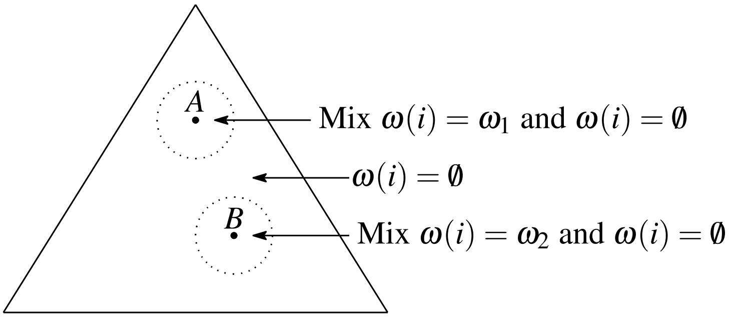
A
Mix $\omega(i) = \omega_1$ and $\omega(i) = \emptyset$
$\omega(i) = \emptyset$
B
Mix $\omega(i) = \omega_2$ and $\omega(i) = \emptyset$

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 $\pi_i^{\omega_2}(a^G)$.

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 ω, players play a^* for T periods, and player i's signal frequency during these periods is f_i. Then we have the following lemma:

Lemma 2. *Suppose that Condition 3 holds. Then there is $\overline{T}$ such that for any $T > \overline{T}$, ω, $\tilde{\omega} \neq \omega$, 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 ω, the opponent's observation is also unusual and not close to the ex-ante distribution $\pi^{\omega}_{-i}(a^*)$. 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),

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

Accordingly, in order to have $\left|\pi^{\omega}_{-i}(a^*) - f_{-i}\right| < \varepsilon$, 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 ω_1, the opponent's signal frequency f_{-i} is typical of the state ω_1 so that the opponent believes that player i's inference is $\omega(i) = \omega_1$, but (ii) conditional on the state ω_2, the opponent's signal frequency f_{-i} is typical of the state ω_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 ω_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 ω_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.

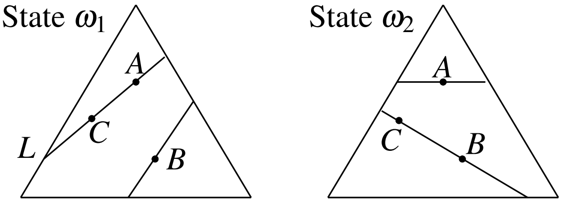

Figure 6: Each line in the left triangle is the set of signal frequencies f_i which give the same expectation about the opponent's signal frequency at the state ω_1. That is, $C_i^{\omega_1}(a^*)f_i = C_i^{\omega_1}(a^*)\tilde{f}_i$ for any f_i and $\tilde{f}_i$ on the same line. At the point $f_i = A$, player i believes that the opponent's observation is typical of the state ω_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 ω_2. At the point $f_i = B$, player i believes that the opponent's observation is typical of the state ω_2, and the same is true at the point $f_i = C$.

4.3 Transition Rule ρ_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 the continuation game must be $\overline{v}_i^{\omega_1}$ conditional on the state ω_1 (since player i is rewarded at ω_1) and $\underline{v}_i^{\omega_2}$ conditional on the state ω_2 (since player i is punished at ω_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 ω, 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 ω, 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 ω, then player i's payoff in the continuation game is exactly $\overline{v}_i^{\omega}$ conditional on the state ω. 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 ω, player i's payoff in the continuation game is $\underline{v}_i^{\omega}$ conditional on the state ω. 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 ω, i, $s_i^{T_b}$, t, h_i^t, and x with $x_{-i}^{\omega} = G$, and

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

for each ω, 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 ω 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 ω 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 ω is publicly revealed at the beginning of the game.

Second, the transfer U_i^ω is state-specific, that is, we use different transfer rules U_i^ω for different states ω. 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 ω is solely determined by the punishment plan x_{-i}^ω for the state ω (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^ω should depend on ω.

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 continuation 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 ω, 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 ω_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 ω_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 ω, 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 ω 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 ω. 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 ω, 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 ω 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 ω_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 ω_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 ω (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 ω ($|f_{-i} - \pi_{-i}^{\omega}(a^G)| < \varepsilon$), 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

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

    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 ω, 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 ω 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 ω) the opponent believes that player i's signal frequency is also typical of ω, 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 detailed 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 ω.

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 ω, 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 $\overline{v}_i^{\omega}$, 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 ω), she believes that the opponent's observation f_{-i} is not typical of ω, 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^ω 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 ω. Since we choose such a transfer rule U_i^ω for each state ω, reporting the null inference $\omega(i) = \emptyset$ is a best reply regardless of the state ω, 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 ω, the opponent's signal frequency f_{-i} is typical of ω almost surely given that player i has the correct inference $\omega(i) = \omega$, while f_{-i} is not typical of ω 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.

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.

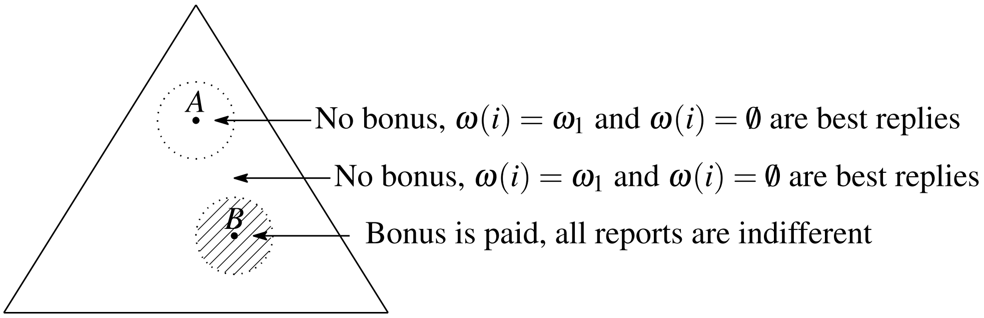
A
B
No bonus, $\omega(i)=\omega_1$ and $\omega(i)=\emptyset$ are best replies
No bonus, $\omega(i)=\omega_1$ and $\omega(i)=\emptyset$ are best replies
Bonus is paid, all reports are indifferent

Figure 7: Player i's incentive in the complete-information game with ω_1, assuming that $f_i = \hat{f}_i$.

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^{ω} 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 22 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 $\overline{v}_i^\omega$, 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.

5 Private Monitoring of Actions

In this section, we consider the case in which actions are not observable, so that players need to monitor the opponents' actions via noisy private signals. Since the distribution of the signals depends on the unknown state ω, the monitoring structure is unknown to players. So the model is now a repeated game with *private monitoring* and *unknown monitoring structure*.

5.1 Setup and Weak Ex-Post Equilibrium

We consider infinitely repeated games in which the set of players is denoted by $I = \{1, \cdots, N\}$. As in the case with observed actions, we assume that Nature chooses the state of the world ω from a finite set $\Omega = \{\omega_1, \cdots, \omega_o\}$. Assume that players cannot observe the true state ω, and let $\mu \in \triangle\Omega$ denote their common prior over ω. Each period, players move simultaneously, and player $i \in I$ chooses an action a_i from a finite set A_i and observes a private signal y_i from a finite set Y_i. Note that player i does not observe the opponents' actions a_{-i}. Let $A \equiv \times_{i\in I} A_i$ and $Y \equiv \times_{i\in I} Y_i$. The distribution of the signal profile $y \in Y$ depends on the state of the world ω and on the action profile $a \in A$, and is denoted by $\pi^\omega(\cdot|a) \in \triangle Y$. Let $\pi_i^\omega(\cdot|a)$ denote the marginal distribution of $y_i \in Y_i$ at state ω conditional on a, that is, $\pi_i^\omega(y_i|a) = \sum_{y_{-i}\in Y_{-i}} \pi^\omega(y|a)$. Player i's actual payoff is $u_i^\omega(a_i, y_i)$, so her expected payoff at state ω given an action profile a is $g_i^\omega(a) = \sum_{y_i \in Y_i} \pi_i^\omega(y_i|a) u_i^\omega(a_i, y_i)$. We write $\pi^\omega(\alpha)$ and $g_i^\omega(\alpha)$ for the signal distribution and 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 expected payoff when the opponents play a mixed action $\alpha_{-i} \in \times_{j\neq i}\triangle A_j$. Let $g^\omega(a)$ denote the vector of expected payoffs at state ω given an action profile a.

Note that the model studied in the earlier sections is a special case of the one above. To see this, let $Y_i = A \times Z_i$ and assume that $\pi_i^\omega(y_i|a) = 0$ for each i, ω, a, and y_i such that $y_i = (a', z_i)$ where $a' \neq a$. Then actions are perfectly observable (as y_i must be consistent with the action profile a), and players learn the true state ω from private signals z_i. Other examples that fit our model include:

- Secret price-cutting of Stigler (1964) with unknown demand function. I is the set of firms in an oligopoly market, a_i is firm i's price, and y_i is firm i's sales level. Often times, firms do not have precise information about the demand function, and hence do not know the distribution π of sales levels.

- Moral hazard with subjective evaluation and unknown evaluation distribution. I is the set of agents working in a joint project, a_i is agent i's effort level, and y_i is agent i's subjective evaluation about the opponents' performance. Often times, agents do not know how the opponents form their subjective evaluations, and hence do not know the distribution π.

In the infinitely repeated game, players have a common discount factor $\delta \in (0,1)$. Let (a_i^τ, y_i^τ) be player i's pure action and signal in period τ, and we denote player i's private history from period one to period $t \geq 1$ by $h_i^t = (a_i^\tau, y_i^\tau)_{\tau=1}^t$. 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. Also, we denote a profile of t-period histories by $h^t = (h_i^t)_{i \in I}$, and let 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 $S = \times_{i \in I} S_i$.

In this section, we use the following equilibrium concept:

Definition 2. A strategy profile s is a *weak ex-post equilibrium* if it is a Nash equilibrium in the infinitely repeated game where ω is common knowledge for each ω.

By the definition, in a weak ex-post equilibrium, player i's continuation play after any on-path history h_i is optimal regardless of the true state ω. On the other hand, player i's play after off-path history h_i may be suboptimal for some state ω. Therefore, a weak ex-post equilibrium is not necessarily a sequential equilibrium for some initial prior. However, if the full support assumption holds so that $\pi^\omega(y|a) > 0$ for all ω, a, and y, then given any initial prior, any payoff achievable by weak ex-post equilibria is also achievable by sequential equilibria. Indeed, given any weak ex-post equilibrium s and given any initial prior, there is a sequential equilibrium $\tilde{s}$ in which the play on the equilibrium path is identical with that of s (so s and $\tilde{s}$ yield the same equilibrium payoffs given any state ω). See Sekiguchi (1997) for a proof.

5.2 Identifiability Conditions

Now we will state a set of assumptions under which the folk theorem holds. We will use the following notation. Let $\pi^{\omega}(a_{-i},y_{-i}|a_i,\alpha_{-i},y_i)$ be the probability of (a_{-i},y_{-i}) given that the profile (a_i,α_{-i}) is chosen and player i observes y_i. Let $\tilde{\pi}_i^{\omega}(\alpha)$ be the distribution of (a_i,y_i) when players play α at state ω. Likewise, let $\tilde{\pi}_{-i}^{\omega}(\alpha)$ be the distribution of (a_{-i},y_{-i}) when players play α at state ω. Let $\Pi_i^{(j,\omega)}(\alpha)$ be the affine hull of $\tilde{\pi}_i^{\omega}(a_j,\alpha_{-j})$ for all a_j. Roughly, $\Pi_i^{(j,\omega)}(\alpha)$ includes the set of all possible distributions of (a_i,y_i) when the true state is ω, players $-j$ choose α, but player j may deviate from α by taking an arbitrary action a_j. Likewise, let $\Pi_{-i}^{(j,\omega)}(\alpha)$ be the affine hull of $\tilde{\pi}_{-i}^{\omega}(a_j,\alpha_{-j})$ for all a_j. Let $C_i^{\omega}(\alpha)$ be the matrix which maps player i's private observation $f_i \in \triangle(A_i \times Y_i)$ to her estimate about the opponents' private observation $f_{-i} \in \triangle(A_{-i} \times Y_{-i})$ conditional on ω and α. Note that f_i here is a frequency of (a_i,y_i), rather than a frequency of y_i.

When players know the monitoring structure (i.e., $|\Omega| = 1$), Sugaya (2019) shows that the folk theorem holds in repeated games with private monitoring under a set of assumptions on the signal structure π. In this paper, we impose the same assumptions on the signal structure π^{ω} for each state ω. The following is the assumption for games with two players. (We will not state the assumptions for games with more than two players, as they involve a complex and lengthy notation. See Sugaya (2019) for details.)

Condition 4. (Regular Environment When $|I| = 2$) For each ω, the following conditions hold:

(i) $\pi^{\omega}(y|a) > 0$ for each a and y.

(ii) For each i and a_i, the marginal distributions $\{\pi_i^{\omega}(a)|a_{-i} \in A_{-i}\}$ are linearly independent .

(iii) For each i, a, and $\tilde{a}_{-i} \neq a_{-i}$, we have $\pi_i^\omega(y_i|a) \neq \pi_i^\omega(y_i|a_i,\tilde{a}_{-i})$ for all y_i.

(iv) For each i, there is α_{-i} such that for each (a_i,y_i) and $(\tilde{a}_i,\tilde{y}_i)$ with $(a_i,y_i) \neq (\tilde{a}_i,\tilde{y}_i)$, there is (a_{-i},y_{-i}) such that $\pi^\omega(a_{-i},y_{-i}|a_i,\alpha_{-i},y_i) \neq \pi^\omega(a_{-i},y_{-i}|\tilde{a}_i,\alpha_{-i},\tilde{y}_i)$.

Clause (i) is the full support assumption, which requires that each signal profile y can happen with positive probability given any state ω and given any action profile a. Clause (ii) is a version of individual full-rank of Fudenberg, Levine, and Maskin (1994), which requires that player i can statistically distinguish the opponent's actions through her private signal. Clause (iii) ensures that different actions of the opponent induce different probability on each signal y_i. Clause (iv) requires that when the opponent chooses a particular mixed action α_{-i}, different histories of player i induce different beliefs about the opponent's history. Note that Condition 4 holds for generic choice of π, if $|Y_i| \geq |A_{-i}|$ for each i.

The next condition extends the statewise full-rank condition and correlated learning condition for games with observable actions to the current setup. When monitoring is imperfect, player i's deviation is not directly observable, and she may secretly deviate to manipulate the opponents' state learning and/or the opponents' belief about player i's belief. The following condition is useful in order to deter such a manipulation.

Condition 5. (Statewise Full Rank and Correlated Learning) For each i, ω, and $\tilde{\omega} \neq \omega$, there is player i's pure action a_i and the opponents' (possibly mixed) action α_{-i} which satisfy the following conditions:

(i) $\Pi_i^{(j,\omega)}(a_i,\alpha_{-i}) \cap \Pi_i^{(l,\tilde{\omega})}(a_i,\alpha_{-i}) = \emptyset$ for each $j \neq i$ and $l \neq i$ (possibly $j = l$).

(ii) If $f_i \in \bigcup_{j\neq i} \Pi_i^{(j,\tilde{\omega})}(a_i,\alpha_{-i})$ then $C_i^\omega(a_i,\alpha_{-i})f_i \notin \Pi_{-i}^{(i,\omega)}(a_i,\alpha_{-i})$. Likewise, if $f_i \in \bigcup_{j\neq i} \Pi_i^{(j,\omega)}(a_i,\alpha_{-i})$ then $C_i^{\tilde{\omega}}(a_i,\alpha_{-i})f_i \notin \Pi_{-i}^{(i,\tilde{\omega})}(a_i,\alpha_{-i})$.

Condition 5(i) is the statewise full-rank condition, which generalizes Condition 2 to the private-monitoring case. To see its implication, suppose that players play the action profile (a_i,α_{-i}) for T periods and that player i tries to distinguish ω from $\tilde{\omega}$ using private signals during this T-period interval. Note that when Condition 5(i) holds, we have

$$\left(\bigcup_{j\neq i} \Pi_i^{(j,\omega)}(a_i,\alpha_{-i})\right) \cap \left(\bigcup_{j\neq i} \Pi_i^{(j,\tilde{\omega})}(a_i,\alpha_{-i})\right) = \emptyset.$$

This implies that player i can distinguish ω from $\tilde{\omega}$ even if someone else secretly and unilaterally deviates from (a_i, α_{-i}). In other words, the opponents' deviation cannot manipulate player i's state learning. This condition is similar to the statewise full-rank condition of Yamamoto (2014).

Condition 5(ii) generalizes the correlated learning condition (Condition 3) to the private-monitoring case. Intuitively, it requires that signals are correlated across players so that if player i observes an "unusual" signal frequency, then the opponent also observes an "unusual" signal frequency. To be more precise, recall that Condition 3 requires that if player i's signal frequency is the true distribution at state $\tilde{\omega}$ (which is unusual at state $\omega \neq \tilde{\omega}$), then she believes that conditional on the state ω, the opponent's signal frequency is not close to the true distribution at ω (which is also unusual at state ω). Condition 5(ii) strengthens this condition, and requires that if player i's signal frequency is the true distribution at state $\tilde{\omega}$ *or the ones induced by the opponent's unilateral deviation*, then she believes that the opponent's signal frequency is not close to the true distribution at ω *or the ones induced by player i's unilateral deviation*.

So Condition 5(ii) is stronger than Condition 3 in two respects. First, Condition 5(ii) imposes a restriction on player i's posterior belief when her signal frequency is the ones induced by the opponent's deviation (i.e., $f_i \in \bigcup_{j \neq i} \Pi_i^{(j,\tilde{\omega})}(a_i, \alpha_{-i})$). This implies that the opponent's secret deviation at state $\tilde{\omega}$ cannot manipulate whether player i's posterior falls into the set $\Pi_{-i}^{(i,\omega)}(a_i, \alpha_{-i})$.

Second, Condition 5(ii) requires that player i's posterior be different from not only the true distribution $\pi_{-i}^{\omega}(a_i, \alpha_{-i})$ at state ω, but also the ones induced by player i's deviation. This implies that player i cannot pretend to have an "unusual" signal frequency $f_i \in \bigcup_{j \neq i} \Pi_i^{(j,\tilde{\omega})}(a_i, \alpha_{-i})$ by deviating from (a_i, α_{-i}) secretly. This property is useful when we bound player i's gain by deviating during her own learning round.

The following proposition shows that Condition 5 is generically satisfied if each player's signal space is large enough. The proof can be found in Appendix A.

Proposition 3. *Suppose that $|Y_i| \geq 2|A_j| - 1$ and $|A_{-i}| \times |Y_{-i}| \geq |A_i| + |A_j| - 1$ for each i and $j \neq i$. Then Condition 5 is satisfied for generic choice of π.*

In order to obtain the folk theorem under private monitoring, we need one more condition. Recall that, in the perfect-monitoring game, we have constructed an equilibrium in which each player makes a summary inference about the state ω in the learning round, and then reports it in the summary report round. To prove the folk theorem for games with private monitoring, we will construct an equilibrium with a similar structure, that is, in our equilibrium, there is an summary report round in which players report their private inferences. However, it is a priori unclear whether communication via actions can be meaningful when monitoring is imperfect. A major problem is that each player needs to make an inference about the message of the opponents based on noisy, ambiguous signals. Also, since signals are private, a player can deviate in the continuation game by pretending as if she received a wrong message in the summary report round. It turns out that the following condition is enough to avoid these difficulties:

Condition 6. For each i and $(\omega,\tilde{\omega})$ with $\omega \neq \tilde{\omega}$, there is player i's pure action a_i and two (possible mixed) actions of the opponents, m_{-i}^{ω} and $m_{-i}^{\tilde{\omega}}$, such that for each f_i, there is $\omega^* \in \{\omega,\tilde{\omega}\}$ such that $C_i^{\omega^*}(a_i,m_{-i}^{\omega^*})f_i \notin \Pi_{-i}^{(i,\omega^*)}(a_i,m_{-i}^{\omega^*})$.

To interpret this condition, suppose that there are only two players and two states, ω and $\tilde{\omega}$. In the proof, we will construct an equilibrium in which the opponent reports her inference $\omega(-i)$ to player i via actions; specifically, the opponent chooses m_{-i}^{ω} for T periods when her inference is $\omega(-i)=\omega$, and she chooses $m_{-i}^{\tilde{\omega}}$ for T periods when $\omega(-i)=\tilde{\omega}$. Player i needs to distinguish these two cases based on her signal frequency during this T-period interval. Let f_i denote this signal frequency. Condition 6 requires that regardless of the realized signal frequency f_i, player i believes that in at least one of the two cases above, (conditional on that the state matches the opponent's inference) the opponent's signal frequency is "unusual" in the sense that it is different from the ex-ante distribution or the ones induced by player i's deviation.

The next proposition shows that Condition 6 is generically satisfied if

$$|Y_i| \geq 2(|A_i| + \max\{0, |Y_i| - |A_{-i}| \times |Y_{-i}|\}) - 1 \qquad (13)$$

for each i. Roughly speaking, this rank condition (13) requires that the numbers of private signals be similar across all players. For example, consider the extreme case in which the signal space is identical for all players, i.e., $|Y_i| = |Y_j|$ for each i and j. (Note that this assumption is common in the mechanism design literature, see Crémer and Mclean (1988), for example.) Then we have $|Y_i| - |A_{-i}| \times |Y_{-i}| < 0$, and hence (13) reduces to $|Y_i| \geq 2|A_i| - 1$. On the other hand, if player i's signal space is much larger than the others' so that $|Y_i| > 2|A_{-i}| \times |Y_{-i}|$, then it is easy to check that (13) never holds. The proof of the proposition can be found in Appendix A.

Proposition 4. *Suppose that (13) holds for each i. Then Condition 6 is satisfied for generic choice of π.*

Now we are ready to state the folk theorem for games with private monitoring:

Proposition 5. *Suppose that Conditions 1, 5, and 6 holds. Suppose also that the assumption in Sugaya (2019) is satisfied for each ω (When $|I| = 2$, it is precisely Condition 4). 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 a weak ex-post equilibrium with payoff v.*

6 Proof of Proposition 5 with $|I| = |\Omega| = 2$

6.1 Automaton with State-Specific Punishments

Fix an arbitrary payoff vector $v \in \text{int}V^*$. To prove the proposition, we need to construct a weak ex-post equilibrium with payoff v. In what follows, we will briefly describe how to construct such an equilibrium for the case in which there are only two players and two states, ω_1 and ω_2. Take $\underline{v}_i^\omega$ and $\overline{v}_i^\omega$ as in the perfect-monitoring case, that is, for each ω, take $\underline{v}_i^\omega$ and $\overline{v}_i^\omega$ for each i so that $m_i^\omega < \underline{v}_i^\omega < v_i^\omega < \overline{v}_i^\omega$ for each i and that the product set $\times_{i \in I}[\underline{v}_i^\omega, \overline{v}_i^\omega]$ is in the interior of the set $V(\omega)$.

As in the perfect-monitoring case, our equilibrium strategy is an automaton over blocks with length T_b. (See Figure 2.) Each player i's automaton state is denoted by $x_i = (x_i^{\omega_1}, x_i^{\omega_2}) \in \{G, B\}^2$, and it can be interpreted as her state-contingent plan about whether to reward or punish the opponent. That is, player i plans to reward the opponent at state ω if $x_i^\omega = G$, and punish the opponent if $x_i^\omega = B$. Pick K such that $K \geq \log_2 |A_i||Y_i|$ for each i, and then let $T_b = 4T + T^3 + 8KT^2 + 8LK^2T^2$. The parameters L and T will be specified later.

Let $s_i^{x_i}$ denote the block strategy induced by the automaton state x_i, and ρ_i denote the transition rule. Our goal is to find $s_i^{x_i}$ and ρ_i which satisfy the promise-keeping condition (7) and (8), and the incentive compatibility condition which is now simplified to:

$$\overline{v}_i^\omega \geq (1-\delta^{T_b}) \sum_{t=1}^{T_b} \delta^{t-1} E[g_i^\omega(a^t)|\omega, s_i^{T_b}, s_{-i}^{x_{-i}}] + \delta^{T_b} \left\{ \overline{v}_i^\omega - E[\rho_{-i}^\omega(B|G, h_{-i}^{T_b})|\omega, s_i^{T_b}, s_{-i}^{x_{-i}}](\overline{v}_i^\omega - \underline{v}_i^\omega) \right\} \tag{14}$$

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

$$\underline{v}_i^\omega = (1-\delta^{T_b}) \sum_{t=1}^{T_b} \delta^{t-1} E[g_i^\omega(a^t)|\omega, s_i^{T_b}, s_{-i}^{x_{-i}}] + \delta^{T_b} \left\{ \underline{v}_i^\omega + E[\rho_{-i}^\omega(G|B, h_{-i}^{T_b})|\omega, s_i^{T_b}, s_{-i}^{x_{-i}}](\overline{v}_i^\omega - \underline{v}_i^\omega) \right\} \tag{15}$$

for each ω, i, $s_i^{T_b}$, and x_{-i} with $x_{-i}^\omega = B$. Unlike (9) and (10), the incentive conditions above do not require sequential rationality of $s_i^{x_i}$; they require only that the strategy $s_i^{x_i}$ be a best reply in the normal-form game. However, this difference is not essential. As shown by Sekiguchi (1997), Nash equilibria and sequential equilibria are payoff-equivalent under the full support assumption. Accordingly, if (7), (8), (14), and (15) hold, then any payoff in the set $\times_{\omega\in\Omega} \times_{i\in I} [\underline{v}_i^\omega, \overline{v}_i^\omega]$ can be achievable by sequential equilibria, just as in the perfect-monitoring case.

6.2 Block Strategy $s_i^{x_i}$

6.2.1 Brief Description

As described in Figure 2, in our equilibrium, player i's play in each block is solely determined by the automaton state x_i; she will play a block strategy $s_i^{x_i}$ if the current automaton state is x_i. In what follows, we will explain how to construct this block strategy $s_i^{x_i}$.

As in the perfect-monitoring case, each block is divided into four parts: the learning round, the summary report round, the main round, and the detailed report round. Specifically:

Learning Round: The first T periods of the block are player 1's learning round, and the next T periods are player 2's learning round. In player i's learning round, players play the profile (a_i, α_{-i}) which satisfies Condition 5. Then, based on the realized signal frequency f_i, player i makes an inference $\omega(i)$ about the true state. Roughly, she chooses $\omega(i) = \omega_1$ if f_i is close to $\Pi_i^{(j,\omega_1)}(a_i, \alpha_{-i})$, and chooses $\omega(i) = \omega_2$ if f_i is close to $\Pi_i^{(j,\omega_2)}(a_i, \alpha_{-i})$. Otherwise, she chooses $\omega(i) = \emptyset$. Note that the opponent cannot manipulate this inference $\omega(i)$ much; e.g., if the true state is ω_1, then regardless of the opponent's action, player i's signal frequency will be in the neighborhood of $\Pi_i^{(-i,\omega_1)}(a_i, \alpha_{-i})$, in which case player i makes the inference $\omega(i) = \omega_1$.

Summary Report Round: The next $2T$ periods are the summary report round. The first T periods are player 1's summary report round, in which player 1 reports her summary inference $\omega(1)$ about the true state. The next T periods are player 2's summary report round, in which player 2 reports her inference $\omega(2)$. Choose a_i, $m_{-i}^{\omega_1}$, and $m_{-i}^{\omega_2}$ as in Condition 6. In player $-i$'s summary report round, player i chooses a_i every period, while player $-i$'s chooses either $m_{-i}^{\omega_1}$ or $m_{-i}^{\omega_2}$ depending on her inference $\omega(-i)$. Specifically, she chooses $m_{-i}^{\omega_1}$ if her inference is $\omega(-i) = \omega_1$, and she chooses $m_{-i}^{\omega_2}$ if her inference is $\omega(-i) = \omega_2$. If $\omega(-i) = \emptyset$, she randomly selects $m_{-i}^{\omega_1}$ or $m_{-i}^{\omega_2}$ and chooses it for T periods. After T periods, player i makes an inference $\hat{\omega}(-i)$ about the opponent's report, using the private observation f_i. Specifically:

- If player i does not deviate from a_i and if $C_i^{\omega_1}(a_i, m_{-i}^{\omega_1})f_i$ is ε-close to $\Pi_{-i}^{(i,\omega_1)}(a_i, m_{-i}^{\omega_1})$, then let $\hat{\omega}(-i) = \omega_1$.[15]
- If player i does not deviate from a_i and if $C_i^{\omega_2}(a_i, m_{-i}^{\omega_2})f_i$ is ε-close to $\Pi_{-i}^{(i,\omega_2)}(a_i, m_{-i}^{\omega_2})$, then let $\hat{\omega}(-i) = \omega_2$.
- For all other cases, choose $\hat{\omega}(-i) = \omega_1$ or $\hat{\omega}(-i) = \omega_2$ randomly.

In words, player i chooses $\hat{\omega}(-i) = \omega$ if she believes that (conditional on that the true state is ω and the opponent reported $\omega(-i) = \omega$) the opponent's observation f_{-i} is typical of ω. The above inference rule is well-defined, because Condition 6 ensures that the events stated in the first two bullet points never occur at the same time. So if player i chooses $\hat{\omega}(-i) = \tilde{\omega}$ (which is unusual given the state ω and the opponent's report $\omega(-i) = \omega$), then she believes that almost surely, the opponent's

[15] Here, a distribution $f_{-i} \in \triangle(A_{-i} \times Y_{-i})$ is *ε-close to* a set $B \subseteq \triangle(A_{-i} \times Y_{-i})$ if there is $\tilde{f}_{-i} \in B$ such that $|f_{-i} - \tilde{f}_{-i}| \leq \varepsilon$.

observation is also unusual in the sense that it is not close to $\Pi_{-i}^{(i,\omega)}(a_i, m_{-i}^{\omega})$. This property plays a crucial role when we prove Lemma 11 later.

Main Round: The next T^3 periods are the main round. This round is much longer than the other rounds, so the average payoff in the block is approximated by the one in the main round, as in the perfect-monitoring case. During the main round, each player i plays one of the four strategies: $s_i^{\omega_1,G}$, $s_i^{\omega_1,B}$, $s_i^{\omega_2,G}$, or $s_i^{\omega_2,B}$. The formal definition of these strategies will be given shortly, but roughly speaking, the strategy $s_i^{\omega_1,G}$ yields a high payoff to the opponent conditional on the state ω_1, so it can be used to reward the opponent at ω_1. Similarly, $s_i^{\omega_2,G}$ yields a high payoff at ω_2, so player i uses it when she plans to reward the opponent at ω_2. On the other hand, $s_i^{\omega_1,B}$ and $s_i^{\omega_2,B}$ are used for state-specific punishments; $s_i^{\omega_1,B}$ yields a low payoff at ω_1, while $s_i^{\omega_2,B}$ yields a low payoff at ω_2. At the beginning of the main round, each player i chooses one of these strategies, depending on her current plan $x_i = (x_i^{\omega_1}, x_i^{\omega_2})$ and the history in the learning and summary report rounds. Specifically,

- If player i's inference is $\omega(i) = \omega_1$, then play $s_i^{\omega_1, x_i^{\omega_1}}$, where $x_i^{\omega_1}$ is player i's current plan for the state ω_1.
- Similarly, if her inference is $\omega(i) = \omega_2$, then play $s_i^{\omega_2, x_i^{\omega_2}}$.
- If she has the null inference $\omega(i) = \emptyset$, then...
  - If $\hat{\omega}(-i) = \omega_1$, play $s_i^{\omega_1, x_i^{\omega_1}}$.
  - If $\hat{\omega}(-i) = \omega_2$, play $s_i^{\omega_2, x_i^{\omega_2}}$.

That is, if player i could learn the state in the learning round (i.e., $\omega(i) = \omega$), then she ignores the opponent's report in the summary report round and chooses the strategy $s_i^{\omega,G}$ or $s_i^{\omega,B}$, depending on her current plan x_i^{ω} for the state ω. If she has the null inference $\omega(i) = \emptyset$, then she chooses a strategy depending on the opponent's report in the summary report round, in order to coordinate the continuation play.

Detailed Report Round: The last $8KT^2 + 8LK^2T^2$ periods of the block are the report round, in which each player i reports her private histories in her own learning round and the opponent's summary report round. (She does not report the history during the opponent's learning round or her own summary report round.) The way she reports is similar to that of Sugaya (2019) and will be explained later.

The description of the block strategy $s_i^{x_i}$ above is informal, as we have not specified the inference rule $\omega(i)$, the strategies for the main round, or the strategies for the detailed report round. In what follows, we will explain how to choose them.

6.2.2 Inference Rule $\omega(i)$

We will explain how each player i forms the inference $\omega(i)$ from signals 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. As in the perfect-monitoring case, 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 $P(\cdot|h_i^T)$.

Given an inference rule P, let $\hat{P}(\cdot|\omega, \alpha^1, \cdots, \alpha^T)$ denote the conditional distribution of $\omega(i)$ given that the true state is ω and players play the action sequence $(\alpha^1, \cdots, \alpha^T)$ during player i's learning round. Likewise, for each $t \in \{0, \cdots, T-1\}$ and h_{-i}^t, let $\hat{P}(\cdot|\omega, h^t, \alpha^{t+1}, \cdots, \alpha^T,)$ be the conditional distribution of $\omega(i)$ given that the true state is ω, the history profile up to the tth period is h^t, and players play $(\alpha^{t+1}, \cdots, \alpha^T)$ thereafter. Given h_i^T, let $f_i(h_i^T) \in \triangle(A_i \times Y_i)$ denote player i's outcome frequency induced by h_i^T.

The following lemma shows that there is an inference rule P which satisfies some useful properties.

Lemma 6. *Suppose that Condition 5 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_i, α_{-i}), the inference $\omega(i)$ coincides with the true state almost surely: For each ω,

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

(ii) *Regardless of the past history, the opponent's deviation cannot manipulate player i's inference almost surely: For each ω, $t \in \{0,\cdots,T-1\}$, 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 $\tau \geq t+1$,*

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

(iii) *Suppose that player i does not deviate from (a_i,α_{-i}). Then she has the inference $\omega(i)=\omega$, only if her observation is close to the set $\Pi_i^{(j,\omega)}(a_i,\alpha_{-i})$: For all $h_i^T = (a_i^t,y_i^t)_{t=1}^T$ such that $a_i^t = a_i$ for all t and such that $P(\omega(i)=\omega|h_i^T)>0$, $f_i(h_i^T)$ is ε-close to the set $\Pi_i^{(j,\omega)}(a_i,\alpha_{-i})$.*

Clauses (i) and (ii) are exactly the same as those in Lemma 1, which ensure that state learning is almost perfect and robust to the opponent's deviation. Clause (iii) is slightly different from that in Lemma 1; now player i makes the inference $\omega(i)=\omega$ not only when her signal frequency is close to the true distribution $\pi_i^\omega(a_i,\alpha_{-i})$ at state ω, but also when her signal frequency is close to the ones induced by the opponent's deviation. This change is needed in order to obtain clause (ii) under imperfect monitoring. Indeed, the opponent can generate any signal frequency f_i in the set $\Pi_i^{(j,\omega)}(a_i,\alpha_{-i})$ by secretly deviating from α_{-i}, so for the inference $\omega(i)$ to be non-manipulable by the opponent, player i must make the same inference for all f_i in this set.

Proof. The proof is very similar to that of Lemma 1; we define the base score, the random score, and the final score, and the inference $\omega(i)$ is determined by these scores. Here we will illustrate only the definition of the base score when actions are not observable. The rest of the proof is exactly the same as that of Lemma 1.

Let $f_i \in \triangle(A_i \times Y_i)$ denote player i's observation during her learning round. Then we compute a *base score* $q_i^{\text{base}} \in \mathbb{R}^{|Y_i|}$ using the following formula:

$$q_i^{\text{base}} = Q_i f_i$$

where Q_i is a $|A_i \times Y_i| \times |Y_i|$ matrix, so it is a linear operator which maps an observation f_i to a score vector q_i^{base}. (Here, both f_i and q_i^{base} are column vectors.) From Condition 5(i), there is a matrix Q_i and $|Y_i|$-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 a_{-i},

$$Q_i \pi_i^\omega(a_i,a_{-i}) = \begin{cases} q_i^{\omega_1} & \text{if } \omega=\omega_1 \\ q_i^{\omega_2} & \text{if } \omega=\omega_2 \end{cases}.$$

That is, we choose a matrix Q_i so that the opponent cannot influence the expected value of the base score, as in the perfect-monitoring case. The only difference is that the matrix Q_i cannot depend on a_{-i}, as actions are not observable. *Q.E.D.*

6.2.3 Strategy s_i^{ω,x_i^ω} for the Main Round

As noted earlier, during the main round, each player i plays one of the four strategies, $s_i^{\omega_1,G}$, $s_i^{\omega_1,B}$, $s_i^{\omega_2,G}$, or $s_i^{\omega_2,B}$. The following lemma shows how to choose these strategies. Roughly speaking, these strategies are "block strategies" in Sugaya (2019), and they are chosen so that (with some appropriate transfer functions) the target payoffs are achieved when the state ω is common knowledge. The lemma directly follows from the main theorem of Sugaya (2019), and hence we omit the proof.

Lemma 7. *Suppose that $|I| = 2$ and Condition 4 holds. For each ω, there is $\underline{C} > 0$, $\overline{C} > 0$, and $\overline{T}$ such that for each $T > \overline{T}$, there is $\overline{\delta} \in (0,1)$ such that for each $\delta \in (\overline{\delta},1)$ and i, there are T^3-period strategies $s_i^{\omega,G}$ and $s_i^{\omega,B}$ and transfers $\underline{U}_i^{\omega,G} : H_{-i}^{T^3} \to \boldsymbol{R}$ and $\underline{U}_i^{\omega,B} : H_{-i}^{T^3} \to \boldsymbol{R}$ such that the following conditions hold for each i:*

(i) In the T^3-period complete-information transfer game with $(\omega, \underline{U}_i^{\omega,G})$, both $s_i^{\omega,G}$ and $s_i^{\omega,B}$ are best replies against $s_{-i}^{\omega,G}$ and yield $\overline{v}_i^{\omega}$.

(ii) In the T^3-period complete-information transfer game with $(\omega, \underline{U}_i^{\omega,B})$, both $s_i^{\omega,G}$ and $s_i^{\omega,B}$ are best replies against $s_{-i}^{\omega,B}$ and yield $\underline{v}_i^{\omega}$.

(iii) $-\overline{C}T^3 < \underline{U}_i^{\omega,G}(h_{-i}^{T^3}) < -\underline{C}T^3$ and $\underline{C}T^3 < \underline{U}_i^{\omega,B}(h_{-i}^{T^3}) < \overline{C}T^3$ for each $h_{-i}^{T^3}$.

To interpret this lemma in our context, suppose that the true state is ω. Suppose also that both players could learn the state in the learning round (i.e., $\omega(i) = \omega$ for each i), so that each player i plays either $s_i^{\omega,G}$ or $s_i^{\omega,B}$ during the main round, depending on her current plan x_i^{ω}. The lemma above ensures that playing these strategies $s_i^{\omega,G}$ and $s_i^{\omega,B}$ during the main round is indeed incentive compatible, and that each player i's payoff is solely determined by the opponent's plan x_{-i}^{ω} about whether to reward or punish player i. Specifically, clause (i) considers the case in which the opponent chooses $s_{-i}^{\omega,G}$ (i.e., the opponent plans to reward player i at state ω), and it asserts that if player i can receive an additional transfer $\underline{U}_i^{\omega,G}(h_{-i}^{T^3})$

after the main round, then both $s_i^{\omega,G}$ and $s_i^{\omega,B}$ are best replies for player i, and yield the payoff $\overline{v}_i^{\omega}$. Likewise, clause (ii) considers the case in which the opponent plans to punish player i at state ω, and it asserts that if there is an additional transfer $\underline{U}_i^{\omega,B}(h_{-i}^{T^3})$, then both $s_i^{\omega,G}$ and $s_i^{\omega,B}$ are best replies for player i, and yield the payoff $\underline{v}_i^{\omega}$.

6.2.4 Strategy for the Detailed Report Round

To complete the definition of the block strategies s_i^{GG}, s_i^{GB}, s_i^{BG}, and s_i^{BB}, we have to specify the play during the detailed report round, in which each player i reports her observation (a_i^t, y_i^t) during her own learning round and the opponent's summary report round. With an abuse of notation, let $h_i^{2T} = (a_i^t, y_i^t)_{t=1}^{2T}$ denote the information player i should report, that is, $(a_i^t, y_i^t)_{t=1}^{T}$ denotes player i's history during her own learning round, and $(a_i^t, y_i^t)_{t=T+1}^{2T}$ denotes her history during the opponent's summary report round. Let h_{-i}^{2T} denote the opponent's history during these rounds. Note that h_{-i}^{2T} is informative about player i's history h_i^{2T}, as signals are correlated across players.

The detailed report round is divided into four parts: Player 1's detailed report round for state ω_1, player 1's detailed report round for state ω_2, player 2's detailed report round for state ω_1, and player 2's detailed report round for state ω_2. In player i's detailed report round for state ω, she reports her private information h_i^{2T}, using a strategy which effectively transmits this information to the opponent conditional on the state ω. Since there is a report round for each state ω, she can effectively transmit the information h_i^{2T} regardless of the true state ω. For notational convenience, let $R = 2KT^2 + 2LK^2T^2$ denote the length of each round.

Specifically, player i's *reporting strategy* for state ω is a mapping $\sigma_i^{\text{report},\omega} : H_i^{2T} \to S_i^R$. That is, given private information h_i^{2T}, player i plays a pure strategy $\sigma_i^{\text{report},\omega}(h_i^{2T})$ during her detailed report round for state ω. Meanwhile, the opponent mixes all actions each period and makes an inference about player i's detailed report, using her observation h_{-i}^R. Formally, the opponent's inference rule is given by a mapping $P^{\omega} : H_{-i}^R \to \triangle H_i^{2T}$.

The following lemma shows that there is a reporting strategy $\sigma_i^{\text{report},\omega}$ and an inference rule P^{ω} such that the opponent's inference matches player i's detailed report almost surely, and such that player i has a strict incentive to report truthfully if she receives an additional transfer $U_i^{\text{report},\omega}(h_{-i}^{2T},h_{-i}^R)$. Here the transfer $U_i^{\text{report},\omega}$ depends not only on the opponent's history h_{-i}^R during the detailed report round, but also on the history h_{-i}^{2T} during the time in which player i observes the history h_i^{2T}. The idea is very similar to the one for the perfect monitoring case; the opponent's observation h_{-i}^{2T} is informative about player i's observation h_i^{2T}, and hence useful to detect player i's misreport. Let $\Pr(h_{-i}^{2T}|\omega,h_i^{2T},\alpha_{-i},m_{-i}^{\omega})$ denote the probability of h_{-i}^{2T} conditional on that the true state is ω, player i's history is h_i^{2T}, and the opponent plays α_{-i} during player i's learning round and m_{-i}^{ω} during her own summary report round. Also, let $\Pr(h_{-i}^R|\omega,s_i^R)$ denote the probability of the opponent's history during the detailed report round being h_{-i}^R conditional on that the true state is ω, player i plays s_i^R during the detailed report round, and the opponent mixes all actions each period. Similarly, let $\Pr(h_{-i}^R|\omega,s_i^R,h_i^\tau)$ denote the probability of h_{-i}^R given that the history during the first τ periods of the detailed report round is h_i^τ.

Lemma 8. *There are $C>0$, L, and $\overline{T}$ such that for any $T>\overline{T}$, ω, and i, there is a reporting strategy $\sigma_i^{report,\omega}$, an inference rule P^{ω}, and a transfer rule $U_i^{report,\omega}$ which satisfy the following properties:*

(i) For each h_i^{2T},

$$\sum_{h_{-i}^R\in H_{-i}^R}\Pr(h_{-i}^R|\omega,\sigma_i^{report,\omega}(h_i^{2T}))P^{\omega}(h_{-i}^R)[h_i^{2T}]\geq(1-\exp(-T^{\frac{1}{2}}))^{2T}.$$

(ii) For each h_i^{2T},

$$\sum_{h_{-i}^{2T}\in H_{-i}^{2T}}\sum_{h_{-i}^R\in H_{-i}^R}\Pr(h_{-i}^{2T}|\omega,h_i^{2T},\alpha_{-i},m_{-i}^{\omega})\Pr(h_{-i}^R|\omega,\sigma_i^{report,\omega}(h_i^{2T}))U_i^{report,\omega}(h_{-i}^{2T},h_{-i}^R)=0.$$

(iii) For each h_i^{2T}, for each on-path history h_i^τ with $\Pr(h_i^\tau|\omega,\sigma_i^{report,\omega}(h_i^{2T}))>0$, and for each pure strategy s_i^R with $s_i^R(h_i^\tau)\neq\sigma_i^{report,\omega}(h_i^{2T})[h_i^\tau]$,

$$\sum_{h_{-i}^{2T}\in H_{-i}^{2T}}\sum_{h_{-i}^R\in H_{-i}^R}\Pr(h_{-i}^{2T}|\omega,h_i^{2T},\alpha_{-i},m_{-i}^{\omega})\Pr(h_{-i}^R|\omega,\sigma_i^{report,\omega}(h_i^{2T}),h_i^\tau)U_i^{report,\omega}(h_{-i}^{2T},h_{-i}^R)$$

$$\geq\sum_{h_{-i}^{2T}\in H_{-i}^{2T}}\sum_{h_{-i}^R\in H_{-i}^R}\Pr(h_{-i}^{2T}|\omega,h_i^{2T},\alpha_{-i},m_{-i}^{\omega})\Pr(h_{-i}^R|\omega,s_i^R,h_i^\tau)U_i^{report,\omega}(h_{-i}^{2T},h_{-i}^R)+\frac{1}{T^{18}}$$

(iv) $|U_i^{report,\omega}(h_{-i}^{2T}, h_{-i}^R)| < CT^2$ *for all* h_{-i}^{2T} *and* h_{-i}^R.

Clause (i) of the lemma asserts that communication is almost perfect, in the sense that the opponent's inference matches player i's detailed report almost surely. (Note that the right-hand side of the inequality is at least $1 - 2T\exp(-T^{\frac{1}{2}})$, which converges to one according to l'Hôpital's rule.) Clause (ii) implies that the expected value of the transfer $U_i^{\text{report},\omega}$ is zero, if player i reports truthfully and the opponent plays m_{-i}^{ω} in her summary report round (which happens when the opponent's inference is $\omega(-i) = \omega$). Clause (iii) ensures that in each period of the detailed reporting round, player i has a strict incentive to follow the reporting strategy $\sigma_i^{\text{report},\omega}$, if we ignore the stage-game payoffs. Clause (iv) gives a bound on the transfer function $U_i^{\text{report},\omega}$.

Proof. The proof is very similar to the one for Lemma 14 of Sugaya (2019). Specifically, Sugaya considers an equilibrium in which each player i reports both (i) her history in the "review block" and (ii) her history in the "non-review block." The reporting strategy for (ii) is much simpler than the one for (i), and our reporting strategy is exactly the same as this simpler one. So we provide only the outline of the proof.[16]

To illustrate the idea, consider the case in which each player i has two actions, a_i^G and a_i^B, and two signals, y_i^G and y_i^B. Fix ω. Player i's detailed report round for ω consists of two stages:

Stage 1: Player i reports $h_i^{2T} = (a_i^t, y_i^t)_{t=1}^{2T}$ using her actions. Specifically, in the first T periods of this stage, she reports the first component a_i^1 of her message; she chooses a_i^G for T periods if $a_i^1 = a_i^G$, and a_i^B for T periods if $a_i^1 = a_i^B$. By the law of large numbers, the opponent can obtain very accurate information about player i's message. Similarly, in the next T periods, she reports the second component y_i^1 of her message; she chooses a_i^G for T periods if $y_i^1 = y_i^G$, and a_i^B for T periods if $y_i^1 = y_i^B$. In this way, she reports each component of h_i^{2T} sequentially. The total length of this stage is $T \cdot 4T = 4T^2$, because h_i^{2T} has $4T$ components.

[16] Sugaya's proof uses public randomization, in two places. First, in his block strategy, only one of the players reports the history, and this player is chosen by public randomization. Second, public randomization is used when a player reports her history in the review block. In our lemma, we do not need public randomization because we have a report round for each player i and we use the communication protocol for the history in the non-review block, rather than the one for the review block.

If signals are conditionally independent (i.e., if signals are not correlated across players), the above communication protocol is enough for the result we want: We can find an inference rule and a transfer rule such that the opponent's inference coincides with player i's message almost surely, and the truthful report (taking a constant action for each T-period interval) is a best reply for player i. The point here is that under the conditionally independence assumption, player i's signal has no information about the opponent's signal and hence no information about the opponent's inference. This property greatly simplifies player i's incentive problem (Matsushima (2004) and Lemma 53 of Sugaya (2019)).

When signals are not conditionally independent, player i's signal is informative about the opponent's inference, which complicates player i's incentive problem. To avoid this problem, we need an additional communication stage:

Stage 2: Player i reports her private history during the first stage. With an abuse of notation, let $h_i^{4T^2} = (a_i^t, y_i^t)_{t=1}^{4T^2}$ denote this history. In the first L periods of this second stage, she reports a_i^1, the first component of the history $h_i^{4T^2}$. Specifically, she plays some L-period strategy σ_i^G if $a_i^1 = a_i^G$, and plays σ_i^B if $a_i^1 = a_i^B$. The strategy σ_i^G is very similar to but slightly different the constant action a_i^G; it chooses a_i^G in most periods, but after some histories, it chooses different actions. Likewise, the strategy σ_i^B is quite similar to the constant action a_i^B. Since these strategies induce different actions in most periods, the opponent can obtain very accurate information about player i's message. Player i reports other components of $h_i^{4T^2}$ in the same way. The total length of this stage is $L \cdot 8T^2 = 8LT^2$, as $h_i^{4T^2}$ has $8T^2$ components.

After the second stage, the opponent makes an inference about player i's message h_i^{2T}, according to some inference rule $P : H_{-i}^{4T^2+8LT^2} \to \triangle H_i^{2T}$. Very roughly, using the information during the second stage, the opponent adjusts the inference carefully so that player i's signals during the first stage has no information about the opponent's inference P. Then player i's incentive problem during the first stage becomes essentially the same as the one for the conditionally-independent case, so that we can find a transfer rule with which the truthful report during the first stage is a strict best reply for player i.

We also need to choose a transfer rule in such a way that the truthful report during the second stage is a strict best reply for player i. This is a delicate problem because signals are not conditionally independent; but it can be done if we choose the reporting strategies σ_i^G and σ_i^B carefully. See Lemma 55 of Sugaya (2019) for more details. *Q.E.D.*

6.3 Transition Rule ρ_i

So far we have defined the four block strategies s_i^{GG}, s_i^{GB}, s_i^{BG}, and s_i^{BB}. What remains is to find the transition rule ρ_i which satisfies the promise-keeping condition (7) and (8) and the incentive compatibility condition (14) and (15). As in the perfect-monitoring case, this problem is equivalent to find appropriate "transfer rules." So consider the complete-information transfer game with length T_b in which the state ω is common knowledge and player i receives a transfer after the game. Our goal is to show the following lemma, which is a counterpart to Lemmas 3 and 4.

Lemma 9. *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 ω, there are transfer rules $U_i^{\omega,G} : H_{-i}^{T_b} \to \boldsymbol{R}$ and $U_i^{\omega,B} : H_{-i}^{T_b} \to \boldsymbol{R}$ which satisfies the following properties.*

(i) For each x,

$$\frac{1-\delta}{1-\delta^{T_b}} G_i^{\omega}(s^x, U_i^{\omega,x_{-i}^{\omega}}) = \begin{cases} \overline{v}_i^{\omega} & \text{if } x_{-i}^{\omega} = G \\ \underline{v}_i^{\omega} & \text{if } x_{-i}^{\omega} = B \end{cases}$$

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

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

6.4 Proof of Lemma 9

The outline of the proof is somewhat similar to that of Lemma 3 for the perfect-monitoring case. As a first step, we will construct a transfer rule $\tilde{U}_i^{\omega,x_{-i}^{\omega}}$ which satisfies clause (ii) "approximately," i.e., the prescribed strategy $s_i^{x_i}$ is an approximate best reply given this transfer rule. Then we will modify this transfer rule so that clause (ii) holds exactly, in the sense that the prescribed strategy $s_i^{x_i}$ is an exact best reply. Also we will show that this transfer rule satisfies clause (i). Then we will modify the transfer rule further so that clause (iii) holds.

6.4.1 Step 1: Construction of $\tilde{U}_i^{\omega,x_{-i}^\omega}$

In this step, we will construct a transfer rule $\tilde{U}_i^{\omega,x_{-i}^\omega}$ such that the prescribed strategy $s_i^{x_i}$ is an approximate best reply in the complete-information transfer game.

For each ω and i, let $\tilde{u}_i^{\omega,G} : A_{-i} \times Y_{-i} \to \boldsymbol{R}$ be such that

$$g_i^\omega(a) + \sum_{y \in Y} \pi^\omega(y|a)\tilde{u}_i^{\omega,G}(a_{-i},y_{-i}) = \overline{v}_i^\omega.$$

That is, $\tilde{u}_i^{\omega,G}$ is chosen in such a way that player i becomes indifferent over all actions in the one-shot game with the true state ω, if she maximizes the stage-game payoff $g_i^\omega(a)$ plus the transfer $\tilde{u}_i^{\omega,G}(a_{-i},y_{-i})$. We choose this function carefully so that the resulting payoff is exactly equal to the target payoff $\overline{v}_i^\omega$ in Lemma 9(i). Likewise, let $\tilde{u}_i^{\omega,B} : A_{-i} \times Y_{-i} \to \boldsymbol{R}$ be such that

$$g_i^\omega(a) + \sum_{y \in Y} \pi^\omega(y|a)\tilde{u}_i^{\omega,B}(a_{-i},y_{-i}) = \underline{v}_i^\omega.$$

The existence of these functions is guaranteed under Condition 4.

The opponent's block history $h_{-i}^{T_b}$ is *regular given* ω if all the following conditions hold:

(R1) The opponent's inference is $\omega(-i) = \omega$.

(R2) The opponent's signal frequency during player i's learning round is ε-close to $\Pi_{-i}^{(i,\omega)}(a_i,\alpha_{-i})$.

(R3) The opponent's signal frequency during her summary report round is ε-close to $\Pi_{-i}^{(i,\omega)}(a_i,m_{-i}^\omega)$.

The opponent's history h_{-i}^{4T} during the learning and summary report rounds is *regular given* ω if the above three conditions hold. A history $h_{-i}^{T_b}$ (or h_{-i}^{4T}) is irregular if it is not regular. Intuitively, the opponent's history is irregular when her observation during the learning or summary report round is not typical of the state ω.

For each ω and $x_{-i}^\omega \in \{G,B\}$, consider the following transfer rule $\tilde{U}_i^{\omega,x_{-i}^\omega}$:

- For any regular history $h_{-i}^{T_b}$,

$$\tilde{U}_i^{\omega,x_{-i}^\omega}(h_{-i}^{T_b}) = \sum_{t=1}^{4T} \frac{\tilde{u}_i^{\omega,x_{-i}^\omega}(a_{-i}^t,y_{-i}^t)}{\delta^{T_b-t+1}} + \frac{\underline{U}_i^{\omega,x_{-i}^\omega}(h_{-i}^{\text{main}})}{\delta^{4R}} + \sum_{t=4T+T^3+1}^{T_b} \frac{\tilde{u}_i^{\omega,x_{-i}^\omega}(a_{-i}^t,y_{-i}^t)}{\delta^{T_b-t+1}} \tag{16}$$

where h_{-i}^{main} denotes the opponent's history during the main round.

- For any irregular history $h_{-i}^{T_b}$,

$$\tilde{U}_i^{\omega,x_{-i}^{\omega}}(h_{-i}^{T_b}) = \sum_{t=1}^{T_b} \frac{\tilde{u}_i^{\omega,x_{-i}^{\omega}}(a_{-i}^t, y_{-i}^t)}{\delta^{T_b-t+1}} \tag{17}$$

So if the history is irregular, then the opponent makes player i indifferent over all actions each period, using the transfer rule (17). If the history is regular, the transfer (16) is used. The first term offsets the stage-game payoffs during the learning and summary report rounds. Similarly, the last term offsets the stage-game payoffs during the detailed report round. The second term is the transfer defined in Lemma 7, which is useful to discipline player i's incentive during the main round.

6.4.2 Step 2: $\tilde{U}_i^{\omega,x_{-i}^{\omega}}$ approximately satisfies clause (ii)

Take a state ω and the opponent's automaton state x_{-i} as given, and suppose that player i receives the above transfer $\tilde{U}_i^{\omega,x_{-i}^{\omega}}$ after the block. In what follows, we will show that in such an complete-information transfer game, the block strategy $s_i^{x_i}$ is an approximate best reply for player i. To do so, we will first find player i's optimal strategy s_i^* in this complete-information transfer game, and then show that this optimal strategy s_i^* and the block strategy $s_i^{x_i}$ yield almost the same payoff.

It turns out that the following strategy s_i^* is a best reply for player i in the complete-information transfer game with the state ω:

- The play during the learning, summary report, and detailed report rounds are exactly the same as the one induced by $s_i^{x_i}$.
- During the main round, play $s_i^{\omega,G}$ or $s_i^{\omega,B}$ regardless of the past history.

This optimal strategy s_i^* differs from the prescribed strategy $s_i^{x_i}$ regarding the play during the main round. While the optimal strategy s_i^* always induces $s_i^{\omega,G}$ or $s_i^{\omega,B}$ regardless of the past history, the prescribed strategy $s_i^{x_i}$ may induce $s_i^{\tilde{\omega},G}$ or $s_i^{\tilde{\omega},B}$, depending on the history in player i's learning round and the opponent's summary report round. For example, if player i has the wrong inference $\omega(i) = \tilde{\omega}$, the prescribed strategy induces $s_i^{\tilde{\omega},G}$ or $s_i^{\tilde{\omega},B}$. Let $H_i^{2T,\tilde{\omega}}$ denote the set of all such on-path histories h_i^{2T}, that is, it is the set of histories h_i^{2T} such that player i did not deviate during these rounds and such that the strategy $s_i^{x_i}$ induces $s_i^{\tilde{\omega},G}$ or $s_i^{\tilde{\omega},B}$.

The following lemma shows that the above strategy s_i^* is indeed a best reply in the complete-information transfer game.

Lemma 10. *Take ω and x_{-i} as given, and consider the complete-information transfer game with $(\omega,\tilde{U}_i^{\omega,x_{-i}^{\omega}})$. Then s_i^* is a best reply against $s_{-i}^{x_{-i}}$, and yields a payoff of $\overline{v}_i^{\omega}$ if $x_{-i}^{\omega}=G$, and a payoff of $\underline{v}_i^{\omega}$ if $x_{-i}^{\omega}=B$. In particular, playing s_i^* is a best reply in each period of the block, regardless of the past history (even if player i has deviated from s_i^* in the past).*

Proof. We prove the lemma by backward induction. Consider player i's incentive during the detailed report round. Here, player i is indifferent over all actions each period, because the stage-game payoffs are offset by the term $\tilde{u}_i^{\omega,x_{-i}^{\omega}}$ in the transfer rule $\tilde{U}_i^{\omega,x_{-i}^{\omega}}$. Hence playing s_i^* in the detailed report round is a best reply, regardless of the past history. By the definition of $\tilde{u}_i^{\omega,x_{-i}^{\omega}}$, player i's per-period payoff during the detailed report round (augmented with the transfer $\tilde{u}_i^{\omega,x_{-i}^{\omega}}$) is $\overline{v}_i^{\omega}$ if $x_{-i}^{\omega}=G$, and is $\underline{v}_i^{\omega}$ if $x_{-i}^{\omega}=B$.

Next, consider player i's incentive during the main round. Suppose for now that player i knows the opponent's past history h_{-i}^{4T}. There are two cases to be considered:

Case 1: The opponent's history h_{-i}^{4T} is irregular. In this case, the transfer rule (17) will be used, so player i is indifferent over all actions each period. Hence playing s_i^* is a best reply. Again, by the definition of $\tilde{u}_i^{\omega,x_{-i}^{\omega}}$, player i's per-period payoff during the main round is exactly equal to the target payoff; it is $\overline{v}_i^{\omega}$ if $x_{-i}^{\omega}=G$, and is $\underline{v}_i^{\omega}$ if $x_{-i}^{\omega}=B$.

Case 2: The opponent's history h_{-i}^{4T} is regular. In this case, (R1) holds so that the opponent will play $s_{-i}^{\omega,x_{-i}^{\omega}}$ during the main round. Also, the transfer rule (16) will be used, so during the main round, player i maximizes the sum of the stage-game payoffs and the second term $\underline{U}_i^{\omega,x_{-i}^{\omega}}$ of the transfer. Then from Lemma 7, $s_i^{\omega,G}$ and $s_i^{\omega,B}$ are both best replies for player i during the main round, and her per-period payoff during the round is $\overline{v}_i^{\omega}$ if $x_{-i}^{\omega}=G$, and is $\underline{v}_i^{\omega}$ if $x_{-i}^{\omega}=B$.

In sum, regardless of the opponent's past history h_{-i}^{4T}, playing s_i^* during the main round is a best reply. Hence, playing s_i^* is a best reply even if player i does not know the opponent's history h_{-i}^{4T}. Note that player i's continuation payoff from the main round equals the target payoff (i.e., it is $\overline{v}_i^{\omega}$ if $x_{-i}^{\omega} = G$, and $\underline{v}_i^{\omega}$ if $x_{-i}^{\omega} = B$), regardless of the opponent's past history h_{-i}^{4T}.

Finally, consider player i's incentive in the learning and summary report rounds. Actions in these rounds have two effects: First, they influence the stage-game payoffs. Second, they influence the opponent's history h_{-i}^{4T}, which influences the opponent's continuation strategy from the main round. However, the first effect is offset by the transfer $\tilde{u}_i^{\omega, x_{-i}^{\omega}}$. Also, the second effect does not impact player i's incentive, because as noted above, player i's continuation payoff from the main round does not depend on the opponent's history h_{-i}^{4T}. Hence player i is indifferent over all actions during the learning and summary report rounds, and playing s_i^* is a best reply. Player i's per-period payoff during these rounds is equal to the target payoff, and thus her per-period payoff in the whole block is also equal to the target payoff, as desired. *Q.E.D.*

Recall that the prescribed strategy $s_i^{x_i}$ and the optimal strategy s_i^* above induce the same play after almost all on-path histories; they induce different actions only in the main round, and only when $h_i^{2T} \in H_i^{2T,\tilde{\omega}}$. So the prescribed strategy $s_i^{x_i}$ is suboptimal only in the main round, and only when $h_i^{2T} \in H_i^{2T,\tilde{\omega}}$; only in such a case, player i obtain a positive gain by deviating. In what follows, we will show that this gain is small, so the prescribed strategy $s_i^{x_i}$ is an approximate best reply.

Let $p_i^{\omega}(h_i^{2T})$ denote the conditional probability that (R2) and (R3) hold given player i's history h_i^{2T} and the opponent's inference $\omega(-i) = \omega$, i.e., let

$$p_i^{\omega}(h_i^{2T}) = \sum_{h_{-i}^{2T}:\text{(R2) and (R3) hold}} \Pr(h_{-i}^{2T} | \omega, h_i^{2T}, \alpha_{-i}, m_{-i}^{\omega}).$$

The following lemma shows that this probability is small for each history $h_i^{2T} \in H_i^{2T,\tilde{\omega}}$. This is a counterpart to Lemma 2, and Conditions 5(ii) and 6 play a crucial role in the proof.

Lemma 11. $p_i^{\omega}(h_i^{2T}) \leq \exp(-T^{\frac{1}{2}})$ *for any* $h_i^{2T} \in H_i^{2T,\tilde{\omega}}$.

Proof. If $h_i^{2T} \in H_i^{2T,\tilde{\omega}}$, we must have $\omega(i) = \tilde{\omega}$ or $\hat{\omega}(-i) = \tilde{\omega}$. (Otherwise, the prescribed strategy does not induce $s_i^{\tilde{\omega}, x_{-i}^{\tilde{\omega}}}$ in the main round.) We will prove the result for each case.

Case 1: $\omega(i)=\tilde{\omega}$. From Lemma 6(iii), player i's observation f_i during her own learning round must be ε-close to the set $\Pi_i^{(j,\omega)}(a_i,\alpha_{-i})$. Then from Condition 5(ii), player i should believe that the opponent's observation f_{-i} during player i's learning round is not in the ε-neighborhood of $\Pi_{-i}^{(i,\omega)}(a_i,\alpha_{-i})$ almost surely. This means that player i believes that (R2) does not hold almost surely, and hence the result follows. (Use Hoeffding's inequality to get the bound $\exp(-T^{\frac{1}{2}})$.)

Case 2: $\hat{\omega}(-i)=\tilde{\omega}$. Let f_i be player i's observation during the opponent's summary report round. Then by the definition of $\hat{\omega}(-i)$, $C_i^{\tilde{\omega}}(a_i,m_{-i}^{\tilde{\omega}})f_i$ must be ε-close to $\Pi_{-i}^{(i,\tilde{\omega})}(a_i,m_{-i}^{\tilde{\omega}})$. Then from Condition 6, $C_i^{\omega}(a_i,m_{-i}^{\omega})f_i$ is not ε-close to $\Pi_{-i}^{(i,\omega)}(a_i,m_{-i}^{\omega})$. This implies that player i believes that (R3) does not hold almost surely, and hence the result. *Q.E.D.*

The next lemma is the main result in this step: It shows that the prescribed strategy $s_i^{x_i}$ is an approximate best reply when T is large.

Lemma 12. *Take ω and x as given, and consider the complete-information transfer game with $(\omega,\tilde{U}_i^{\omega,x_{-i}^{\omega}})$. Suppose hypothetically that player i knows the opponent's inference $\omega(-i)$. Consider the main round, and suppose that player i's past history is $h_i^{2T}\in H_i^{2T,\tilde{\omega}}$. Then the following results hold:*

- *If $\omega(-i)\neq\omega$, player i is indifferent over all actions in the main round, so playing the prescribed strategy $s_i^{x_i}$ is an exact best reply.*

- *If $\omega(-i)=\omega$, the prescribed strategy $s_i^{x_i}$ is not optimal for player i; by playing $s_i^{\omega,G}$ or $s_i^{\omega,B}$ in the main round. she can improve her expected (unnormalized) payoff by*

$$p_i^{\omega}(h_i^{2T})\left(\begin{array}{c} E\left[\sum_{t=1}^{T^3}\delta^{t-1}g_i^{\omega}(a^t)+\delta^{T^3}\underline{U}_i^{\omega,x_{-i}^{\omega}}(h_{-i}^{main})\right|\omega,s^{\omega,x^{\omega}}\Big] \\ -E\left[\sum_{t=1}^{T^3}\delta^{t-1}g_i^{\omega}(a^t)+\delta^{T^3}\underline{U}_i^{\omega,x_{-i}^{\omega}}(h_{-i}^{main})\right|\omega,s_i^{\tilde{\omega},x_i^{\tilde{\omega}}},s_{-i}^{\omega,x_{-i}^{\omega}}\Big] \end{array}\right).$$

  From Lemma 11, $p_i^{\omega}(h_i^{2T})\leq\exp(-T^{\frac{1}{2}})$ for any $h_i^{2T}\in H_i^{2T,\tilde{\omega}}$, so this gain is approximately zero for large T.

To interpret this deviation gain, note that $\sum_{t=1}^{T^3}\delta^{t-1}g_i^{\omega}(a^t)+\delta^{T^3}\underline{U}_i^{\omega,x_{-i}^{\omega}}(h_{-i}^{\text{main}})$ is the (unnormalized) payoff in the main round, augmented with the transfer $\underline{U}_i^{\omega,x_{-i}^{\omega}}$. From Lemma 7, the strategies $s^{\omega,G}$ and $s^{\omega,B}$ maximize this value, while the strategies $s_i^{\tilde{\omega},G}$ and $s_i^{\tilde{\omega},B}$ do not. Hence the term in the brackets is positive. It is multiplied by the probability $p_i^{\omega}(h_i^{2T})$, because player i believes that the transfer $\underline{U}_i^{\omega,x_{-i}^{\omega}}$ is used with this probability (see the proof below for more details).

Proof. If $\omega(-i) \neq \omega$, the opponent's history is irregular and the transfer rule (17) will be used for sure. Hence player i is indifferent over all actions in the main round.

Now suppose that $\omega(-i) = \omega$. In this case, player i believes that the opponent's history is regular with probability $p_i^{\omega}(h_i^{2T})$, and irregular with $1 - p_i^{\omega}(h_i^{2T})$. In the latter case, the transfer rule (17) will be used and player i becomes indifferent over all actions in the main round. So the gain by deviating is zero in this case. In the former case, the transfer rule (16) will be used, so in the main round, player i maximizes the sum $\sum_{t=1}^{T^3} \delta^{t-1} g_i^{\omega}(a^t) + \delta^{T^3} \underline{U}_i^{\omega,x_{-i}^{\omega}}(h_{-i}^{\text{main}})$ of the stage-game payoffs and the transfer $\underline{U}_i^{\omega,x_{-i}^{\omega}}$. From Lemma 7, the strategies $s_i^{\omega,G}$ and $s_i^{\omega,B}$ maximize the expected value of this sum, while $s_i^{\tilde{\omega},G}$ and $s_i^{\tilde{\omega},B}$ do not. Hence deviating from the prescribed strategy is profitable. The expected deviation gain is the difference between the payoff yielded by $s_i^{\omega,G}$ and the one by $s_i^{\tilde{\omega},x_{-i}^{\tilde{\omega}}}$, multiplied by the probability $p_i^{\omega}(h_i^{2T})$. This is precisely the value stated in the lemma. *Q.E.D.*

6.4.3 Step 3: Construction of $\hat{U}_i^{\omega,x_{-i}^{\omega}}$ and clauses (i) and (ii)

In the previous step, we have constructed the transfer rule $\tilde{U}_i^{\omega,x_{-i}^{\omega}}$ such that the prescribed strategy $s_i^{x_i}$ is an approximate best reply. In particular, the prescribed strategy $s_i^{x_i}$ is suboptimal only in the main round, and only when the past history is $h_i^{2T} \in H_i^{2T,\tilde{\omega}}$.

In what follows, we will slightly modify the transfer rule $\tilde{U}_i^{\omega,x_{-i}^{\omega}}$ so that the prescribed strategy $s_i^{x_i}$ is an exact best reply even after histories $h_i^{2T} \in H_i^{2T,\tilde{\omega}}$. Also, we will show that the new transfer rule satisfies clause (i) of the lemma, i.e., the prescribed strategy achieves the target payoff.

To simplify the notation, let $\hat{h}_i^{2T,\omega}$ denote the opponent's inference about player i's detailed report h_i^{2T} during player i's detailed report round for ω.[17] That is, let $\hat{h}_i^{2T,\omega} = P^{\omega}(h_{-i}^R)$ where h_{-i}^R is the opponent's history during player i's detailed

[17] Here h_i^{2T} does not have the superscript ω, because player i will report the same message h_i^{2T} in each report round.

report round for ω. Suppose for now that communication in the detailed report round is perfect; that is, $\hat{h}_i^{2T,\omega} = h_i^{2T}$ with probability one if player i reports h_i^{2T}. Later on, we will explain how the idea can be extended to the case in which communication is imperfect.

We will add the following adjustment term to the transfer:

$$U_i^{\text{adjust},\omega}(h_{-i}^{T_b})$$
$$= \begin{cases} p_i^{\omega}(\hat{h}_i^{2T}) \left(\displaystyle\sum_{t=4T+1}^{4T+T^3} \frac{\tilde{u}_i^{\omega,x_{-i}^{\omega}}(a_{-i}^t, y_{-i}^t)}{\delta^{T_b-t+1}} - \frac{\underline{U}_i^{\omega,x_{-i}^{\omega}}(h_{-i}^{\text{main}})}{\delta^{4R}} \right) & \text{if } \omega(-i) = \omega \text{ and } \hat{h}_i^{2T,\omega} \in H_i^{2T,\tilde{\omega}} \\ 0 & \text{otherwise} \end{cases}$$

The idea of this adjustment term is as follows. Lemma 12 shows that the prescribed strategy is not a best reply, because the transfer $\underline{U}_i^{\omega,x_{-i}^{\omega}}(h_{-i}^{\text{main}})$ is used with probability $p_i^{\omega}(h_i^{2T})$, in which case playing $s_i^{\tilde{\omega},x_{-i}^{\tilde{\omega}}}$ in the main round is suboptimal. As will be explained, the adjustment term above fixes this problem because (i) it eliminates the effect of $\underline{U}_i^{\omega,x_{-i}^{\tilde{\omega}}}(h_{-i}^{\text{main}})$ by subtracting $p_i^{\omega}(\hat{h}_i^{2T})\underline{U}_i^{\omega,x_{-i}^{\omega}}(h_{-i}^{\text{main}})$ and (ii) it makes player i indifferent in the main round by adding the term $\tilde{u}_i^{\omega,x_{-i}^{\omega}}(a_{-i}^t, y_{-i}^t)$.

We do not make an adjustment (i.e., $U_i^{\text{adjust},\omega}(h_{-i}^{T_b}) = 0$) if $\omega(-i) \neq \omega$. Indeed, in this case, the prescribed strategy is a best reply even without an adjustment (Lemma 12). Likewise, we do not make an adjustment if $h_i^{2T,\omega} \notin H_i^{2T,\tilde{\omega}}$. In this case, the prescribed strategy induces $s_i^{\omega,G}$ or $s_i^{\omega,B}$ in the main round, which is a best reply even without an adjustment.

Note that the adjustment term $U_i^{\text{adjust},\omega}(h_{-i}^{T_b})$ above is small, because Lemma 11 ensures that the probability $p_i^{\omega}(h_i^{2T})$ is small for each $h_i^{2T} \in H_i^{2T,\tilde{\omega}}$. Formally, we have the following lemma.

Lemma 13. *There is $\overline{T}$ such that for any $T > \overline{T}$ and $h_{-i}^{T_b}$, we have $|U_i^{adjust,\omega}(h_{-i}^{T_b})| < \overline{C}T^3 \exp(-T^{\frac{1}{2}})$, where $\overline{C}$ is chosen as in Lemma 7.*

Now define the new transfer rule as

$$\hat{U}_i^{\omega,x_{-i}^{\omega}}(h_{-i}^{T_b}) = \begin{cases} \tilde{U}_i^{\omega,x_{-i}^{\omega}}(h_{-i}^{T_b}) + \frac{U_i^{\text{report},\omega}(h_{-i}^{T_b})}{\delta^{4R}} + U_i^{\text{adjust},\omega}(h_{-i}^{T_b}) & \text{if } \omega(-i) = \omega \\ \tilde{U}_i^{\omega,x_{-i}^{\omega}}(h_{-i}^{T_b}) & \text{otherwise} \end{cases}.$$

Here we add two additional terms $U_i^{\text{report},\omega}$ and $U_i^{\text{adjust},\omega}$ to the original transfer rule $\tilde{U}_i^{\omega,x_{-i}^{\omega}}$, if $\omega(-i) = \omega$. We write $U_i^{\text{report},\omega}(h_{-i}^{T_b})$ rather than $U_i^{\text{report},\omega}(h_{-i}^{2T}, h_{-i}^R)$ for shorthand notation.

In the rest of this step, we will show that the prescribed strategy $s_i^{x_i}$ is an exact best reply if we use this new transfer rule $\hat{U}_i^{\omega,x_{-i}^{\omega}}$. We first show that the truthful report during the detailed report round is optimal for player i:

Lemma 14. *Take ω and x_{-i} as given. If T is sufficiently large, then regardless of the past history (even if player i has deviated before the detailed report round), the truthful report of h_i^{2T} is a best reply for player i in the detailed report round.*

Proof. During the detailed report round for $\tilde{\omega}$, player i is indifferent over all actions. This is so because actions during this round does not influence the additional terms $U_i^{\text{report},\omega}$ and $U_i^{\text{adjust},\omega}$, and player i's problem is essentially the same as the one with the original transfer rule $\tilde{U}_i^{\omega,x_{-i}^{\omega}}$.

So we will focus on the detailed report round for ω. Suppose for now that player i can observe the opponent's inference $\omega(-i)$. If $\omega(-i) \neq \omega$, then the transfer rule is $\hat{U}_i^{\omega,x_{-i}^{\omega}}(h_{-i}^{T_b}) = \tilde{U}_i^{\omega,x_{-i}^{\omega}}(h_{-i}^{T_b}) = \sum_{t=1}^{T_b} \frac{\tilde{u}_i^{\omega,x_{-i}^{\omega}}(a_{-i}^t,y_{-i}^t)}{\delta^{T_b-t+1}}$, so player i is indifferent over all actions during the detailed report round.

If $\omega(-i) = \omega$, then player i's actions during the detailed report round influence $U_i^{\text{report},\omega}$ and $U_i^{\text{adjust},\omega}$. If player i deviates in the detailed report round, it may increase the adjustment term $U_i^{\text{adjust},\omega}$, but from Lemma 13, this effect is of order $T^3\exp(-T^{\frac{1}{2}})$. On the other hand, such a deviation will decrease the expected value of $U_i^{\text{report},\omega}$, and this effect is at least $\frac{1}{T}$ from Lemma 8(iii). So for sufficiently large T, the loss is greater than the gain, and a deviation is not profitable for player i.

In sum, regardless of the opponent's inference $\omega(-i)$, the truthful report is a best reply for player i in the detailed report round. So even if player i cannot observe $\omega(-i)$, the truthful report is still a best reply. *Q.E.D.*

Next, we will show that the prescribed strategy $s_i^{x_i}$ is optimal in the main round after every *on-path* history h_i^{4T}, thanks to the adjustment term $U_i^{\text{adjust},\omega}$. (The result does not extend to off-path histories, that is, the prescribed strategy may be suboptimal at off-path histories h_i^{4T} in which player i has deviated in the learning or summary report rounds.)

Lemma 15. *Take ω, x_i, and x_{-i} as given, and pick T as in Lemma 14. After every history h_i^{4T} such that player i did not deviate in the learning or summary report* *rounds, playing the continuation strategy $s_i^{x_i}|_{h_i^{4T}}$ is a best reply for player i in the continuation game.*

Proof. Suppose for now that player i can observe the opponent's inference $\omega(-i)$. As explained in the proof of Lemma 8, if $\omega(-i) \neq \omega$, player i is indifferent over all actions each period of the block, so playing $s_i^{x_i}|_{h_i^{4T}}$ is optimal for player i.

If $\omega(-i) = \omega$, player i faces a more complex problem. There are two cases to be considered.

Case 1: Player i's past history is $h_i^{2T} \in H_i^{2T,\tilde{\omega}}$. Note that this is the case in which the prescribed strategy $s_i^{x_i}$ induces $s_i^{\tilde{\omega},x_{-i}^{\tilde{\omega}}}$, which is suboptimal under the original transfer rule $\tilde{U}_i^{\omega,x_{-i}^{\omega}}$.

Recall that the original transfer rule $\tilde{U}_i^{\omega,x_{-i}^{\omega}}$ takes the form (16) if the conditions (R1)-(R3) hold, and takes the form (17) otherwise. Since we assume that player i observes $\omega(-i) = \omega$, she knows that (R1) holds. Hence, player i with history h_i^{2T} assigns probability $p_i^{\omega}(h_i^{2T})$ on the transfer rule (16), and the remaining probability $1 - p_i^{\omega}(h_i^{2T})$ on (17). Hence, in expectation, player i faces the following transfer rule:

$$\begin{aligned}
& p_i^{\omega}(h_i^{2T}) \left\{ \sum_{t=1}^{4T} \frac{\tilde{u}_i^{\omega,x_{-i}^{\omega}}(a_{-i}^t, y_{-i}^t)}{\delta^{T_b-t+1}} + \frac{\underline{U}_i^{\omega,x_{-i}^{\omega}}(h_{-i}^{\text{main}})}{\delta^{4R}} + \sum_{t=4T+T^3+1}^{T_b} \frac{\tilde{u}_i^{\omega,x_{-i}^{\omega}}(a_{-i}^t, y_{-i}^t)}{T_b - t + 1} \right\} \\
& + (1 - p_i^{\omega}(h_i^{2T})) \sum_{t=1}^{T_b} \frac{\tilde{u}_i^{\omega,x_{-i}^{\omega}}(a_{-i}^t, y_{-i}^t)}{\delta^{T_b-t+1}} \\
& + \frac{U_i^{\text{report},\omega}(h_{-i}^{T_b})}{\delta^{4R}} + U_i^{\text{adjust},\omega}(h_{-i}^{T_b}) \qquad (18)
\end{aligned}$$

Since player i reports h_i^{2T} truthfully in the detailed report round (see Lemma 14), by the definition of the adjustment term $U_i^{\text{adjust},\omega}$, we have

$$U_i^{\text{adjust},\omega}(h_{-i}^{T_b}) = p_i^{\omega}(h_i^{2T}) \left(\sum_{t=4T+1}^{4T+T^3} \frac{\tilde{u}_i^{\omega,x_{-i}^{\omega}}(a_{-i}^t, y_{-i}^t)}{\delta^{T_b-t+1}} - \frac{\underline{U}_i^{\omega,x_{-i}^{\omega}}(h_{-i}^{\text{main}})}{\delta^{4R}} \right).$$

Plugging this into the above display, the transfer rule can be simplified to

$$\sum_{t=1}^{T_b} \frac{\tilde{u}_i^{\omega,x_{-i}^{\omega}}(a_{-i}^t, y_{-i}^t)}{\delta^{T_b-t+1}} + \frac{U_i^{\text{report},\omega}(h_{-i}^{T_b})}{\delta^{4R}}.$$

Since the second term $U_i^{\text{report},\omega}$ is not influenced by the history during the main round, this transfer rule makes player i indifferent over all actions in the main round. Hence playing $s_i^{x_i}|_{h_i^{4T}}$ is a best reply for player i.

Case 2: Player i's past history is $h_i^{2T} \notin H_i^{2T,\tilde{\omega}}$. This is the case in which the prescribed strategy $s_i^{x_i}$ induces $s_i^{\omega,x_{-i}^{\omega}}$ in the main round.

Again, player i faces the transfer rule (18), but now $U_i^{\text{adjust},\omega}(h_{-i}^{T_b}) = 0$, since we consider the case with $h_i^{2T} \notin H_i^{2T,\tilde{\omega}}$. So essentially player i faces the transfer rule (16) with probability $p_i^{\omega}(h_i^{2T})$, and (17) with the remaining probability. As discussed in the proof of Lemma 10, in any case, playing $s_i^{\omega,x_i^{\omega}}$ during the main round is optimal.

In sum, regardless of the opponent's inference $\omega(-i)$, playing the continuation strategy $s_i^{x_i}|_{h_i^{4T}}$ is a best reply for player i. Hence the same is true even if player i does not observe $\omega(-i)$. *Q.E.D.*

Finally, we consider player i's incentives during the learning and summary report rounds. The following lemma shows that $s_i^{x_i}$ is an exact reply so that clause (ii) holds. It also shows that the target payoff is exactly achieved, and hence clause (i) holds.

Lemma 16. *Take ω and x_{-i} as given, and pick T as in Lemma 14. Then for any automaton state x_i, the corresponding block strategy $s_i^{x_i}$ is a best reply, and yields a payoff of $\overline{v}_i^{\omega}$ if $x_{-i}^{\omega} = G$ and $\underline{v}_i^{\omega}$ if $x_{-i}^{\omega} = B$.*

Proof. As in the case with the original transfer $\tilde{U}_i^{\omega,x_{-i}^{\omega}}$, the strategy s_i^* is a best reply in the entire block game when the new transfer rule $\hat{U}_i^{\omega,x_{-i}^{\omega}}$ is used. (The proof is very similar to that of Lemma 10 and hence omitted.) Now, recall that $s_i^{x_i}$ and s_i^* induce the same actions during the learning and summary report rounds. This, together with Lemma 15, implies that s_i^* and $s_i^{x_i}$ yield the same payoff in the block. Hence $s_i^{x_i}$ is a best reply in the whole block game with the transfer rule $\hat{U}_i^{\omega,x_{-i}^{\omega}}$.

To compute the payoff, recall that the strategy s_i^* achieves the target payoff when the original transfer rule $\tilde{U}_i^{\omega,x_{-i}^{\omega}}$ is used (Lemma 10). Note also that given this strategy s_i^* and the opponent's inference $\omega(-i) = \omega$, the expected value of the additional terms is zero. (Specifically, the expected value of $U_i^{\text{report},\omega}$ is zero because of Lemma 8. The expected value of $U_i^{\text{adjust},\omega}$ is zero because both $\underline{U}_i^{\omega,x_{-i}^{\omega}}$

and $\sum_{t=4T+1}^{4T+T^3} \frac{\tilde{u}_i^{\omega,x_{-i}^{\omega}}(a_{-i}^t,y_{-i}^t)}{\delta^{T_b-t+1}}$ yield the same expected value if $s^{\omega,x^{\omega}}$ is played during the main round.) Accordingly, the strategy s_i^* achieves the target payoff even when the new transfer rule $\hat{U}_i^{\omega,x_{-i}^{\omega}}$ is used. This immediately implies the result, as $s_i^{x_i}$ and s_i^* yield the same payoff against this transfer rule $\hat{U}_i^{\omega,x_{-i}^{\omega}}$. *Q.E.D.*

So far we have assumed that communication in the detailed report round is perfect. When communication is imperfect, we need to perturb the adjustment term $U_i^{\text{adjust},\omega}$. We will consider the adjustment term which depends only on $\omega(-i)$, h_{-i}^{main}, and $\hat{h}_i^{2T,\omega}$, and we will write $U_i^{\text{adjust},\omega}(\omega(-i),h_{-i}^{\text{main}},\hat{h}_i^{2T,\omega})$ to emphasize this dependence.

When $\omega(-i) \neq \omega$, we let $U_i^{\text{adjust},\omega}(\omega(-i),h_{-i}^{\text{main}},\hat{h}_i^{2T,\omega}) = 0$, just as in the perfect-communication case. When $\omega(-i) = \omega$, we slightly perturb the adjustment term so that it solves

$$\sum_{\hat{h}_i^{2T,\omega}} \Pr(\hat{h}_i^{2T,\omega}|\omega,\sigma_i^{\text{report},\omega}(h_i^{2T}))U_i^{\text{adjust},\omega}(\omega(-i)=\omega,h_{-i}^{\text{main}},\hat{h}_i^{2T})$$

$$= \begin{cases} p_i^{\omega}(\hat{h}_i^{2T}) \left(\sum_{t=4T+1}^{4T+T^3} \frac{\tilde{u}_i^{\omega,x_{-i}^{\omega}}(a_{-i}^t,y_{-i}^t)}{\delta^{T_b-t+1}} - \frac{\underline{U}_i^{\omega,x_{-i}^{\omega}}(h_{-i}^{\text{main}})}{\delta^{8T^2}} \right) & \text{if } h_i^{2T} \in H_i^{2T,\tilde{\omega}} \\ 0 & \text{otherwise} \end{cases}.$$

for each h_{-i}^{main} and h_i^{2T}. That is, the expected value of the adjustment $U_i^{\text{adjust},\omega}$ after the main round (so h_{-i}^{main} is given) but before player i reports h_i^{2T} is exactly the same as the one for the perfect-communication case. Obviously, player i's incentive with this new adjustment term is the same as the one for the perfect-communication case, and hence $s_i^{x_i}$ is a best reply for player i. Also clause (i) still holds, as the expected value of the adjustment term does not change.

6.4.4 Step 4: Construction of $U_i^{\omega,x_{-i}^{\omega}}$ and clause (iii)

The transfer rule $\hat{U}_i^{\omega,x_{-i}^{\omega}}$ in the previous step satisfies clauses (i) and (ii) of the lemma. However, it does not satisfy clause (iii). To see this, consider the transfer rule $\hat{U}_i^{\omega,G}$, and suppose that the opponent's history $h_{-i}^{T_b}$ is irregular. Then the transfer takes the form

$$\hat{U}_i^{\omega,G}(h_{-i}^{T_b}) = \tilde{U}_i^{\omega,G}(h_{-i}^{T_b}) = \sum_{t=1}^{T_b} \frac{\tilde{u}_i^{\omega,G}(a_{-i},y_{-i})}{\delta^{T_b-t+1}}.$$

In general, the function $\tilde{u}_i^{\omega,G}(a_{-i}^t, y_{-i}^t)$ takes a positive value for some (a_{-i}, y_{-i}), and thus for some history $h_{-i}^{T_b}$, we have $\hat{U}_i^{\omega,G}(h_{-i}^{T_b}) > 0$. This implies that $\hat{U}_i^{\omega,G}$ does not satisfy clause (iii) of the lemma in general. The same argument applies to $\hat{U}_i^{\omega,B}$, and it is easy to see that $\hat{U}_i^{\omega,B}(h_{-i}^{T_b}) < 0$ for some $h_{-i}^{T_b}$.

In what follows, we will further modify the transfer rule to fix this problem. Suppose that after the block, the opponent randomly chooses a number $\theta^\omega \in \{0,1,2,3,4,5,6\}$, depending on the block history. Formally, the distribution of this random variable θ^ω is described by a mapping $Q^\omega : H_{-i}^{T_b} \to \triangle\{0,1,2,3,4,5,6\}$; given a block history $h_{-i}^{T_b}$, the opponent chooses θ^ω according to the distribution $Q^\omega(\cdot|h_{-i}^{T_b}) \in \triangle\{0,1,2,3,4,5,6\}$. We choose this mapping Q^ω as in the following lemma. The proof will be given in the next step.

Lemma 17. *For each ω, there is $Q^\omega : H_{-i}^{T_b} \to \triangle\{0,1,2,3,4,5,6\}$ which satisfies the following properties.*

(i) $\theta^\omega \geq 1$ if the opponent's history $h_{-i}^{T_b}$ is irregular: $Q^\omega(\theta^\omega = 0|h_{-i}^{T_b}) = 0$ for each irregular history $h_{-i}^{T_b}$.

(ii) Given a state ω, the expected value of θ^ω is small, and it is independent of player i's strategy $s_i^{T_b}$ and of the opponent's automaton state x_{-i}: Let

$$E[\theta^\omega|\omega, s_i^{T_b}, s_{-i}^{x_{-i}}] = \sum_{h_{-i}^{T_b} \in H_{-i}^{T_b}} \Pr(h_{-i}^{T_b}|s_i^{T_b}, s_{-i}^{x_{-i}}) \sum_{\theta^\omega \in \{0,1,2,3,4,5,6\}} Q^\omega(\theta^\omega|h_{-i}^{T_b}).$$

Then $E[\theta^\omega|\omega, s_i^{T_b}, s_{-i}^{x_{-i}}]$ is independent of $s_i^{T_b}$ and x_{-i} (so we denote it by $E[\theta^\omega|\omega]$), and $E[\theta^\omega|\omega] \leq 6\exp(-T^{\frac{1}{2}})$.

Now we define a new transfer rule $U_i^{\omega,G}$ such that for each $h_{-i}^{T_b}$,

$$U_i^{\omega,G}(h_{-i}^{T_b}) = \hat{U}_i^{\omega,G}(h_{-i}^{T_b}) - 2T_b\overline{u}_i\theta^\omega + 2T_b\overline{u}_iE[\theta^\omega|\omega],$$

where $\overline{u} > 0$ is a constant such that $\overline{u}_i > |\tilde{u}_i^{\omega,x_{-i}^\omega}(a_{-i}, y_{-i})|$ for all x_{-i}^ω, a_{-i}, and y_{-i}. That is, we add two terms, $-2T_b\overline{u}_i\theta^\omega$ and $2T_b\overline{u}_iE[\theta^\omega|\omega]$. Similarly, for each $h_{-i}^{T_b}$, let

$$U_i^{\omega,B}(h_{-i}^{T_b}) = \hat{U}_i^{\omega,B}(h_{-i}^{T_b}) + 2T_b\overline{u}_i\theta^\omega - 2T_b\overline{u}_iE[\theta^\omega|\omega].$$

Note that these additional terms do not influence player i's incentive. Indeed, the first additional term $2T_b\overline{u}_i\theta^\omega$ is independent of player i's strategy (Lemma

17(ii)), and the second one is just a constant. Hence, the new transfer rule $U_i^{\omega,x_{-i}^{\omega}}$ still satisfies Lemma 9(ii). Also this transfer rule satisfies Lemma 9(i), because the sum of these additional terms is zero in expectation.

So what remains is to prove Lemma 9(iii). We will first show that $-\frac{\overline{v}_i^{\omega}-\underline{v}_i^{\omega}}{1-\delta} < U_i^{\omega,G}(h_{-i}^{T_b}) < 0$ for each $h_{-i}^{T_b}$. There are three cases to be considered.

Case 1: $h_{-i}^{T_b}$ does not satisfy (R1). In this case, Lemma 17(i) implies that $\theta^{\omega} \geq 1$ with probability one. Also, by the construction, $\hat{U}_i^{\omega,G}(h_{-i}^{T_b}) = \sum_{t=1}^{T_b} \frac{\tilde{u}_i^{\omega,G}(a_{-i}^t,y_{-i}^t)}{\delta^{T_b-t+1}}$. Hence, using $\overline{u}_i > |\tilde{u}_i^{\omega,G}(a_{-i},y_{-i})|$ and Lemma 17(ii), we have

$$-\sum_{t=1}^{T_b} \frac{\overline{u}_i}{\delta^{T_b-t+1}} - 12T_b\overline{u}_i - 12T_b\overline{u}_i\exp(-T^{\frac{1}{2}})$$

$$< U_i^{\omega,G}(h_{-i}^{T_b}) < \sum_{t=1}^{T_b} \frac{\overline{u}_i}{\delta^{T_b-t+1}} - 2T_b\overline{u}_i + 12T_b\overline{u}_i\exp(-T^{\frac{1}{2}}),$$

where the lower bound is derived by considering the case with $\theta^{\omega} = 6$, and the upper bound is derived by considering the case with $\theta^{\omega} = 1$. Taking the limit as $\delta \to 1$,

$$-13T_b\overline{u}_i - 12T_b\overline{u}_i\exp(-T^{\frac{1}{2}}) < \lim_{\delta\to 1} U_i^{\omega,G}(h_{-i}^{T_b}) < -T_b\overline{u}_i + 12T_b\overline{u}_i\exp(-T^{\frac{1}{2}}).$$

When T is sufficiently large, $T_b\exp(-T^{\frac{1}{2}})$ is almost negligible, so

$$-14T_b\overline{u}_i < \lim_{\delta\to 1} U_i^{\omega,G}(h_{-i}^{T_b}) < 0.$$

So in the limit as $\delta \to 1$, the transfer $U_i^{\omega,G}(h_{-i}^{T_b})$ satisfies the inequality in Lemma 9(iii). By the continuity, the same inequality holds for δ close to one.

Case 2: $h_{-i}^{T_b}$ satisfies (R1) but it is irregular (so (R2) or (R3) does not hold). As in the previous case, Lemma 17(i) implies that $\theta^{\omega} \geq 1$ with probability one. Also, by the construction, we have

$$\hat{U}_i^{\omega,G}(h_{-i}^{T_b}) = \sum_{t=1}^{T_b} \frac{\tilde{u}_i^{\omega,G}(a_{-i}^t,y_{-i}^t)}{\delta^{T_b-t+1}} + U_i^{\text{adjust},\omega}(h_{-i}^{T_b}) + \frac{U_i^{\text{report},\omega}(h_{-i}^{T_b})}{\delta^{4R}}.$$

Thus from $\overline{u}_i > |\tilde{u}_i^{\omega,G}(a_{-i},y_{-i})|$ and Lemmas 8, 13, and 17(ii), we obtain

$$-\sum_{t=1}^{T_b} \frac{\overline{u}_i}{\delta^{T_b-t+1}} - T^3\exp(-T^{\frac{1}{2}}) - \frac{CT^2}{\delta^{4R}} - 12T_b\overline{u}_i - 12T_b\overline{u}_i\exp(-T^{\frac{1}{2}})$$

$$< U_i^{\omega,G}(h_{-i}^{T_b}) < \sum_{t=1}^{T_b} \frac{\overline{u}_i}{\delta^{T_b-t+1}} + T^3\exp(-T^{\frac{1}{2}}) + \frac{CT^2}{\delta^{4R}} - 2T_b\overline{u}_i + 12T_b\overline{u}_i\exp(-T^{\frac{1}{2}}).$$

Again, we take $\theta^{\omega} = 6$ and $\theta^{\omega} = 1$ to obtain the lower and upper bounds, respectively. Taking the limit as $\delta \to 1$,

$$\begin{aligned}
&- T^3 \exp(-T^{\frac{1}{2}}) - CT^2 - 13T_b\overline{u}_i - 12T_b\overline{u}_i \exp(-T^{\frac{1}{2}}) \\
&< \lim_{\delta \to 1} U_i^{\omega,G}(h_{-i}^{T_b}) < T^3 \exp(-T^{\frac{1}{2}}) + CT^2 - T_b\overline{u}_i + 12T_b\overline{u}_i \exp(-T^{\frac{1}{2}}).
\end{aligned}$$

When T is sufficiently large, $T^3 \exp(-T^{\frac{1}{2}})$ and $T_b \exp(-T^{\frac{1}{2}})$ are almost negligible, and also $T_b\overline{u}_i > CT^2$. Hence,

$$-14T_b\overline{u}_i < \lim_{\delta \to 1} U_i^{\omega,G}(h_{-i}^{T_b}) < 0.$$

As in Case 1, this implies the desired inequality for δ close to one.

Case 3: $h_{-i}^{T_b}$ is regular. In this case, we have

$$\begin{aligned}
\hat{U}_i^{\omega,G}(h_{-i}^{T_b}) = &\sum_{t=1}^{4T} \frac{\tilde{u}_i^{\omega,G}(a_{-i}^t, y_{-i}^t)}{\delta^{T_b - t + 1}} + \frac{\underline{U}_{-i}^{\omega,G}(h_{-i}^{\text{main}})}{\delta^{8T^2}} + \sum_{t=4T+T^3+1}^{T_b} \frac{\tilde{u}_i^{\omega,G}(a_{-i}^t, y_{-i}^t)}{\delta^{T_b - t + 1}} \\
&+ U_i^{\text{adjust},\omega}(h_{-i}^{T_b}) + \frac{U_i^{\text{report},\omega}(h_{-i}^{T_b})}{\delta^{4R}}.
\end{aligned}$$

Hence from $\overline{u}_i > |\tilde{u}_i^{\omega,G}(a_{-i}, y_{-i})|$ and Lemmas 7, 8, 13, and 17(ii),

$$\begin{aligned}
&-\sum_{t=1}^{4T} \frac{\overline{u}_i}{\delta^{T_b - t + 1}} - \overline{C}T^3 - \sum_{t=4T+8T^2+1}^{T_b} \frac{\overline{u}_i}{\delta^{T_b - t + 1}} - T^3 \exp(-T^{\frac{1}{2}}) - \frac{CT^2}{\delta^{4R}} - 12T_b\overline{u}_i - 12T_b\overline{u}_i \exp(-T^{\frac{1}{2}}) \\
&< U_i^{\omega,G}(h_{-i}^{T_b}) \\
&< \sum_{t=1}^{4T} \frac{\overline{u}_i}{\delta^{T_b - t + 1}} - \underline{C}T^3 + \sum_{t=4T+8T^2+1}^{T_b} \frac{\overline{u}_i}{\delta^{T_b - t + 1}} + T^3 \exp(-T^{\frac{1}{2}}) + \frac{CT^2}{\delta^{4R}} + 12T_b\overline{u}_i \exp(-T^{\frac{1}{2}}).
\end{aligned}$$

Here we take $\theta^{\omega} = 0$ (rather than $\theta^{\omega} = 1$) to obtain the upper bound; this is so because Lemma 17(i) does not apply when the opponent's history is regular. Taking the limit as $\delta \to 1$,

$$\begin{aligned}
&-\overline{C}T^3 - T^3 \exp(-T^{\frac{1}{2}}) - CT^2 - 13T_b\overline{u}_i - 12T_b\overline{u}_i \exp(-T^{\frac{1}{2}}) \\
&< \lim_{\delta \to 1} U_i^{\omega,G}(h_{-i}^{T_b}) < (4T + 8T^2)\overline{u}_i - \underline{C}T^3 + T^3 \exp(-T^{\frac{1}{2}}) + CT^2 + 12T_b\overline{u}_i \exp(-T^{\frac{1}{2}}).
\end{aligned}$$

When T is sufficiently large, $T^3 \exp(-T^{\frac{1}{2}})$ and $T_b \exp(-T^{\frac{1}{2}})$ are almost negligible, and also $\underline{C}T^3 > (4T + 8T^2)\overline{u}_i + CT^2$. Hence

$$-\overline{C}T^3 - CT^2 - 14T_b\overline{u}_i < \lim_{\delta \to 1} U_i^{\omega,G}(h_{-i}^{T_b}) < 0.$$

As in the previous cases, this implies the desired inequality for δ close to one.

In sum, regardless of the opponent's history $h_{-i}^{T_b}$, the transfer $U_i^{\omega,G}(h_{-i}^{T_b})$ satisfies the inequality in Lemma 9(iii). The same argument applies to the transfer rule $U_i^{\omega,B}$.

6.4.5 Step 5: Proof of Lemma 17

We define θ^{ω} as a sum of three random variables:

$$\theta^{\omega} = \theta^{\omega,1} + \theta^{\omega,2} + \theta^{\omega,3}.$$

For each $k \in \{1,2,3\}$, we will choose the random variable $\theta^{\omega,k}$ so that

(i) $\theta^{\omega,k} \geq 1$ if (Rk) does not hold.

(ii) Given that the true state is ω, the expected value of $\theta^{\omega,k}$ is less than $2\exp(-T^{\frac{1}{2}})$, and it is independent of player i's strategy $s_i^{T_b}$ and of the opponent's automaton state x_{-i}.

If there are such $\theta^{\omega,1}$, $\theta^{\omega,2}$, and $\theta^{\omega,3}$, then the result immediately follows. So we will explain how to find such $\theta^{\omega,1}$, $\theta^{\omega,2}$, and $\theta^{\omega,3}$.

Substep 1: Construction of $\theta^{\omega,1}$

In the opponent's learning round, the opponent computes the base score, the random score, and the final score, and determines the inference $\omega(-i)$ depending on these scores. (See the proofs of Lemmas 1 and 6.) By the definition of $\omega(-i)$, (R1) does not hold if and only if

(a) $|q_{-i}^{\omega} - q_{-i}^{\text{random}}| \geq 2\tilde{\varepsilon}$, or

(b) $|q_{-i}^{\text{base}} - q_{-i}^{\text{random}}| \geq \tilde{\varepsilon}$.

Let $\xi \in \{0,1\}$ be such that $\xi = 1$ if and only if (a) holds, and let $\eta \in \{0,1\}$ be such that $\eta = 1$ if and only if (b) holds. Let $\tilde{\theta}^{\omega,1} = \xi + \eta$.

Obviously this variable $\tilde{\theta}^{\omega,1}$ satisfies the property (i) above, that is, $\tilde{\theta}^{\omega,1} \geq 1$ if (R1) does not hold. Also, the expected value of ξ is independent of player i's strategy $s_i^{T_b}$ and of the opponent's automaton state x_{-i}, as the distribution of the random score q_{-i}^{random} is independent of $s_i^{T_b}$ and x_{-i}. However, the variable

$\tilde{\theta}^{\omega,1} = \xi + \eta$ does not satisfy the property (ii) above, because the expected value of η depends on player i's strategy $s_i^{T_b}$. In what follows, we will modify this variable η so that its expected value is independent of $s_i^{T_b}$.

Let h_{-i}^T denote the opponent's history during her own learning round. For each h_{-i}^T, let

$$\hat{p}^\omega(h_{-i}^T) = \Pr(|q_{-i}^{\text{base}} - q_{-i}^{\text{random}}| \geq \tilde{\varepsilon}|\omega, h_{-i}^T)$$

be the probability that (b) occurs conditional on h_{-i}^T. By Hoeffding's inequality, $\hat{p}^\omega(h_{-i}^T) \leq \exp(-T^{\frac{1}{2}})$.

Then define $\hat{\eta}^1 \in \{0,1\}$ such that

- If (b) holds, then let $\hat{\eta}^1 = 1$.
- If not, then let $\hat{\eta}^1 = 1$ with probability $\frac{\exp(-T^{\frac{1}{2}}) - \hat{p}^\omega(h_{-i}^T)}{1-\hat{p}^\omega(h_{-i}^T)}$, and let $\hat{\eta}^1 = 0$ with the remaining probability.

That is, we let $\hat{\eta}^1 = 1$ not only when (b) occurs, but also when (b) does not occur, with positive probability. This probability is adjusted depending on the opponent's history h_{-i}^T so that all histories h_{-i}^T induce the same probability of $\hat{\eta} = 1$. Indeed, given h_{-i}^T, the probability of $\hat{\eta} = 1$ is

$$\Pr(|q_{-i}^{\text{base}} - q_{-i}^{\text{random}}| \geq \tilde{\varepsilon}|\omega, h_{-i}^T) + (1 - |q_{-i}^{\text{base}} - q_{-i}^{\text{random}}| \geq \tilde{\varepsilon}|\omega, h_{-i}^T) \frac{\exp(-T^{\frac{1}{2}}) - \hat{p}^\omega(h_{-i}^T)}{1-\hat{p}^\omega(h_{-i}^T)}$$

$$= \hat{p}^\omega(h_{-i}^T) + (1 - \hat{p}^\omega(h_{-i}^T)) \frac{\exp(-T^{\frac{1}{2}}) - \hat{p}^\omega(h_{-i}^T)}{1-\hat{p}^\omega(h_{-i}^T)}$$

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

Accordingly, the expected value of the random variable $\theta^{\omega,1} = \xi^1 + \hat{\eta}^1$ is independent of player i's strategy $s_i^{T_b}$ and of the opponent's automaton state x_{-i}. Also this expected value is at most $2\exp(-T^{\frac{1}{2}})$, as the expected value of $\hat{\eta}^1$ is $\exp(-T^{\frac{1}{2}})$ (this is shown in the above display) and the expected value of ξ^1 is at most $\exp(-T^{\frac{1}{2}})$ (this follows from Hoeffding's inequality). Hence this random variable $\theta^{\omega,1} = \xi^1 + \hat{\eta}^1$ satisfies the desired properties (i) and (ii).

Substep 2: Construction of $\theta^{\omega,2}$

To define $\theta^{\omega,2}$, consider player i's learning round, in which players are supposed to play (a_i, α_{-i}). Let $f_{-i} \in \triangle(A_{-i} \times Y_{-i})$ denote the opponent's observation during this round, and let $\pi^{\omega}_{-i}(\tilde{a}_i, \alpha_{-i})$ denote the distribution of (a_{-i}, y_{-i}) given ω and $(\tilde{a}_i, \alpha_{-i})$ for shorthand notation.

After this round, the opponent computes a base score $q^{\text{base}}_{-i} \in \boldsymbol{R}^{|A_{-i}| \times |Y_{-i}|}$ using the formula

$$q^{\text{base}}_{-i} = Q_{-i} f_{-i},$$

where Q_{-i} is a $|A_{-i} \times Y_{-i}| \times |A_{-i} \times Y_{-i}|$ matrix. We choose this matrix so that there is some $q^{\omega}_{-i} \in \boldsymbol{R}^{|A_{-i}| \times |Y_{-i}|}$ such that

$$Q_{-i} \pi^{\omega}_{-i}(\tilde{a}_i, \alpha_{-i}) = q^{\omega}_{-i}$$

for each $\tilde{a}_i$. That is, we choose Q_{-i} so that player i cannot influence the expected value of the base score. Condition 4 ensures that such a matrix exists. Then the opponent generates a random score q^{random}_{-i}, just as explained in the proof of Lemma 1.

Take $\tilde{\varepsilon}$ smaller than ε. Then (R2) does not hold only if $|q^{\omega}_{-i} - q^{\text{base}}_{-i}| \geq \tilde{\varepsilon}$. This in turn implies that (R2) does not hold only if

(a) $|q^{\omega}_{-i} - q^{\text{random}}_{-i}| \geq \frac{\tilde{\varepsilon}}{2}$, or

(b) $|q^{\text{base}}_{-i} - q^{\text{random}}_{-i}| \geq \frac{\tilde{\varepsilon}}{2}$.

Define ξ^2 and $\hat{\eta}^2$ as in the previous substep, and let $\theta^{\omega,2} = \xi^2 + \hat{\eta}^2$. Then this random variable satisfies the desired properties (i) and (ii). The proof is very similar to the one in the previous substep and hence omitted.

Substep 3: Construction of $\theta^{\omega,3}$

The argument is exactly the same as the one for $\theta^{\omega,2}$, and hence omitted.

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.

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^{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 ω, the score q_i^{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 ω_1 from ω_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 18. *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 ω, the expected base score is q_i^{ω} 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 18, 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 kth 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 kth 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 ω, the random score is close to q_i^{ω} 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 18, 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^{ω} so that the signals indicate that ω 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^{ω}. 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^{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^{random} is equal to the base score q_i^{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

$$\left| q_i^{\omega_1} - q_i^{\text{final}} \right| < 2\tilde{\varepsilon}, \tag{19}$$

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

$$\left| q_i^{\omega_2} - q_i^{\text{final}} \right| < 2\tilde{\varepsilon}. \tag{20}$$

If neither (19) nor (20) holds, then we set $\omega(i) = \emptyset$. In words, if the score is in the $2\tilde{\varepsilon}$-neighborhood of the expected score at ω, 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^{base}, which is equal to $Q_i(a^*)f_i$. Hence, from (19) and (20), we must have

$$|q_i^{\omega} - Q_i(a^*)f_i| < 3\tilde{\varepsilon}. \tag{21}$$

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

$$|\pi_i^\omega(a^*) - f_i| < K\tilde{\varepsilon} \tag{22}$$

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* (ω, 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* (ω, 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 ω. 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 (ω, G), choose $\tilde{U}_i^{\omega,G}(h_{-i}^{T_b})$ so that it solves

$$\frac{1-\delta}{1-\delta^{T_b}}\left[\sum_{t=1}^{T_b}\delta^{t-1}g_i^{\omega}(a^t)+\delta^{T_b}\tilde{U}_i^{\omega,G}(h_{-i}^{T_b})\right]=\overline{v}_i^{\omega}. \tag{23}$$

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

$$\frac{1-\delta}{1-\delta^{T_b}}\left[\sum_{t=1}^{T_b}\delta^{t-1}g_i^{\omega}(a^t)+\delta^{T_b}\tilde{U}_i^{\omega,G}(h_{-i}^{T_b})\right]=-2\overline{g}_i^{\omega}. \tag{24}$$

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 ω, 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 ω, 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 ω 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 19. *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 20. *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 21. *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^{ω}, 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 22. *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 21.

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^{ω}, the truthful report

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

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

for this misreport $\hat{z}_i^t$, 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 23. *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 22. 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 20 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^{ω} 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^ω 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}^ω denote the probability of the opponent's block history $h_{-i}^{T_b}$ being regular given (ω, G), conditional on that the state is ω 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

$$c^G = (1 - p_{-i}^\omega)(\overline{v}_i^\omega + 2\overline{g}_i^\omega) + E\left[\frac{\varepsilon}{T} \sum_{t \in T(i)} \left| e(z_{-i}^t) - C_i^\omega(a^*)e(\hat{z}_i^t) \right|^2 - b_i^\omega(h_{-i}^{T_b}) \right| \omega, s^x \left. \vphantom{\sum_{t}} \right]. \tag{25}$$

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 (23) and (24) 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}^ω 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

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

Plugging (25) 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 (23) and (24), 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 δ 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 (ω, 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 δ 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 (23),

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

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

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

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

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

Note that when T is large, p^{ω}_{-i} approximates 1 and $b^{\omega}_i(h^{T_b}_{-i})$ approximates 0 for all $h^{T_b}_{-i}$. (This follows from Lemma 21.) Hence for sufficiently large T,

$$\frac{(1-\delta)\delta^{T_b}}{1-\delta^{T_b}} U^{\omega,G}_i(h^{T_b}_{-i}) < -(2-\sqrt{2})\varepsilon < 0$$

as desired.

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

A.3 Proof of Lemma 4

Fix i and ω arbitrarily. In what follows, we will construct a transfer rule $U^{\omega,B}_i$ 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 $s^{x_{-i}}_{-i}$ and she makes the correct inference $\omega(-i) = \omega$. Formally, the opponent's block history $h^{T_b}_{-i}$ is *regular given* (ω, 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 $x^{\omega}_{-i} = B$ in the first period of the main round.

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

A history $h^{T_b}_{-i}$ is *irregular given* (ω, 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 (ω, B), choose $U_i^{\omega,B}(h_{-i}^{T_b})$ so that it solves

$$\frac{1-\delta}{1-\delta^{T_b}}\left[\sum_{t=1}^{T_b}\delta^{t-1}g_i^{\omega}(a^t)+\delta^{T_b}U_i^{\omega,B}(h_{-i}^{T_b})\right]=\underline{v}_i^{\omega}-\frac{\tau\varepsilon}{T}-c^B \tag{26}$$

  where τ 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

$$\frac{1-\delta}{1-\delta^{T_b}}\left[\sum_{t=1}^{T_b}\delta^{t-1}g_i^{\omega}(a^t)+\delta^{T_b}U_i^{\omega,B}(h_{-i}^{T_b})\right]=2\overline{g}_i^{\omega}-\frac{\tau\varepsilon}{T}-c^B. \tag{27}$$

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

$$c^B = (1 - p_{-i}^{\omega})(2\overline{g}_i^{\omega} - \underline{v}_i^{\omega}) > 0. \tag{28}$$

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 (26) and (27) 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}^{ω} is the probability of the opponent's history being regular. Plugging (28) 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 (26), (27), 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 δ, 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 (ω,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 δ 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 (26), we obtain $U_i^{\omega,B}(h_{-i}^{T_b}) > 0$.

Case 2: $h_{-i}^{T_b}$ is irregular given (ω,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 (27), we obtain $U_i^{\omega,B}(h_{-i}^{T_b}) > 0$.

Appendix B: Proof of Proposition 1 for the General Case

In this appendix, we will explain how to prove Proposition 1 when we have more than two players and more than two states.

Fix an arbitrary payoff vector $v \in \text{int}V^*$. We will construct an equilibrium which achieves this payoff v. For now, we maintain the assumption $|A_i| \geq |Z_i|$ for each i. At the end of this appendix, we will explain how to drop this assumption.

As in Section 4, for each state ω, choose payoffs $\underline{v}_i^\omega$ and $\overline{v}_i^\omega$ for each i so that $\underline{v}_i^\omega < v_i^\omega < \overline{v}_i^\omega$ for each i and that the product set $\times_{i \in I}[\underline{v}_i^\omega, \overline{v}_i^\omega]$ is in the interior of the set $V^*(\omega)$. Then for each ω and $x^\omega \in \{G,B\}^N$, choose an action profile a^{ω,x^ω} such that $g_i^\omega(a^{\omega,x^\omega}) > \overline{v}_i^\omega$ for each i with $x_{i-1}^\omega = G$ and $g_i^\omega(a^{\omega,x^\omega}) < \underline{v}_i^\omega$ for each i with $x_{i-1}^\omega = B$, where $i-1 = N$ for $i = 1$. That is, we choose this action profile a^{ω,x^ω} so that player i's payoff is lower than $\underline{v}_i^\omega$ if player $i-1$ plans to punish player i, while the payoff is higher than $\overline{v}_i^\omega$ if player $i-1$ plans to reward player i. Looking ahead, in our equilibrium, player i's payoff is determined solely by player $i-1$'s plan about whether to reward or punish player i.

Then as in Section 4, we pick $\varepsilon > 0$ sufficiently small so that all the following conditions hold:

- For each ω, i, x^ω, and $\tilde{x}^\omega$ such that $x_{i-1}^\omega = B$ and $\tilde{x}_{i-1}^\omega = G$,

$$\max\{g_i^\omega(a^{\omega,x^\omega}), m_i^\omega\} < \underline{v}_i - \varepsilon < \overline{v}_i + 2\varepsilon < g_i^\omega(a^{\omega,\tilde{x}^\omega}).$$

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

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

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

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

B.1 Automaton with State-Contingent Punishment

Let $T_b = \frac{(N+1)NT|\Omega|(|\Omega|-1)}{2} + T^2 + 1$, where T is to be specified later. As in Section 4, our equilibrium is described as an automaton over blocks. Specifically, the infinite horizon is divided into a sequence of *blocks* with T_b periods. At the beginning of each block, each player i chooses an automaton state $x_i = (x_i^\omega)_{\omega \in \Omega} \in \{G,B\}^{|\Omega|}$.

As in Section 4, this automaton state x_i can be interpreted as player i's state-contingent plan about whether to reward or punish player $i+1$: The automaton state x_i has $|\Omega|$ components, and each component x_i^{ω} represents her plan at state ω. Specifically, when $x_i^{\omega} = G$, player i plans to reward player $i+1$ at state ω. Likewise, when $x_i^{\omega} = B$, player i plans to punish player $i+1$ at state ω. Each player i's play during the block is solely determined by this automaton state x_i. Let $s_i^{x_i}$ denote the block strategy induced by an automaton state x_i.

After each block, each player i randomly chooses a new automaton state (plan) $\tilde{x}_i$ for the next block. Specifically. a new plan x_i^{ω} for state ω is chosen according to some distribution $\rho_i^{\omega}(\cdot|x_i^{\omega}, h_i^{T_b}) \in \triangle\{G,B\}$.

B.2 Block Strategy $s_i^{x_i}$

B.2.1 Brief Description

We will describe the block strategy $s_i^{x_i}$ for each automaton state x_i. As in Section 4, each block with length T_b is further divided into the *Learning Round*, the *Summary Report Round*, the *Main Round*, and the *Detailed Report Round*. Specifically:

Learning Round: The first $\frac{|\Omega|(|\Omega|-1)}{2}T$ periods of the block are player 1's learning round, in which player 1 collects private signals and makes an inference $\omega(i) \in \Omega \cup \emptyset$ about the state. Then there is player 2's learning round, player 3's learning round, and so on. So in total, the learning round consists of $\frac{|\Omega|(|\Omega|-1)}{2}T$ periods. The way each player i makes the inference $\omega(i)$ will be specified later. Let $T(i)$ denote the set of the periods included in player i's learning round. Throughout the learning round, players play a^*, so that Condition 2 ensures that state learning is indeed possible.

Summary Report Round: The next period is the summary report round, in which each player i reports her summary inference $\omega(i)$ through actions. For simplicity, we assume that each player has at least $|\Omega|+1$ actions so that she can indeed represent $\omega(i)$ through one-shot actions; but this assumption is dispensable, as discussed in Section 4.

Main Round: The next T^2 periods are the main round. As in Section 4, players' play during the main round depends on the information reported in the summary report round. Specifically:

- If all players report ω in the summary report round (i.e., if their inferences coincide), then in the first period of the main round, each player i reveals her plan x_i^ω for this state ω through her action. After that, players choose a^{ω,x^ω} until the main round ends, where $x^\omega = (x_i^\omega)_{i\in I}$ is the reported plan. If someone (say player i) unilaterally deviates from this action profile a^{ω,x^ω}, she will be minimaxed by $\underline{\alpha}^\omega(i)$.

- If $N-1$ players report ω but one reports the null inference $\emptyset$, then the play during the main round is the same as above. (Intuitively, reporting $\omega(i) = \emptyset$ is treated as an abstention.)

- If all players report ω but one (say player j) reports $\tilde{\omega} \neq \omega$, then during the main round, each player i reveals x_i^ω, and then chooses the minimax action $\underline{\alpha}_i^\omega(j)$,

- Otherwise, the play during the main round is the same as the case in which all players report ω_1.

Detailed Report Round: The remaining $N^2 T^{\frac{|\Omega|(|\Omega|-1)}{2}}$ periods of the block are the detailed report round. In the first $NT^{\frac{|\Omega|(|\Omega|-1)}{2}}$ periods of this round, each player i reports the signal sequence $(z_i^t)_{t\in T(i)}$ observed during her own learning round. After that, each player i reports the signal sequence $(z_i^t)_{t\in T(j)}$ observed during player j's learning rounds, for each $j \neq i$.

For each automaton state x_i, let $s_i^{x_i}$ denote the block strategy which chooses actions as described above. This definition is informal, because we have not yet specified how player i forms the inference $\omega(i)$.

B.2.2 Inference Rule

We will explain how each player i makes an inference $\omega(i)$ in her own learning round. The technique is very similar to the one for the two-state case, but the notation is more involved.

We regard player i's learning round as a sequence of T-period intervals; since player i's learning round consists of $\frac{|\Omega|(|\Omega|-1)}{2}T$ periods, there are $\frac{|\Omega|(|\Omega|-1)}{2}$ such intervals. In each interval, player i compares two states, ω and $\tilde{\omega}$, and makes an inference about which one is more likely to be the true state. For example, when there are three states ω_1, ω_2, and ω_3, there are three intervals, and player i compares ω_1 with ω_2 in the first interval, ω_1 with ω_3 in the second interval, and ω_2 with ω_3 in the last interval, Let $T(i,\omega,\tilde{\omega})$ denote the T-period interval in which player i compares ω with $\tilde{\omega}$. By the definition, $T(i)$ is the union of $T(i,\omega,\tilde{\omega})$ over all possible pairs $(\omega,\tilde{\omega})$.

More specifically, in the interval $T(i,\omega,\tilde{\omega})$, player i makes an inference $r_i(\omega,\tilde{\omega}) \in \{\omega,\tilde{\omega},\emptyset\}$, depending on her private history h_i^T. The inference rule is a mapping $P_i^{(\omega,\tilde{\omega})} : H_i^T \to \triangle\{\omega,\tilde{\omega},\emptyset\}$, that is, given a history h_i^T during the interval, player i randomly selects the inference $r_i(\omega,\tilde{\omega})$ according to the distribution $P_i^{(\omega,\tilde{\omega})}(h_i^T)$. Given an inference rule $P_i^{(\omega,\tilde{\omega})}$, let $\hat{P}(\cdot|\omega^*,a^1,\cdots,a^T)$ denote the probability distribution of $r_i(\omega,\tilde{\omega})$, conditional on that the state is ω^* and players play the action sequence $(a^1,\cdots,a^T)$. Also, define $\hat{P}(\cdot|\omega,h_{-i}^t,a^{t+1},\cdots,a^T)$ as in Section 4.

We choose this inference rule $P_i^{(\omega,\tilde{\omega})}$ as in the following lemma. The proof is very similar to Lemma 1 and hence omitted.

Lemma 26. *Suppose that Condition 2 holds. Then there is $\overline{T}$ such that for any $T > \overline{T}$, ω, and $\tilde{\omega} \neq \omega$, there is an inference rule $P_i^{(\omega,\tilde{\omega})} : H_i^T \to \triangle\{\omega,\tilde{\omega},\emptyset\}$ which satisfies the following conditions:*

(i) If players do not deviate from a^*, the inference $r_i(\omega,\tilde{\omega})$ coincides with the true state almost surely: For each ω,

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

(ii) Regardless of the past history, player j's deviation cannot manipulate player i's inference almost surely: For each ω, $t \in \{0,\cdots,T-1\}$, h_{-i}^t, $(a^\tau)_{\tau=t+1}^T$, and $(\tilde{a}^\tau)_{\tau=t+1}^T$ such that $a_{-j}^\tau = \tilde{a}_{-j}^\tau = a_{-j}^*$ for all τ,

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

(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 ω: For

all $h_i^T = (a^t, z_i^t)_{t=1}^T$ *such that* $a^t = a^*$ *for all* t *and such that* $P(r_i(\omega,\tilde{\omega}) = \omega | h_i^T) > 0$,

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

Clause (i) asserts that player i's state learning is almost perfect, and clause (ii) implies that player j's gain is almost negligible even if she deviates in the interval $T(i,\omega,\tilde{\omega})$. Clause (iii) implies that player i forms the inference $r_i(\omega,\tilde{\omega}) = \omega$ only if her signal frequency is close to the true distribution $\pi_i^{\omega}(a^*)$ at ω. So if her signal frequency is not close to $\pi_i^{\omega}(a^*)$ or $\pi_i^{\tilde{\omega}}(a^*)$, she forms the inference $r_i(\omega,\tilde{\omega}) = \emptyset$.

So far we have explained how each player i makes an inference $r_i(\omega,\tilde{\omega})$ for each pair $(\omega,\tilde{\omega})$. At the end of the learning round, she summarizes all these inferences and makes a "final inference" $\omega(i) \in \Omega \cup \{\emptyset\}$. Specifically, we set $\omega(i) = \omega$ if $r_i(\omega,\tilde{\omega}) = \omega$ for all $\tilde{\omega} \neq \omega$. In words, player i's final inference is $\omega(i) = \omega$ if the state ω beats all the other states $\tilde{\omega} \neq \omega$ in the relevant comparisons. If such ω does not exist, then we set $\omega(i) = \emptyset$.

It is easy to see that Lemma 2 still holds in this environment:

Lemma 27. *Suppose that Condition 3 holds. Then there is* $\overline{T}$ *such that for any* $T > \overline{T}$, ω, $\tilde{\omega} \neq \omega$, *and* h_i^T *such that* $|f_i(h_i^T) - \pi_i^{\tilde{\omega}}(a^*)| < \varepsilon$, *we have*

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

To interpret this lemma, suppose that player i's final inference is $\omega(i) = \tilde{\omega}$. Then we must have $r_i(\omega,\tilde{\omega}) = \tilde{\omega}$, and thus Lemma 26(iii) implies that $|f_i(h_i^T) - \pi_i^{\tilde{\omega}}(a^*)| < \varepsilon$, where h_i^T is player i's history during $T(i,\omega,\tilde{\omega})$. Then from the lemma above, player i must believe that "If my inference is wrong and the true state is ω, then the opponents' signal frequency during the interval $T(i,\omega,\tilde{\omega})$ must be also distorted and not close to $\pi_{-i}^{\omega}(a^*)$." As in Section 4, this property plays a crucial role in order to induce the truthful summary report.

B.3 Equilibrium Conditions

We have specified the block strategies $s_i^{x_i}$, so what remains is to find the transition rules ρ_i in such a way that the resulting automaton strategy is an equilibrium. Formally, as in the two-player case, we will choose the transition rules ρ_i which satisfy both the promise-keeping condition and the incentive-compatibility condition. The promise-keeping condition requires

$$\overline{v}_i^\omega = (1-\delta^T)\sum_{t=1}^{T_b}\delta^{t-1}E[g_i^\omega(a^t)|\omega,s^x] + \delta^{T_b}\left\{\overline{v}_i^\omega - E[\rho_{i-1}^\omega(B|G,h_{i-1}^{T_b})|\omega,s^x](\overline{v}_i^\omega - \underline{v}_i^\omega)\right\} \tag{29}$$

for each ω, i, and x with $x_{i-1}^\omega = G$, and

$$\underline{v}_i^\omega = (1-\delta^T)\sum_{t=1}^{T_b}\delta^{t-1}E[g_i^\omega(a^t)|\omega,s^x] + \delta^{T_b}\left\{\underline{v}_i^\omega + E[\rho_{i-1}^\omega(G|B,h_{i-1}^{T_b})|\omega,s^x](\overline{v}_i^\omega - \underline{v}_i^\omega)\right\} \tag{30}$$

for each ω, i, and x with $x_{i-1}^\omega = B$. These conditions assert that player i's repeated-game payoff is determined by the plan x_{i-1} chosen by player $i-1$. Specifically, player i's payoff is $\overline{v}_i^\omega$ if player $i-1$ plans to reward player i, while the payoff is $\underline{v}_i^\omega$ if player $i-1$ plans to punish player i.

The incentive-compatibility condition requires that

$$\begin{aligned}&(1-\delta^{T_b-t})\sum_{\tau=t+1}^{T_b}\delta^{\tau-1}\left(E[g_i^\omega(a^\tau)|\omega,s_i^{T_b},s_{-i}^{x_{-i}},h_i^t] - E[g_i^\omega(a^\tau)|\omega,s^x,h_i^t]\right)\\ &\leq \delta^{T_b-t}\left(E[\rho_{i-1}^\omega(B|G,h_{i-1}^{T_b})|\omega,s_i^{T_b},s_{-i}^{x_{-i}},h_i^t] - E[\rho_{i-1}^\omega(B|G,h_{i-1}^{T_b})|\omega,s^x,h_i^t]\right)(\overline{v}_i^\omega - \underline{v}_i^\omega)\end{aligned} \tag{31}$$

for each ω, i, $s_i^{T_b}$, h_i^t, and x with $x_{i-1}^\omega = G$, and

$$\begin{aligned}&(1-\delta^{T_b-t})\sum_{\tau=t+1}^{T_b}\delta^{\tau-1}\left(E[g_i^\omega(a^\tau)|\omega,s_i^{T_b},s_{-i}^{x_{-i}},h_i^t] - E[g_i^\omega(a^\tau)|\omega,s^x,h_i^t]\right)\\ &\leq \delta^{T_b-t}\left(E[\rho_{i-1}^\omega(B|B,h_{i-1}^{T_b})|\omega,s_i^{T_b},s_{-i}^{x_{-i}},h_i^t] - E[\rho_{i-1}^\omega(B|B,h_{i-1}^{T_b})|\omega,s^x,h_i^t]\right)(\overline{v}_i^\omega - \underline{v}_i^\omega)\end{aligned} \tag{32}$$

for each ω, i, $s_i^{T_b}$, h_i^t, and x with $x_{i-1}^\omega = B$. That is, deviating to any other block strategy $s_i^{T_b} \neq s_i^{x_i}$ is not profitable, regardless of the state ω and the past history h_i^t.

As discussed in Section 4, if the transition rules satisfy the above conditions, then the resulting automaton is indeed an ex-post equilibrium, and the payoff v is achievable by choosing the initial automaton state carefully. So in what follows, we will find such transition rules ρ_i.

B.4 Complete-Information Transfer Game

As discussed in Section 4, finding appropriate transition rules is equivalent to finding appropriate "transfer rules," as continuation payoffs after the block can play a role like that of transfers in the mechanism design. So consider the complete-information transfer game in which (i) a state ω is given and common knowledge and (ii) after T_b periods, the game ends and player i receives a transfer according to some transfer rule $U_i : H_{i-1}^{T_b} \to \boldsymbol{R}$. Note that the amount of the transfer depends only on player $(i-1)$'s history $h_{i-1}^{T_b}$. This restriction comes from the fact that player i's continuation payoff, which is represented by the second terms of the right-hand sides of (29) and (30) depends only on $h_{i-1}^{T_b}$. Let $G_i^{\omega}(s^{T_b}, U_i)$ denote player i's expected payoff in this auxiliary scenario game, when players play s^{T_b}. Also, for each history h_i^t with $t \leq T_b$, let $G_i^{\omega}(s^{T_b}, U_i, h_i^t)$ denote player i's payoff in the continuation game after history h_i^t. Our goal in this subsection is to prove the following two lemmas. The first lemma is:

Lemma 28. *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 ω, there is a transfer rule $U_i^{\omega,G} : H_{i-1}^{T_b} \to \boldsymbol{R}$ 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 such that $x_{i-1}^{\omega} = G$.*

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

To interpret this lemma, suppose that the opponents play the block strategy $s_{-i}^{x_{-i}}$ with $x_{i-1}^{\omega} = G$. That is, player $i-1$ plans to rewards player i at state ω. Clauses (i) and (ii) in the above lemma ensures that there is a transfer rule $U_i^{\omega,G}$ such that playing the prescribed block strategy $s_i^{x_i}$ is a best reply for player i and yields the payoff $\overline{v}_i$. Clause (iii) requires that this transfer be non-negative and bounded by $\frac{\overline{v}_i^{\omega} - \underline{v}_i^{\omega}}{1-\delta}$.

Once we have this lemma, we can construct a transition rule $\rho_{i-1}^{\omega}(\cdot | G, h_{i-1}^{T_b})$ which satisfies the desired properties (29) and (31), by setting

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

for each $h_{i-1}^{T_b}$.

The second lemma is a counterpart to the above lemma, and considers the case in which player $i-1$ plans to punish player i.

Lemma 29. *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 ω, there is a transfer rule $U_i^{\omega,B} : H_{i-1}^{T_b} \to \boldsymbol{R}$ 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 $x_{i-1}^{\omega} = B$.*

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

Again, once we have this lemma, we can construct a transition rule $\rho_{i-1}^{\omega}(\cdot|B, h_{i-1}^{T_b})$ which satisfies the desired properties (30) and (32), by setting

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

So in order to complete the proof of Proposition 1, it is sufficient to prove the above two lemmas.

B.5 Proof of Lemma 28

As in the proof of Lemma 3, we first construct a transfer rule $\tilde{U}_i^{\omega,G}$ which "approximately" satisfies clause (ii) of the lemma. That is, we will construct $\tilde{U}_i^{\omega,G}$ such that playing the prescribed strategy $s_i^{x_i}$ is an approximate best reply for player i in the summary report round, and is an exact best reply in other rounds. After that, we modify this transfer rule $\tilde{U}_i^{\omega,G}$ and construct $U_i^{\omega,G}$ which satisfies clause (ii) exactly. Then we show that the this transfer rule $U_i^{\omega,G}$ satisfies clauses (i) and (iii).

Recall that in the detailed report round, each player $j \neq i$ reports her signal sequence $(z_j^t)_{t \in T(i,\omega,\tilde{\omega})}$ observed during the T-period interval $T(i,\omega,\tilde{\omega})$. Let $(\hat{z}_j^t)_{t \in T(i,\omega,\tilde{\omega})}$ denote the reported signal sequence, and let $\hat{f}_{-i}^{(\omega,\tilde{\omega})} \in \triangle Z_{-i}$ denote the signal frequency computed from the reported signals $(\hat{z}_{-i}^t)_{t \in T(i,\omega,\tilde{\omega})}$. Due to signal correlation, this signal frequency $\hat{f}_{-i}^{(\omega,\tilde{\omega})}$ is informative about player i's signals during $T(i,\omega,\tilde{\omega})$.

A block history $h_{i-1}^{T_b}$ is *regular given* (ω, G) if it satisfies all the following conditions:

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

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

(G3) Player $i-1$ reports $x_{i-1}^{\omega} = G$ 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 opponents' detailed report $\hat{f}_{-i}$ satisfies $\left|\pi_{-i}^{\omega}(a^*) - \hat{f}_{-i}^{(\omega,\tilde{\omega})}\right| < \varepsilon$ for all $\tilde{\omega} \neq \omega$

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

As in the two-player case, the last condition (G5) requires that player i's summary report be consistent with the opponents' signals during the learning round. If $\hat{f}_{-i}^{(\omega,\tilde{\omega})}$ is close to the true distribution $\pi_{-i}^{\omega}(a^*)$ at state ω, then the opponents believe that conditional on the state ω, player i's signal frequency during $T(i,\omega,\tilde{\omega})$ is also close to the true distribution $\pi_i^{\omega}(a^*)$ at state ω. Thus, if $\hat{f}_{-i}^{(\omega,\tilde{\omega})}$ is close to the true distribution $\pi_{-i}^{\omega}(a^*)$ for each $(\omega,\tilde{\omega})$, then the opponents believe that player i's signal frequency is close to the true distribution in each T-period interval within player i's learning round, in which case player i's inference is indeed $\omega(i) = \omega$ or $\omega(i) = \emptyset$.

B.5.1 Step 1: Construction of $\tilde{U}_i^{\omega,G}$

Choose a transfer rule $\tilde{U}_i^{\omega,G} : H_{i-1}^{T_b} \to \boldsymbol{R}$ such that

- For each regular history $h_{i-1}^{T_b}$, choose $\tilde{U}_i^{\omega,G}(h_{i-1}^{T_b})$ so that it solves

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

- For each irregular history $h_{i-1}^{T_b}$, choose $\tilde{U}_i^{\omega,G}(h_{i-1}^{T_b})$ so that

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

In words, after regular histories (which requires player i not to deviate from the prescribed strategy and not to report a wrong state), we set the transfer $\tilde{U}_i^{\omega,G}$ so that player i's average payoff in the complete-information transfer game is equal to $\overline{v}_i^{\omega}$. After irregular histories, we choose a huge negative transfer $\tilde{U}_i^{\omega,G}$ so that player i's payoff goes down to $-2\overline{g}_i^{\omega}$.

B.5.2 Step 2: $\tilde{U}_i^{\omega,G}$ approximately satisfies clause (ii)

With the transfer rule $\tilde{U}_i^{\omega,G}$ above, playing the prescribed strategy $s_i^{x_i}$ is an approximate best reply for player i in the summary report round, and is an exact best reply in other rounds. The proof is very similar to the one for Lemma 3, and hence omitted.

B.5.3 Step 3: Construction of $U_i^{\omega,G}$ and Clause (ii)

The transfer rule $\tilde{U}_i^{\omega,G}$ approximately satisfies Lemma 3(ii), but not exactly. The reason is that the truthful report of $\omega(i)$ in the summary report round is not an exact best reply at some histories. Specifically, if player i's inference is $\omega(i) = \tilde{\omega} \neq \omega$ (i.e., her inference is incorrect), then the truthful report is not an exact best reply.

So in order to satisfy (ii) exactly, we need to modify the transfer rule $\tilde{U}_i^{\omega,G}$. As in the proof of Lemma 4, the idea is to give a "bonus" to player i when she reports $\omega(i) = \tilde{\omega}$. This gives her an extra incentive to report $\omega(i) = \tilde{\omega}$ truthfully.

Recall that in the detailed report round, player i reports her signal sequence $(z_i^t)_{t\in T(i,\omega,\tilde{\omega})}$ during her own learning round. Let $(\hat{z}_i^t)_{t\in T(i,\omega,\tilde{\omega})}$ denote the reported signal sequence, and let $\hat{f}_i^{(\omega,\tilde{\omega})} \in \triangle Z_i$ denote the signal frequency computed from this sequence. Let $\Pr(f_{-i}|\omega,a^*,\cdots,a^*,f_i)$ denote the conditional probability of the opponents' signal frequency over T periods being f_{-i}, given that the true state is ω, no one deviates from a^* each period, and player i's signal frequency is f_i. Let

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

Then define a bonus function $b_i^\omega : H_{i-1}^{T_b} \to \boldsymbol{R}$ as

$$b_i^\omega(h_{i-1}^{T_b}) = \begin{cases} 0 & \text{if player } i \text{ reports } \omega(i) = \omega \text{ or } \omega(i) = \emptyset \\ 0 & \text{if someone deviates in the learning round} \\ 0 & \text{if some } j \neq i \text{ reports } \omega(j) \neq \omega \\ 0 & \text{if player } i \text{ reports } \omega(i) = \tilde{\omega} \text{ and } |\hat{f}_i^{(\omega,\tilde{\omega})} - \pi_i^{\tilde{\omega}}(a^*)| > \varepsilon \\ (\overline{v}_i^\omega + 2\overline{g}_i^\omega)\prod_{\tilde{\omega}\neq\omega} p_i^\omega(\hat{f}_i^{(\omega,\tilde{\omega})}) & \text{otherwise} \end{cases}.$$

This is an extension of the bonus function for the two-player case. To interpret the condition $|\hat{f}_i^{(\omega,\tilde{\omega})} - \pi_i^{\tilde{\omega}}(a^*)| > \varepsilon$, note that from Lemma 26(iii), if player i's inference is $\omega(i) = \tilde{\omega}$, then her signal frequency in the T-period interval $T(i,\omega,\tilde{\omega})$ must be close to the true distribution $\pi_i^{\tilde{\omega}}(a^*)$ at state $\tilde{\omega}$. So if she reports $\omega(i) = \tilde{\omega}$ in the summary report round but reports $|\hat{f}_i^{(\omega,\tilde{\omega})} - \pi_i^{\tilde{\omega}}(a^*)| > \varepsilon$ in the detailed report round, it must be a consequence of player i's misreport, either in the summary report round or the detailed report round (or both). We do not pay a bonus in such a case.

As in the proof of Lemma 3, we can show that the amount of the bonus is small, that is, $b_i^\omega(h_{i-1}^{T_b}) < 3\overline{g}_i^\omega \exp(-T^{\frac{1}{2}})$ for sufficiently large T.

Now we are ready to define the modified transfer rule $U_i^{\omega,G}$ which satisfies Lemma 3(ii) exactly. 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, $e(z_{-i})$ denote the $|Z_{-i}|$-dimensional column vector where the component corresponding to z_{-i} is one and the remaining components are zero. Then define the transfer rule $U_i^{\omega,G}$ as

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

where c^G is a constant term which will be specified later.

This transfer rule satisfies Lemma 28(ii). The proof is very similar to the one for Lemma 3, and hence omitted. (Note that for each $t \in T(i)$, player i reports z_i^t before the opponents report z_{-i}^t; this ensures that the expected value of $\left|e(\hat{z}_{-i}^t) - C_i^\omega(a^*)e(\hat{z}_i^t)\right|^2$ is minimized by the truthful report of z_i^t.) Also, as in the proof of Lemma 3, we can find a constant term c^G such that the resulting transfer rule $U_i^{\omega,G}$ satisfies clauses (i) and (iii).

B.6 Proof of Lemma 29

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

As in the proof of Lemma 4, it is useful to introduce the notion of *regular histories*. Player $(i-1)$'s block history $h_{i-1}^{T_b}$ is *regular given* (ω, B) if it satisfies all the following conditions:

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

(B2) Players $-i$ report ω in the summary report round.

(B3) Player $i-1$ reports $x_{i-1}^{\omega} = B$ in the first period of the main round.

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

In short, the history $h_{i-1}^{T_b}$ is regular if the opponents play the prescribed strategy $s_{-i}^{x_{-i}}$ with $x_{i-1}^{\omega} = B$ and they learn the true state ω correctly. A history $h_{i-1}^{T_b}$ is *irregular given* (ω, B) if it is not regular.

B.6.1 Step 1: Construction of $U_i^{\omega,B}$

Let $\overline{g}_i^{\omega} = \max_{a\in A} |g_i^{\omega}(a)|$, and let $c^B > 0$ be a constant which will be specified later. Choose a transfer rule $U_i^{\omega,B} : H_{i-1}^{T_b} \to \boldsymbol{R}$ such that

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

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

  where τ is the number of periods such that player i deviated from a^* during the learning round.

- For each irregular history $h_{i-1}^{T_b}$, choose $U_i^{\omega,B}(h_{i-1}^{T_b})$ so that

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

In words, if the opponents play the prescribed strategies and learn the true state ω correctly (so that the history $h_{i-1}^{T_b}$ is regular), we adjust the transfer $U_i^{\omega,B}$ so that player i's payoff in the complete-information transfer game equals $\underline{v}_i^{\omega} - c^B$. If not, we give a huge positive transfer to player i so that her payoff goes up to $2\overline{g}_i^{\omega} - c^B$. If player i deviates in the learning round, her payoff decreases due to the term $\frac{\tau\varepsilon}{T}$.

B.6.2 Step 2: Proof of Clause (ii)

The transfer rule $U_i^{\omega,B}$ above satisfies Lemma 29(ii), that is, deviating to any block strategy $s_i^{T_b} \neq s_i^x$ is not profitable. The actual proof is very similar to the one for Lemma 4, and hence omitted.

B.6.3 Step 3: Proof of Clauses (i) and (iii)

Now we choose the constant term c^B such that the resulting transfer rule $U_i^{\omega,B}$ satisfies clauses (i) and (iii) of Lemma 29.

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

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

Then the resulting transfer rule $U_i^{\omega,B}$ satisfies clauses (i) and (iii) of Lemma 29. The proof is very similar to the one for Lemma 4, and hence omitted.

B.7 When $|A_i| < |Z_i|$ for Some i

So far we have assumed $|A_i| \geq |Z_i|$ for all i. This ensures that during the detailed report round, each player i can reveal her signal z_i by choosing some action a_i. If this assumption is not satisfied, player i needs to spend more than one period in order to reveal her signal, and it causes some complications on player i's incentive at off-path histories.[20]

[20] To illustrate the issue, suppose that player i had to choose an action sequence (a_i^1, a_i^2) in order to reveal her signal z_i, but she had deviated from a_i^1 in the first period of the reporting phase. In such a case, choosing a_i^2 in the next period need not be a best reply, even if we use the transfer rule $U_i^{\omega,G}$ defined in the proof of Lemma 28. Also her best reply depends on the state ω in such a history.

To fix this problem, we make the following two changes to the structure of the detailed report round. Pick a natural number K such that $K \geq \log_{|A_i|}|Z_i|$ for each i.

- Each player i uses a sequence of actions $(a_i^1, \cdots, a_i^K)$ (rather than a single action) to report a signal z_i.
- Each player reports the same information $|\Omega|$ times (rather than once). Specifically, after the main round, there is a "detailed report round for ω_1," in which each player reports her history during the learning round. Then there is a "detailed report round for ω_2," and players report the same information again. Then there is a "detailed report round for ω_3," a "detailed report round for ω_4," and so on.

Note that for each ω, the detailed report round for ω consists of $N^2KT\frac{|\Omega|(|\Omega|-1)}{2}$ periods. So we have $N^2KT\frac{|\Omega|^2(|\Omega|-1)}{2}$ periods in total.

Choose a block strategy $s_i^{x_i}$ so that

- The play in the learning, summary report, and main rounds is exactly the same as in the case with $|A_i| \geq |Z_i|$.
- For each ω, in the detailed report round for ω,
  - Player i reports her history truthfully, as long as she has not deviated in some earlier period within the round. (In particular, she reports truthfully even if she has deviated in the learning, summary report, or main round.)
  - If player i has deviated within the round, she may choose other actions, which will be specified later.

Note that the block strategy above induces the same play as the one in the case with $|A_i| \geq |Z_i|$, except histories which are reached after player i's own deviation.

In what follows, we will explain that Lemmas 28 and 29 still hold, if we specify the transfer rules and actions at off-path histories appropriately. This completes the proof, because it ensures that the resulting automaton strategy satisfies the promise-keeping condition (29) and (30), and the incentive compatibility condition (31) and (32).

It is easy to see that Lemma 29 still holds, because the communication in the detailed report round plays no role in the proof of the lemma. Indeed, if we choose the transfer rule $U_i^{\omega,B}$ as in Appendix B.6.1 it satisfies clauses (i) through (iii). (We have not specified actions for some off-path histories in the detailed report round, but this is not a problem because the transfer rule $U_i^{\omega,B}$ makes player i indifferent over all actions during the detailed report round. That is, regardless of the specification of actions in these off-path histories, clause (ii) holds.)

The proof of Lemma 28 needs a minor modification. Fix ω, and fix a transfer rule $\tilde{U}_i^{\omega,G}$ as in Appendix B.5.1. Define the bonus function b_i^{ω} and the transfer rule $U_i^{\omega,G}$ as in Appendix B.5.3, using the information exchanged in the detailed report round for ω. That is, $\hat{z}_i$ and $\hat{f}_i$ which appear in the bonus function b_i^{ω} and the additional terms in $U_i^{\omega,G}$ are player i's report in the detailed report round for ω. (The information exchanged in the detailed report round for $\tilde{\omega} \neq \omega$ is ignored, when we define these terms.) Complete the specification of the strategy $s_i^{x_i}$ (choose actions at off-path histories in the report round) so that for each ω, in the detailed report round for ω, if player i has deviated within that round, let her choose an action which maximizes her payoff in the complete-information game with $(\omega, U_i^{\omega,G})$.

With this modification, clauses (i) through (iii) of Lemma 28 are satisfied. Indeed, clause (i) holds because the play on the equilibrium path is the same as in the previous case. Clause (ii) also holds, because in the complete-information game with $(\omega, U_i^{\omega,G})$,

- The incentive problem in the learning, summary report, and main rounds is the same as in the case with $|A_i| \geq |Z_i|$.
- In the detailed report round for ω, player i chooses a best reply after every history.
- In the detailed report round for $\tilde{\omega} \neq \omega$, player i is indifferent over all actions.

Clause (iii) can be verified as in the case with $|A_i| \geq |Z_i|$.

Appendix C: Common Learning

In this appendix, we will provide the formal statement of Proposition 2 and its proof. That is, we will prove that in our equilibria, common learning occurs and the state ω becomes approximate common knowledge in the sense of Monderer and Samet (1989). For now, we assume full support, i.e., $\pi^{\omega}(z|a) > 0$ for each ω, a, and z. Also we assume that there are only two players. These assumptions are not necessary to obtain the result, but they considerably simplifies our exposition. (See Cripps, Ely, Mailath, and Samuelson (2008) for how to extend the theorem to the case in which there are more than two players and/or the full support assumption is not satisfied.)

Fix a target payoff $v \in \text{int}V^*$, fix a sufficiently large δ, and construct an ex-post equilibrium s as in Section 4. Given a common prior $\mu \in \triangle\Omega$, this equilibrium s induces a probability measure on the set of outcomes $\Xi = \Omega \times (A_1 \times A_2 \times Z_1 \times Z_2)^{\infty}$, where each outcome $\xi = (\omega, (a_1^t, a_2^t, z_1^t, z_2^t)_{t=1}^{\infty}) \in \mathscr{H}$ specifies the state of the world ω and the actions and signals in each period. We use $P \in \triangle\mathscr{F}$ to denote this measure, and use $E[\cdot]$ to denote expectations with respect to this measure. Also, let P^{ω} denote the measure conditional on a given state ω, and let $E^{\omega}[\cdot]$ denote expectations with respect to this measure.

Recall that the set of t-period histories of player i is $H_i^t = (a^{\tau}, z_i^{\tau})_{\tau=1}^{t}$. Let $\{\mathscr{H}_i^t\}_{t=1}^{\infty}$ denote the filtration induced on ξ by player i's histories. For any event $F \subset \Xi$, the ($\mathscr{H}_i^t$-measurable) random variable $E[1_F|\mathscr{H}_i^t]$ is the probability that player i attaches to the event F given her information after period t. Let

$$B_i^{t,q}(F) = \{\xi \in \Xi \mid E[1_F|\mathscr{H}_i^t](\xi) \geq q\},$$

that is, $B_i^{t,q}(F)$ is the set of outcomes ξ where player i attaches at least probability q to event F after period t. Following Cripps, Ely, Mailath, and Samuelson (2008), we say that player i *individually learns the true state* if for each ω and $q \in (0,1)$, there is t^* such that for any $t > t^*$,

$$P^{\omega}(B_i^{t,q}(\{\omega\})) > q,$$

where $\{\omega\}$ denotes the event that the true state is ω.

An event $F \subset \Xi$ is *q-believed* after period t if each player attaches at least probability q to event F. Let $B^{t,q}(F) = B_1^{t,q}(F) \cap B_2^{t,q}(F)$, that is, $B^{t,q}(F)$ is the event that F is q-believed after period t. An event $F \subset \Xi$ is *common q-belief* after period t if F is q-believed, and this event $B^{t,q}(F)$ is q-believed, and this event $B^{t,q}(B^{t,q}(F))$ is q-believed, and so on. Formally, the event that F is common

q-belief after period t is denoted by

$$\mathscr{B}^{t,q}(F) = \bigcap_{n\geq 1} [B^{t,q}]^n(F).$$

Following Cripps, Ely, Mailath, and Samuelson (2008), we say that players *commonly learn the true state* if for each ω and $q \in (0,1)$, there is t^* such that for any $t > t^*$,

$$P^{\omega}(\mathscr{B}^{t,q}(\{\omega\})) > q.$$

The following proposition is a formal version of Proposition 2.

Proposition 6. *Players commonly learn the true state in the equilibrium s.*

In out setup, each player updates her belief about the opponent's signals through two information channels. The first informational channel is private signals. Since signals may be correlated across players, one's private signal may have noisy information about the opponent's signal. The second informational channel is the opponent's actions; since there is a correlation between the opponent's signals and actions, each player can learn the opponent's signals through the action by the opponent. We need to take into account both these effects in order to prove the proposition.

The proof idea is roughly as follows. We begin with considering how signals in the learning rounds influence players' (higher-order) beliefs. To do so, suppose hypothetically that suppose that players observe private signals in the learning rounds only, and do not observe signals in the summary report, main, and detailed report rounds.[21] In our equilibrium, all these signals are publicly revealed in the detailed report rounds, i.e., players' private histories become public information at the end of each block game. This implies that common learning happens if players do not observe signals in the summary report, main, and detailed report rounds.

Next, we consider our original model and investigate what happens if players observe signals in the summary report, main, and detailed report rounds. Since these signals do not influence actions in later periods, the second information channel does not exist, that is, a player can learn the opponent's signal in these

[21] Note that our equilibrium strategy is still an equilibrium in this new setup, as signals in the summary report, main, and detailed report rounds do not influence players' continuation play.

rounds only through the correlation of private signals. Hence the inference problem here reduces to the one considered by Cripps, Ely, Mailath, and Samuelson (2008), and we can apply their result to show that common learning happens if we restrict attention to the effect of signals in these rounds. Taken together, we can conclude that players commonly learn the state in our equilibrium. The formal proof is as follows.

Proof. Given a period t, let $T^{\text{learning}}(t)$ denote the set of periods included in the learning rounds of the past block games. (So $T^{\text{learning}}(t)$ does not include the periods in the learning round of the current block game.) Likewise, let $T^{\text{others}}(t)$ denote the set of periods included in the summary report, main, or detailed report rounds of the past block games. Note that the union $T^{\text{learning}}(t) \cup T^{\text{others}}(t)$ denote the set of periods in the past block game, i.e., $T^{\text{learning}}(t) \cup T^{\text{others}}(t) = \{1, \cdots, kT_b\}$ where k is an integer satisfying $kT_b < t \leq (k+1)T_b$.

By the construction of the equilibrium strategy, players have played the action profile a^* in all the periods in the set $T^{\text{learning}}(t)$, and all the signal profiles in these periods are common knowledge thanks to the communication in the detailed report rounds. For each outcome ξ, let $f^{\text{learning}}(t)[\xi] \in \triangle Z$ denote the empirical distribution of signal profiles z in these periods. We will often omit $[\xi]$ when the meaning is clear. Let $F^{\omega,\text{learning}}(t)$ denote the event that the empirical distribution $f^{\text{learning}}(t)$ is η-close to the true distribution at state ω, i.e.,

$$F^{\omega,\text{learning}}(t) = \{\xi \mid |f^{\text{learning}}(t) - \pi^{\omega}(a^*)| < \eta\}.$$

In the periods in the set $T^{\text{others}}(t)$, players' actions are contingent on the past histories and hence random. Let $A^* \subseteq A$ be the set of action profiles which can be chosen in the summary report, main, or detailed report round with positive probability on the equilibrium path. Then given any outcome ξ, let $\{T^{\text{others}}(t,a)[\xi]\}_{a \in A^*}$ be the partition of $T^{\text{others}}(t)$ with respect to the chosen action profile a, that is, $T^{\text{others}}(t,a)[\xi]$ is the set of the periods in $T^{\text{others}}(t)$ where players played the profile a according to the outcome ξ. Let $f_i(t,a)[\xi]$ be the empirical distribution of player i's signals z_i in the periods in the set $T^{\text{others}}(t,a)[\xi]$, i.e., $f_i(t,a)$ is the empirical distribution of z_i during the periods where players chose the action profile a. Let $F_i^{\omega,1}(t,a)$ be the event that this empirical distribution $f_i(t,a)$ is η-close to the true distribution at state ω, i.e.,

$$F_i^{\omega,1}(t,a) = \{\xi \mid |f_i(t,a) - \pi_i^{\omega}(a)| < \eta\}.$$

Also, let $F_i^{\omega,2}(t,a)$ be the event that player i's estimate (expectation) about the opponent's signal frequency in these periods is close to the true distribution at state ω:

$$F_i^{\omega,2}(t,a) = \{\xi \mid |C_i^{\omega}(a)f_i(t,a) - \pi_j^{\omega}(a)| < \eta - \eta^2\}.$$

Let

$$F_i^{\omega}(t,a) = F_i^{\omega,1}(t,a) \cap F_i^{\omega,2}(t,a).$$

and let

$$F^{\omega,\text{others}}(t) = \bigcap_i \bigcap_{a \in A^*} F_i^{\omega}(t,a).$$

In words, $F^{\omega,\text{others}}(t)$ is the event that for each set of periods $T^{\text{others}}(t,a)$, each player's signal frequency is close to the true distribution at state ω, and her estimate about the opponent's signal frequency is also close to the true distribution at state ω.

Given a natural number τ, let $G(t,\tau)$ denote the event that each action profile $a \in A^*$ is chosen at least τ times in $T^{\text{others}}(t)$, that is,

$$G(t,\tau) = \{\xi \mid |T^{\text{others}}(t,a)| \geq \tau \ \ \forall a \in A^*\}.$$

Then let

$$F^{\omega}(t,\tau) = G(t,\tau) \cap F^{\omega,\text{learning}}(t) \cap F^{\omega,\text{others}}(t).$$

In the following, we will take large t and τ, and hence on the event $G(t,\tau)$, the sets $T^{\text{learning}}(t)$ and $T^{\text{others}}(t,a)$ contain sufficiently many periods. Roughly, this implies that on the event $F^{\omega}(t,\tau)$, (i) the signals in $T^{\text{learning}}(t)$, which are common knowledge among players, reveal that the true state is almost surely ω, (ii) each player's signals in $T^{\text{others}}(t,a)$ reveal that the true state is almost surely ω, and (iii) each player expects that the opponent's signals in $T^{\text{others}}(t,a)$ also reveal that the true state is almost surely ω.

We will establish three lemmas, which are useful to prove Proposition 6. The first lemma shows that on this event $F^{\omega}(t,\tau)$, each player is almost sure that the true state is ω when t and τ are sufficiently large.

Lemma 30. *When η is sufficiently small, for any $q \in (0,1)$, there is t^* and τ such that for any $t > t^*$ and ω, $F^{\omega}(t,\tau) \subseteq B^{t,q}(\{\omega\})$.*

Proof. Let $\mu^t(\omega|h_i^t) = E[1_{\{\omega\}}|h_i^t]$, that is, $\mu^t(\omega|h_i^t)$ is player i's belief on ω after history h_i^t. Given h_i^t, let h_i^{past} and h_i^{current} denote the histories in the past block game and the current block game, respectively. The discussion after Proposition 6 shows that for each ω and $\tilde{\omega} \neq \omega$, we have

$$\frac{\mu^t(\tilde{\omega}|h_i^t)}{\mu^t(\omega|h_i^t)} = \frac{\mu(\tilde{\omega})}{\mu(\omega)} \left(\prod_{\tilde{t} \in T^{\text{learning}}(t)} \frac{\pi^{\tilde{\omega}}(z^{\tilde{t}}|a^*)}{\pi^{\omega}(z^{\tilde{t}}|a^*)} \right) \\ \times \left(\prod_{a \in A^*} \prod_{\tilde{t} \in T^{\text{others}}(t,a)} \frac{\pi_i^{\tilde{\omega}}(z_i^{\tilde{t}}|a)}{\pi_i^{\omega}(z_i^{\tilde{t}}|a)} \right) \frac{\Pr(h_i^{\text{current}}|\tilde{\omega}, h_i^{\text{past}})}{\Pr(h_i^{\text{current}}|\omega, h_i^{\text{past}})}$$

where $\Pr(h_i^{\text{current}}|\omega, h_i^{\text{past}})$ denotes the probability that h_i^{current} occurs given ω and h_i^{past}.

Take a sufficiently small $\eta > 0$. Since $F^{\omega}(t,\tau) \subset F^{\omega,\text{learning}}(t)$, it follows from Lemma 1 of Cripps, Ely, Mailath, and Samuelson (2008) that on the event $F^{\omega}(t,\tau)$, the term in the first set of parenthesis in the right-hand side converges to zero as $t \to \infty$. Similarly, since $F^{\omega}(t,\tau) \subset F_i^{\omega,1}(t,a)$, it follows that on the event $F^{\omega}(t,\tau)$, for any small $\gamma > 0$ there is t^* and τ such that for any $t > t^*$, we have

$$\prod_{\tilde{t} \in T^{\text{others}}(t,a)} \frac{\pi_i^{\tilde{\omega}}(z_i^{\tilde{t}}|a)}{\pi_i^{\omega}(z_i^{\tilde{t}}|a)} < \gamma$$

for each a satisfying $\pi_i^{\omega}(a) \neq \pi_i^{\tilde{\omega}}(a)$. Also it is obvious that for each a satisfying $\pi_i^{\omega}(a) = \pi_i^{\tilde{\omega}}(a)$,

$$\prod_{\tilde{t} \in T^{\text{others}}(t,a)} \frac{\pi_i^{\tilde{\omega}}(z_i^{\tilde{t}}|a)}{\pi_i^{\omega}(z_i^{\tilde{t}}|a)} = 1.$$

Finally, since T_b is fixed, the term after the second set of parenthesis in the right-hand side is bounded from above by some constant. (Note that the probability distribution of x_{-i} in the current block game conditional on $(h_i^{\text{past}}, \omega)$ is the same as that conditional on $(h_i^{\text{past}}, \tilde{\omega})$ since x_{-i} is determined by the action profiles in the past block games and by the signals in the past learning rounds, which are encoded in h_i^{past}.) Taken together, we can conclude that the likelihood $\frac{\mu^t(\tilde{\omega}|h_i^t)}{\mu^t(\omega|h_i^t)}$ is close to zero on the event $F^{\omega}(t,\tau)$, when t and τ are large enough. This proves the lemma. *Q.E.D.*

The second lemma shows that for any τ, the event $F^{\omega}(t,\tau)$ occurs with probability close to one if the true state is ω and t is sufficiently large.

Lemma 31. *For any $\eta \in (0,1)$, τ, and $q \in (0,1)$, there is t^* such that for any $t > t^*$ and ω, $P^{\omega}(F^{\omega}(t,\tau)) > q$.*

Proof. This directly follows from the law of large numbers. Note that there can be $a \in A^*$ which is chosen only when someone make a wrong inference about ω in the learning round and/or players choose a particular automaton state x; but this does not cause any problem because such an event occurs for sure in the long run. *Q.E.D.*

The last lemma shows that the event $\tilde{F}^{\omega}(t,\tau) = \{\omega\} \cap F^{\omega}(t,\tau)$ is *q-evident* in the sense that $\tilde{F}^{\omega}(t,\tau) \subseteq B^{t,q}(\tilde{F}^{\omega}(t,\tau))$.

Lemma 32. *When η is sufficiently small, for any τ, and $q \in (0,1)$, there is t^* such that for any $t > t^*$ and ω, $\tilde{F}^{\omega}(t,\tau) \subseteq B^{t,q}(\tilde{F}^{\omega}(t,\tau))$.*

Proof. It is obvious that $F^{\omega,\text{learning}}(t) \subseteq B^{t,q}(F^{\omega,\text{learning}}(t))$. So it is sufficient to show that $\{\omega\} \cap G(t,\tau) \cap F_i^{\omega}(t,a) \subseteq B_i^{t,q}(\{\omega\} \cap G(t,\tau) \cap F^{\omega}(t,a))$ for each i and $a \in A^*$.

Let $\hat{F}_i^{\omega}(t,a) = \{\xi \mid |C_i^{\omega}(a^*) f_i(t,a) - \pi_j^{\omega}(a^*)| < \eta^2\}$, that is, $\hat{F}_i^{\omega}(t,a)$ is the event that player j's realized signal frequency in $T^{\text{others}}(t,a)$ is close to player i's estimate. The triangle inequality yields

$$F_i^{\omega,2}(t,a) \cap \hat{F}_i^{\omega}(t,a) \subseteq F_j^{\omega,1}(t,a). \tag{33}$$

Let $C_{ij}^{\omega}(a^*) = C_j^{\omega}(a^*) C_i^{\omega}(a^*)$. Since we assume full support, this matrix $C_{ij}^{\omega}(a^*)$ is a contraction mapping when it is viewed as a mapping on $\triangle Z_i$ with fixed point $\pi_i^{\omega}(a^*)$. This means that there is $r \in (0,1)$ such that on the event $F_i^{\omega,1}(t,a)$, we always have

$$|C_{ij}^{\omega}(a^*) f_i(t,a) - \pi_i^{\omega}(a^*)| = |C_{ij}^{\omega}(a^*) f_i(t,a) - C_{ij}^{\omega}(a^*)\pi_i^{\omega}(a^*)| < r\eta.$$

Also, since $C_j^{\omega}(a^*)$ is a stochastic matrix, on the event $\hat{F}_i^{\omega}(t,a)$, we must have

$$|C_{ij}^{\omega}(a^*) f_i(t,a) - C_j^{\omega}(a^*) f_j(t,a)| = |C_j^{\omega}(a^*) C_i^{\omega}(a^*) f_i(t,a) - C_j^{\omega}(a^*) f_j(t,a)| < \eta^2.$$

Taken together, it follows that on the event $F_i^{\omega,1}(t,a) \cap \hat{F}_i^{\omega}(t,a)$,

$$|C_j^{\omega}(a^*) f_j(t,a) - \pi_i^{\omega}(a^*)| < r\eta + \eta^2.$$

Fix a sufficiently small η so that $r\eta + 2\eta^2 < \eta$. Then we obtain

$$|C_j^{\omega}(a^*)f_j(t,a) - \pi_i^{\omega}(a^*)| < \eta - \eta^2$$

on the event $F_i^{\omega,1}(t,a) \cap \hat{F}_i^{\omega}(t,a)$, implying that $F_i^{\omega,1}(t,a) \cap \hat{F}_i^{\omega}(t,a) \subseteq F_j^{\omega,2}(t,a)$. This, together with (33), shows that

$$\{\omega\} \cap G(t,\tau) \cap F_i^{\omega}(t,a) \cap \hat{F}_i^{\omega}(t,a) \subseteq \{\omega\} \cap G(t,\tau) \cap F_j^{\omega}(t,a).$$

Lemma 3 of Cripps, Ely, Mailath, and Samuelson (2008) shows that, for any q, there is t^* and τ such that for any $t > t^*$,

$$\{\omega\} \cap G(t,\tau) \cap F_i^{\omega}(t,a) \subseteq B_i^{t,q}(\{\omega\} \cap G(t,\tau) \cap \hat{F}_i^{\omega}(t,a)).$$

Therefore, we have

$$\begin{aligned}\{\omega\} \cap G(t,\tau) \cap F_i^{\omega}(t,a) &\subseteq B_i^{t,q}(\{\omega\} \cap G(t,\tau) \cap F_i^{\omega}(t,a) \cap \hat{F}_i^{\omega}(t,a)) \\ &\subseteq B_i^{t,q}(\{\omega\} \cap G(t,\tau) \cap F_i^{\omega}(t,a) \cap F_j^{\omega}(t,a)),\end{aligned}$$

as desired. *Q.E.D.*

Now we are ready to prove Proposition 6. Take a sufficiently small η, and fix q. As Monderer and Samet (1989) show, an event $F \subset \Xi$ is common q-belief if it is q-evident. Since Lemma 32 shows that the event $\tilde{F}^{\omega}(t,\tau)$ is q-evident, it is common q-belief whenever it occurs. Lemma 31 shows that this event $\tilde{F}^{\omega}(t,\tau)$ occurs with probability greater than q at state ω, and Lemma 30 shows that the state ω is q-believed on this event. This implies that players commonly learn the true state. *Q.E.D.*

Appendix D: Conditionally Independent Signals

Proposition 1 shows that the folk theorem holds if signals are correlated across players. Here we investigate how the result changes if signals are independently distributed across players. Formally, we impose the following assumption:

Condition 7. (Independent Learning) For each ω, a, and z, $\pi^{\omega}(z|a) = \prod_{i\in I}\pi_i^{\omega}(z_i|a)$.

That is, given any ω and a, signals are independently distributed across players. Under Condition 7, player i's signal has no information about the opponents' signals, and thus we have $C_i^{\omega}(a^*)f_i = \pi_{-i}^{\omega}(a^*)$ for all $f_i \in \triangle Z_i$. This implies that if Condition 7 holds, then Condition 3 is not satisfied.

When signals are independently drawn, player i's signals are not informative about the opponents' signals, and thus player i's best reply after history $h_i^t = (a^{\tau}, z^{\tau})_{\tau=1}^t$ conditional on the true state ω is independent of the past signals $(z^{\tau})_{\tau=1}^t$. Formally, we have the following proposition. Given player i's strategy s_i, let $s_i|_{h_i^t}$ be the continuation strategy after history h_i^t induced by s_i.

Proposition 7. *Suppose that Condition 7 holds. Suppose that players played an ex-post equilibrium s until period t and that the realized history for player i is $h_i^t = (a^{\tau}, z^{\tau})_{\tau=1}^t$. Then for each ω and $\tilde{h}_i^t = (\tilde{a}^{\tau}, \tilde{z}^{\tau})_{\tau=1}^t$ such that $\tilde{a}^{\tau} = a^{\tau}$ for all τ, it is optimal for player i to play $s_i|_{\tilde{h}_i^t}$ in the following periods given any true state ω.*

Proof. Take two different histories h_i^t and $\tilde{h}_i^t$ which shares the same action sequence; i.e., take h_i^t and $\tilde{h}_i^t$ such that $\tilde{a}^{\tau} = a^{\tau}$ for all τ. Since signals are independent, player i's belief about the opponents' history h_{-i}^t conditional on the true state ω and the history h_i^t is identical with the one conditional on the true state ω and the history $\tilde{h}_i^t$. This means that the set of optimal strategies for player i after history h_i^t at ω is the same as the one after history $\tilde{h}_i^t$. Since s is an ex-post equilibrium, $s_i|_{\tilde{h}_i^t}$ is optimal after history $\tilde{h}_i^t$ given ω, and hence the result follows. *Q.E.D.*

The key assumption in this proposition is that s is an ex-post equilibrium. If s is a sequential equilibrium which is not an ex-post equilibrium, then player i's optimal strategy after period t depends on her belief about the true state ω, and such a belief depends on her past signals $(z^{\tau})_{\tau=1}^t$. Hence, her optimal strategy after period t *does* depend on the past signals.

Using this proposition, we will show that when there are only two players, there is an example in which ex-post equilibria cannot approximate some feasible and individually rational payoffs. On the other hand, when there are more than two players, we can prove the folk theorem by ex-post equilibria, as in the case of correlated signals.

D.1 Games with More Than Two Players

When there are more than two players, the folk theorem holds even if signals are independently drawn;

Proposition 8. *Suppose that Conditions 1, 2, and 7 hold. Suppose also that there are at least three players, i.e., $|I| \geq 3$. Then 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.*

An advantage of having more than two players is that a chance that a player can manipulate the continuation play by misreporting in the summary report round is slim, which makes it easier to provide the truth telling incentives in the summary report round. To see this, suppose that there are three players and the true state is ω. By the law of large numbers, each player can make the correct inference ($\omega(i) = \omega$) in the learning round almost surely. If they report truthfully in the summary report round, then everyone reports the same state ω and thus they can agree that the true state is ω. Now suppose that player 1 deviates and reports $\omega(1) = \tilde{\omega}$ in the summary report round. Then the communication outcome is $(\tilde{\omega}, \omega, \omega)$; there are two players reporting ω and one player reporting $\tilde{\omega}$. In such a case, we regard this outcome as a consequence of player 1's deviation and ask players to ignore player 1's report; i.e., in the continuation play, we let players behave as if they could agree that the true state is ω. This implies that player 1 has (almost) no incentive to misreport in the summary report round, since her report cannot influence the continuation play (unless the opponents make a wrong inference in the learning round). Using this property, we can make each player indifferent over all reports in the summary report round so that she is willing to report truthfully.

Note that the above argument does not apply when there are only two players. If player 1 deviates and reports $\omega(1) = \tilde{\omega}$, then the communication outcome is $(\tilde{\omega}, \omega)$ and it is hard to distinguish the identity of the deviator.

Proof. Fix a target payoff $v \in \text{int}V^*$ arbitrarily. The goal is to construct an ex-post equilibrium with payoff v when δ is close to one. As in the proof of Proposition 1, we regard the infinite horizon as a sequence of block games, and each player i's equilibrium strategy is described by an automaton with the state space $\{G,B\}^{|\Omega|}$. A player's strategy within the block game is very similar to the one in the proof of Proposition 1, that is, each player i forms an inference $\omega(i)$ in the learning round, reports the inference $\omega(i)$ in the summary report round, and reveals her private signals in the detailed report round. Only the difference is the behavior in the main round. Specifically, we modify the last bullet point in Section 4.2.1 in the following way:

- If there is j such that all the opponents $l \neq j$ reported the same state ω while player j reported a different state $\tilde{\omega}$ in the summary report round, then we ask players to behave as if everyone reported the same state ω in the summary report round.

That is, in our equilibrium strategy, if players $-j$ could agree in the summary report round that the true state is ω, then players behave as if everyone could agree that the true state is ω, regardless of player j's report $\omega(j)$. Recall that in Section 4.2.1, we have asked players to choose the minimax actions in such histories. Let $s_i^{x_i}$ be the block-game strategy defined above, given the current intention x_i.

What remains is to specify the transition rule of the automaton state x_i^{ω}. As in the proof of Proposition 1, this problem is equivalent to finding appropriate transfer rules. Choose the transfer rule $U_i^{\omega,B}$ as in the proof of Lemma 4. Since all the discussions in the proof of Lemma 4 do not rely on Condition 3, the same result follows; i.e., given any intention profile x with $x_{i-1}^{\omega} = B$, the prescribed strategy $s_i^{x_i}$ is optimal against $s_{-i}^{x_{-i}}$ in the complete-information transfer game with $(\omega, U_i^{\omega,B})$.

As for the transfer function $U_i^{\omega,G}$, we cannot follow the proof of Proposition 1 directly, since Condition 3 plays an important role there. We modify the definition of the regularity in the following way: A block-game history $h_{i-1}^{T_b}$ is *regular with respect to ω and $x_{i-1}^{\omega} = G$* if it satisfies the following properties:

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

(G2) In the summary report round, each player $j \neq i$ reported $\omega(j) = \omega$.

(G3) $x_{i-1}^{\omega} = G$ is reported in the first period of the main round.

(G4) Given the report in the summary report round and in the first period of the main round, everyone followed the prescribed strategy in the second or later periods of the main round.

Note that when the block-game history is regular with respect to ω and $x_{i-1}^{\omega} = G$, player i's average block-game payoff is higher than $\overline{v}_i^{\omega}$ at state ω for sufficiently large T and δ. Let $H_{i-1}^{\omega,G}$ denote the set of all regular histories with respect to ω and $x_{i-1}^{\omega} = G$.

Let $c^G > 0$ be a constant, and we define the transfer rule $U_i^{\omega,G} : H_{i-1}^{T_b} \to \boldsymbol{R}$ in the following way:

- If $h_{i-1}^{T_b} \in H_{i-1}^{\omega,G}$, then let $U_i^{\omega,G}(h_{i-1}^{T_b})$ be such that
$$\frac{1-\delta}{1-\delta^{T_b}}\left[\sum_{t=1}^{T_b}\delta^{t-1}g_i^{\omega}(a^t) + \delta^{T_b}U_i^{\omega,G}(h_{i-1}^{T_b})\right] = \overline{v}_i^{\omega} + c^G.$$

- Otherwise, let $U_i^{\omega,G}(h_{i-1}^{T_b})$ be such that
$$\frac{1-\delta}{1-\delta^{T_b}}\left[\sum_{t=1}^{T_b}\delta^{t-1}g_i^{\omega}(a^t) + \delta^{T_b}U_i^{\omega,G}(h_{i-1}^{T_b})\right] = -2\overline{g}_i^{\omega} + c^G.$$

In words, after regular histories, we set the transfer $U_i^{\omega,G}$ so that the average payoff of the complete-information transfer game is equal to $\overline{v}_i^{\omega} + c^G$. After irregular histories, we set the transfer $U_i^{\omega,G}$ so that the average payoff in the complete-information transfer game is equal to $-2\overline{g}_i^{\omega} + c^G$, which is much lower than $\overline{v}_i^{\omega} + c^G$.

We check whether this transfer function provides appropriate incentives to player i. Note that, by the construction of $U_i^{\omega,G}$, player i's payoff in the complete-information transfer game depends only on whether player $(i-1)$'s block-game history $h_{i-1}^{T_b}$ is regular or not. It is easy to see that player i is willing to follow the prescribed strategy in the learning and main rounds, because (G1) and (G4) imply that the block-game history becomes irregular for sure once player i deviates in these rounds. Also, player i is indifferent among all actions in the summary and detailed report rounds, because actions in these rounds cannot influence whether the resulting history is regular or not. Therefore, for any current intention profile x with $x_{i-1}^{\omega} = G$, the strategy $s_i^{x_i}$ is optimal against $s_{-i}^{x_{-i}}$ in the complete-information transfer game with $(\omega, U_i^{\omega,G})$.

Now we specify the constant term c^G. As in the proof of Proposition 1, we choose c^G in such a way that the expected payoff in the complete-information transfer game when players play the prescribed strategy profile is exactly equal to $\overline{v}_i^\omega$. Then we can prove that there is $\overline{T}$ such that for any $T > \overline{T}$, there is $\overline{\delta} \in (0,1)$ such that for any $\delta \in (\overline{\delta}, 1)$ we have $0 < -(1-\delta)U_i^{\omega,G}(h_{i-1}^{T_b}) < \overline{v}_i^\omega - \underline{v}_i^\omega$. for any block history $h_{i-1}^{T_b}$. The proof is very similar to that of Proposition 4 and hence omitted. *Q.E.D.*

D.2 Two-Player Games

Consider the following two-player games. There are two possible states, ω_1 and ω_2. In each stage game, player 1 chooses either U or D, while player 2 chooses either L or R. Given ω and a chosen action profile a, each player i observes a private signal $z_i \in Z_i = \{z_i(1), z_i(2)\}$. The distribution of player 1's signal z_1 satisfies

$$\pi_1^{\omega_1}(z_1(1)|a) = \pi_1^{\omega_2}(z_1(2)|a) = \frac{2}{3}$$

for all a. That is, the signal $z_1(1)$ is more likely if the true state is ω_1, and the signal $z_1(2)$ is more likely if the true state is ω_2. On the other hand, the distribution of player 2's signal z_2 satisfies

$$\pi_2^{\omega_1}(z_2(1)|a) = \frac{1}{2}, \quad \text{and} \quad \pi_2^{\omega_2}(z_2(2)|a) = 1$$

for all a. That is, the signal $z_2(1)$ reveals that the true state is ω_1. We assume that the signals are independently drawn across players. Assume also that the stage-game payoff for ω_1 is given by the left matrix, and the one for ω_2 is given by the right matrix:

	L	R
U	1, 0	0, 0
D	0, 1	0, 1

	L	R
U	0, 1	0, 1
D	0, 0	2, 0

In this example, the payoff vector $(1,0)$ is feasible and individually rational at ω_1, and the payoff vector $(2,0)$ is feasible and individually rational at ω_2. Hence, the payoff vector $((1,0),(2,0))$ is in the set V^*. However, this payoff vector cannot be approximated by ex-post equilibria even when δ is close to one:

Proposition 9. *Let $\varepsilon < \frac{2}{3}$. Then for any $\delta \in (0,1)$, any feasible and individually rational payoff in the ε-neighborhood of $((1,0),(2,0))$ is not achieved by an ex-post equilibrium.*

The formal proof is given later, but the idea is as follows. To simplify the discussion, let ε be close to zero. Suppose that the above proposition is not true, so that there is an ex-post equilibrium s approximating $((1,0),(2,0))$. Let s_2^* be player 2's strategy such that she deviates from the equilibrium strategy s_2 by pretending as if she observed the signal $z_2(2)$ in all periods, regardless of her true observations. Suppose that the true state is ω_1, and that player 1 follows the equilibrium strategy s_1 while player 2 deviates to s_2^*. Since such a deviation should not be profitable, player 2's average payoff should be less than her equilibrium payoff, which is close to 0. This implies that player 1 must play U with probability close to one in almost all periods. (Otherwise, with a non-negligible probability player 1 takes D, which yields a payoff of 1 to player 2.) Then there must be a sequence of player 1's signals, $(\hat{z}_1^\tau)_{\tau=1}^\infty$, such that player 1 chooses U with probability close to one in almost all periods at state ω_1 if the realized observation is $(\hat{z}_1^\tau)_{\tau=1}^\infty$ and player 2 plays s_2^*. Let s_1^* be player 1's strategy such that she deviates from the equilibrium strategy s_1 by pretending as if her observation was $(\hat{z}_1^\tau)_{\tau=1}^\infty$.

Suppose that the true state is ω_2, and that player 2 follows the equilibrium strategy s_2 while player 1 deviates to s_1^*. Since the true state is ω_2, player 2's play is exactly the same as s_2^*; then by the definition of s_1^*, player 1 must play U, which gives a payoff of 0 to player 1, with a high probability in almost all periods. Thus player 1's average payoff must be close to 0; this is is less than her equilibrium payoff, which approximates 2. Hence deviating to s_1^* is suboptimal when the true state is ω_2. However, this is a contradiction, because Proposition 7 ensures that s_1^* is a best reply to s_2 when the true state is ω_2. Thus we cannot approximate $((1,0),(2,0))$ by ex-post equilibria.

Of course, ex-post incentive compatibility is much stronger than sequential rationality, and thus it may be possible to sustain $((1,0),(2,0))$ using sequential equilibria. However, sequential equilibria are not robust to a perturbation of the initial prior; i.e., a sequential equilibrium may not constitute an equilibrium once the initial prior is perturbed. This can be a problem, because researchers may not know the initial prior to which the model is applied; indeed, the initial prior is private information, which is difficult to observe, and also it may depend on the age of relationship, which may not be known to researchers. On the other hand, ex-post equilibria are robust to a specification of the initial prior, so researchers who do not know such detailed information can regard them as equilibrium strategies.

Proof. Fix $\varepsilon < \frac{2}{3}$. Fix an arbitrary payoff vector v in the ε-neighborhood of $((1,0),(2,0))$, and fix an arbitrary discount factor $\delta \in (0,1)$. Suppose that there is an ex-post equilibrium s with payoff v.

Let s_2^* be player 2's strategy such that she deviates from s_2 by pretending as if she observed $z_2(2)$ in all periods. That is, let s_2^* be such that $s_2^*(h_2^t) = s_2(\tilde{h}_2^t)$ for all t, $h_2^t = (a^\tau, z_2^\tau)_{\tau=1}^t$, and $h_2^t = (\tilde{a}^\tau, \tilde{z}_2^\tau)_{\tau=1}^t$ such that $a^\tau = \tilde{a}^\tau$ and $\tilde{z}_2^\tau = z_2(2)$ for all τ. Suppose that the true state is ω_1 and that player 2 deviates to s_2^*. Then player 2's average payoff is

$$(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}E_{h_1^{t-1}}[s_1(h_1^{t-1})[D]|s_1,s_2^*,\omega_1].$$

Since deviating to s_2^* should not be profitable and the equilibrium payoff v is in the ε-neighborhood of $((1,0),(2,0))$, we have

$$(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}E_{h_1^{t-1}}[s_1(h_1^{t-1})[D]|s_1,s_2^*,\omega_1] \leq \varepsilon.$$

That is, the probability of player 1 choosing D is really small in expectation, if the true state is ω_1 and players play the profile (s_1, s_2^*). Then there must be a sequence $(\hat{z}_1^\tau)_{\tau=1}^\infty$ of player 1's signals such that player 1 chooses D with a very small probability if the realized signal sequence is $(\hat{z}_1^\tau)_{\tau=1}^\infty$; that is, there is $(\hat{z}_1^\tau)_{\tau=1}^\infty$ such that

$$(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}E_{(a^1,\cdots,a^{t-1})}[s_1(a^1,\cdots,a^{t-1},\hat{z}_1^1,\cdots,\hat{z}_1^{t-1})[D]|s_1,s_2^*,\omega_1] \leq \varepsilon. \quad (34)$$

Let s_1^* be player 1's strategy such that she deviates from s_1 by pretending as if her signal sequence is $(\hat{z}_1^\tau)_{\tau=1}^\infty$; that is, let s_1^* be such that $s_1^*(h_1^t) = s_1(\tilde{h}_1^t)$ for all t, $h_1^t = (a^\tau, z_1^\tau)_{\tau=1}^t$, and $h_1^t = (\tilde{a}^\tau, \tilde{z}_1^\tau)_{\tau=1}^t$ such that $a^\tau = \tilde{a}^\tau$ and $\tilde{z}_1^\tau = \hat{z}_1^\tau$ for all τ.

Suppose that the true state is ω_2, and that player 2 follows the equilibrium strategy s_2 while player 1 deviates to s_1^*. Since player 2 always observes $z_2(2)$ at ω_2 and (34) hold, we must have

$$\begin{aligned}&(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}E_{h_1^{t-1}}[s_1(h_1^{t-1})[D]|s_1^*,s_2,\omega_2]\\ &=(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}E_{(a^1,\cdots,a^{t-1})}[s_1(a^1,\cdots,a^{t-1},\hat{z}_1^1,\cdots,\hat{z}_1^{t-1})[D]|s_1,s_2^*,\omega_1] \leq \varepsilon;\end{aligned}$$

that is, player 1 must choose D with a very small probability in this case. Then, because player 1's payoff by taking U is 0, her average payoff is at most

$$(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}E_{h_1^{t-1}}[2s_1(h_1^{t-1})[D]|s_1^*,s_2,\omega_2] \leq 2\varepsilon.$$

This means that deviating to s_1^* at ω_2 is suboptimal, since the equilibrium payoff is at least $2-\varepsilon$ and $\varepsilon < \frac{2}{3}$. However, this contradicts with Proposition 7. *Q.E.D.*