Bounding Equilibrium Payoffs in Repeated Games with Private Monitoring: Online Appendix

Takuo Sugaya and Alexander Wolitzky
Stanford Graduate School of Business and MIT

April 25, 2016

Proof of Proposition 2

We prove that $\overline{E_{\text{talk}}(\delta, p)} = \overline{E_{\text{med}}(\delta)}$. In our construction, players ignore private signals $y_{i,t}$ observed in periods $t = 1, 2, \ldots$. That is, only signal $y_{i,0}$ observed in period 0 is used. Hence we can see p as an ex ante correlation device. Since we consider two-player games, whenever we say players i and j, we assume that they are different players: $i \neq j$.

The structure of the proof is as follows: take any strategy of the mediator, $\tilde{\mu}$, that satisfies inequality (3) in the text (perfect monitoring incentive compatibility); and let $\tilde{v}$ be the value when the players follow $\tilde{\mu}$. Since each $\hat{v} \in E_{\text{med}}(\delta)$ has a corresponding $\hat{\mu}$ that satisfies perfect monitoring incentive compatibility, it suffices to show that, for each $\varepsilon > 0$, there exists a sequential equilibrium whose equilibrium payoff v satisfies $\|v - \tilde{v}\| < \varepsilon$ in the following environment:

1. At the beginning of the game, each player i receives a message m_i^{mediator} from the mediator.

2. In each period t, the stage game proceeds as follows:

   (a) Given player i's history $(m_i^{\text{mediator}}, (m_\tau^{\text{1st}}, a_\tau, m_\tau^{\text{2nd}})_{\tau=1}^{t-1})$, each player i sends the first message $m_{i,t}^{\text{1st}}$ simultaneously.

   (b) Given player i's history $(m_i^{\text{mediator}}, (m_\tau^{\text{1st}}, a_\tau, m_\tau^{\text{2nd}})_{\tau=1}^{t-1}, m_t^{\text{1st}})$, each player i takes action $a_{i,t}$ simultaneously.

   (c) Given player i's history $(m_i^{\text{mediator}}, (m_\tau^{\text{1st}}, a_\tau, m_\tau^{\text{2nd}})_{\tau=1}^{t-1}, m_t^{\text{1st}}, a_t)$, each player i sends the second message $m_{i,t}^{\text{2nd}}$ simultaneously.

We call this environment "perfect monitoring with cheap talk."

To this end, from $\tilde{\mu}$, we first create a strict full-support equilibrium μ with mediated perfect monitoring that yields payoffs close to $\tilde{v}$. We then move from μ to a similar equilibrium μ^*, which will be easier to transform into an equilibrium with perfect monitoring with cheap talk. Finally, from μ^*, we create an equilibrium with perfect monitoring with cheap talk with the same on-path action distribution.

Construction and Properties of μ

In this subsection, we consider mediated perfect monitoring throughout. Since $\mathring{W}^* \neq \emptyset$, by Lemma 2 in the text, there exists a strict full support equilibrium μ^{strict} with mediated perfect monitoring. As in the proof of that lemma, consider the following strategy of the mediator: In period 1, the mediator draws one of two states, $R_{\tilde{v}}$ and R_{perturb}, with probabilities $1-\eta$ and η, respectively. In state $R_{\tilde{v}}$, the mediator's recommendation is determined as follows: If no player has deviated up to period t, the mediator recommends r_t according to $\tilde{\mu}(h_m^t)$. If only player i has deviated, the mediator recommends $r_{-i,t}$ to player j according to α_j^*, and recommends some best response to α_j^* to player i. Multiple deviations are treated as in the proof of Lemma 1 in the text. On the other hand, in state R_{perturb}, the mediator follows the equilibrium μ^{strict}. Let μ denote this strategy of the mediator. From now on, we fix $\eta > 0$ arbitrarily.

With mediated perfect monitoring, since μ^{strict} has full support, player i believes that the mediator's state is R_{perturb} with positive probability after any history. Therefore, by perfect monitoring incentive compatibility and the fact that μ^{strict} is a strict equilibrium, it is always strictly optimal for each player i to follow her recommendation. This means that, for each period t, there exist $\varepsilon_t > 0$ and $T_t < \infty$ such that, for each player i and on-path history h_m^{t+1}, we have

$$\begin{aligned} & (1-\delta)\mathbb{E}^{\mu}\left[u_i(r_t) \mid h_m^t, r_{i,t}\right] + \delta\mathbb{E}^{\mu}\left[(1-\delta)\sum_{\tau=t+1}^{\infty}\delta^{\tau-t-1}u_i(\mu(h_m^\tau)) \mid h_m^t, r_{i,t}\right] \\ > \; & \max_{a_i \in A_i}(1-\delta)\mathbb{E}\left[u_i(a_i, r_{-i,t}) \mid h_m^t, r_{i,t}\right] \\ & +(\delta-\delta^{T_t})\left\{(1-\varepsilon_t)\max_{\hat{a}_i} u_i(\hat{a}_i, \alpha_j^{\varepsilon_t}) + \varepsilon_t \max_{a\in A} u_i(a)\right\} + \delta^{T_t}\max_{a\in A} u_i(a). \end{aligned} \tag{1}$$

That is, suppose that if player i unilaterally deviates from on-path history, then player j virtually minmaxes player i for $T_t - 1$ periods with probability $1-\varepsilon_t$. (Recall that α_j^* is the minmax strategy and α_j^{ε} is a full support perturbation of α_j^*.) Then player i has a strict incentive not to deviate from any recommendation in period t on equilibrium path. Equivalently, since μ is an full support recommendation, player i has a strict incentive not to deviate unless she herself has deviated.

Moreover, for sufficiently small $\varepsilon_t > 0$, we have

$$\begin{aligned} & (1-\delta)\mathbb{E}^{\mu}\left[u_i(r_t) \mid h_m^t, r_{i,t}\right] + \delta\mathbb{E}^{\mu}\left[(1-\delta)\sum_{\tau=t+1}^{\infty}\delta^{\tau-t-1}u_i(\mu(h_m^\tau)) \mid h_m^t\right] \\ > \; & (1-\delta^{T_t})\left\{(1-\varepsilon_t)\max_{\hat{a}_i} u_i(\hat{a}_i, \alpha_j^{\varepsilon_t}) + \varepsilon_t \max_{a\in A} u_i(a)\right\} + \delta^{T_t}\max_{a\in A} u_i(a). \end{aligned} \tag{2}$$

That is, if a deviation is punished with probability $1-\varepsilon_t$ for T_t periods including the current period, then player i believes that the deviation is strictly unprofitable.[1]

For each t, we fix $\varepsilon_t > 0$ and $T_t < \infty$ with (1) and (2). Without loss, we can take ε_t decreasing: $\varepsilon_t \geq \varepsilon_{t+1}$ for each t.

[1] If the current on-path recommendation schedule $\Pr^{\mu}(r_{j,t} \mid h_m^t, r_{i,t})$ is very close to α_j^*, then (2) may be more restrictive than (1).

Construction and Properties of μ^*

In this subsection, we again consider mediated perfect monitoring. We further modify μ and create the following mediator's strategy μ^*: At the beginning of the game, for each i, t, and a^t, the mediator draws $r_{i,t}^{\text{punish}}(a^t)$ according to $\alpha_i^{\varepsilon_t}$. In addition, for each i and t, she draws $\omega_{i,t} \in \{R, P\}$ such that $\omega_{i,t} = R$ (regular) and P (punish) with probability $1 - p_t$ and p_t, respectively, independently across i and t. We will pin down $p_t > 0$ in Lemma 1. Moreover, given $\omega_t = (\omega_{1,t}, \omega_{2,t})$, the mediator chooses $r_t(a^t)$ for each a^t as follows: If $\omega_{1,t} = \omega_{2,t} = R$, then she draws $r_t(a^t)$ according to $\mu(a^t)(r)$. If $\omega_{i,t} = R$ and $\omega_{j,t} = P$, then she draws $r_{i,t}(a^t)$ from $\Pr^{\mu}(r_i \mid r_{j,t}^{\text{punish}}(a^t))$ while she draws $r_{j,t}(a^t)$ randomly from $\sum_{a_j \in A_j} \frac{a_j}{|A_j|}$.[2] Finally, if $\omega_{1,t} = \omega_{2,t} = P$, then she draws $r_{i,t}(a^t)$ randomly from $\sum_{a_i \in A_i} \frac{a_i}{|A_i|}$ for each i independently. Since μ has full support, μ^* is well defined.

As will be seen, we will take p_t sufficiently small. In addition, recall that $\eta > 0$ (the perturbation of $\tilde{\mu}$ to μ) is arbitrarily. In the next subsection and onward, we construct an equilibrium with perfect monitoring with cheap talk that has the same equilibrium action distribution as μ^*. Since p_t is small and $\eta > 0$ is arbitrary, constructing such an equilibrium suffices to prove Proposition 2.

At the start of the game, the mediator draws ω_t, $r_{i,t}^{\text{punish}}(a^t)$, and $r_t(a^t)$ for each i, t, and a^t. Given them, the mediator sends messages to the players as follows:

1. At the start of the game, the mediator sends $\left(\left(r_{i,t}^{\text{punish}}(a^t)\right)_{a^t \in A^{t-1}}\right)_{t=1}^{\infty}$ to player i.

2. In each period t, the stage game proceeds as follows:

   (a) The mediator decides $\bar{\omega}_t(a^t) \in \{R, P\}^2$ as follows: if there is no unilateral deviator (defined below), then the mediator sets $\bar{\omega}_t(a^t) = \omega_t$. If instead player i is a unilateral deviator, then the mediator sets $\bar{\omega}_{i,t}(a^t) = R$ and $\bar{\omega}_{j,t}(a^t) = P$.

   (b) Given $\bar{\omega}_{i,t}(a^t)$, the mediator sends $\bar{\omega}_{i,t}(a^t)$ to player i. In addition, if $\bar{\omega}_{i,t}(a^t) = R$, then the mediator sends $r_{i,t}(a^t)$ to player i as well.

   (c) Given these messages, player i takes an action. In equilibrium, if player i has not yet deviated, then player i takes $r_{i,t}(a^t)$ if $\bar{\omega}_{i,t}(a^t) = R$ and takes $r_{i,t}^{\text{punish}}(a^t)$ if $\bar{\omega}_{i,t}(a^t) = P$. For notational convenience, let

$$r_{i,t} = \begin{cases} r_i(a^t) \text{ if } \bar{\omega}_{i,t}(a^t) = R, \\ r_{i,t}^{\text{punish}}(a^t) \text{ if } \bar{\omega}_{i,t}(a^t) = P \end{cases}$$

   be the action that player i is supposed to take if she has not yet deviated. Her strategy after her own deviation is not specified.

We say that player i has unilaterally deviated if there exist $\tau \leq t - 1$ and a unique i such that (i) for each $\tau' < \tau$, we have $a_{n,\tau'} = r_{n,\tau'}$ for each $n \in \{1, 2\}$ (no deviation happened

[2] As will be seen below, if $\omega_{j,t} = P$, then player j is supposed to take $r_{j,t}^{\text{punish}}(a^t)$. Hence, $r_{j,t}(a^t)$ does not affect the equilibrium action. We define $r_{j,t}(a^t)$ so that, when the mediator sends a message only at the beginning of the game (in the game with perfect monitoring with cheap talk), she sends a "dummy recommendation" $r_{j,t}(a^t)$ so that player j does not realize that $\omega_{j,t} = P$ until period t.

until period $\tau - 1$) and (ii) $a_{i,\tau} \neq r_{i,\tau}$ and $a_{j,\tau} = r_{j,\tau}$ (player i deviates in period τ and player j does not deviate).

Note that μ^* is close to μ on the equilibrium path for sufficiently small p_t. Hence, on-path strict incentive compatibility for player i follows from (1). Moreover, the incentive compatibility condition analogous to (2) also holds.

Lemma 1 *There exists $\{p_t\}_{t=1}^{\infty}$ with $p_t > 0$ for each t such that it is strictly optimal for each player i to follow her recommendation: For each player i and history*

$$h_i^t \equiv \left(\left(\left(r_{i,t}^{\text{punish}}\left(a^t\right) \right)_{a^t \in A^{t-1}} \right)_{t=1}^{\infty}, a^t, \left(\bar{\omega}_\tau(a^\tau)\right)_{\tau=1}^{t-1}, \bar{\omega}_{i,t}(a^t), \left(r_{i,\tau}\right)_{\tau=1}^{t} \right),$$

if player i herself has not yet deviated, we have the following two inequalities:

1. *If a deviation is punished by $\alpha_j^{\varepsilon_t}$ for the next period T_t periods with probability $1 - \varepsilon_t - \sum_{\tau=t}^{t+T_t-1} p_\tau$, then it is strictly unprofitable:*

$$\begin{aligned} & (1-\delta)\mathbb{E}^{\mu^*}\left[u_i(r_{i,t}, a_{j,t}) \mid h_i^t\right] + \delta\mathbb{E}^{\mu^*}\left[(1-\delta)\sum_{\tau=t+1}^{\infty} \delta^{\tau-t-1} u_i(r_{i,\tau}, a_{j,\tau}) \mid h_i^t, a_{i,t} = r_{i,t}\right] \\ > \; & \max_{a_i \in A_i} (1-\delta)\mathbb{E}^{\mu^*}\left[u_i(a_i, a_{j,t}) \mid h_i^t\right] \\ & + (\delta - \delta^{T_t})\left\{ \left(1 - \varepsilon_t - \textstyle\sum_{\tau=t}^{t+T_t-1} p_\tau\right) \max_{\hat{a}_i} u_i(\hat{a}_i, \alpha_j^{\varepsilon_t}) + \left(\varepsilon_t + \textstyle\sum_{\tau=t}^{t+T_t-1} p_\tau\right) \max_{a \in A} u_i(a) \right\} \\ & + \delta^{T_t} \max_{a \in A} u_i(a). \end{aligned} \tag{3}$$

2. *If a deviation is punished by $\alpha_j^{\varepsilon_t}$ from the current period with probability $1 - \varepsilon_t - \sum_{\tau=t}^{t+T_b-1} p_t$, then it is strictly unprofitable:*

$$\begin{aligned} & (1-\delta)\mathbb{E}^{\mu^*}\left[u_i(r_{i,t}, a_{j,t}) \mid h_i^t\right] + \delta\mathbb{E}^{\mu^*}\left[(1-\delta)\sum_{\tau=t+1}^{\infty} \delta^{\tau-t-1} u_i(r_{i,\tau}, a_{j,\tau}) \mid h_i^t, a_{i,t} = r_{i,t}\right] \\ > \; & (1 - \delta^{T_t})\left\{ \left(1 - \varepsilon_t - \textstyle\sum_{\tau=t}^{t+T_t-1} p_\tau\right) \max_{\hat{a}_i} u_i(\hat{a}_i, \alpha_j^{\varepsilon_t}) + \left(\varepsilon_t + \textstyle\sum_{\tau=t}^{t+T_t-1} p_\tau\right) \max_{a \in A} u_i(a) \right\} \\ & + \delta^{T_t} \max_{a \in A} u_i(a). \end{aligned} \tag{4}$$

Moreover, $\mathbb{E}^{\mu^*}$ does not depend on the specification of player j's strategy after player j's own deviation, for each history h_i^t such that player i has not deviated.

Proof. Since μ^* has full support on the equilibrium path, a player i who has not yet deviated always believes that player j has not deviated. Hence, $\mathbb{E}^{\mu^*}$ is well defined without specifying player j's strategy after player j's own deviation.

Moreover, since p_t is small and $\omega_{j,t}$ is independent of $(\omega_\tau)_{\tau=1}^{t-1}$ and $\omega_{i,t}$, given $(\bar{\omega}_\tau(a^\tau))_{\tau=1}^{t-1}$ and $\bar{\omega}_{i,t}(a^t)$ (which are equal to $(\omega_\tau)_{\tau=1}^{t-1}$ and $\omega_{i,t}$ on-path), player i believes that $\bar{\omega}_{j,t}(a^t)$ is equal to $\omega_{j,t}$ and $\omega_{j,t}$ is equal to R with a high probability, unless player i has deviated. Since

$$\Pr^{\mu*}(r_{j,t} \mid \bar{\omega}_{i,t}(a^t), \{\bar{\omega}_{j,t}(a^t) = R\}, h_i^t) = \Pr^{\mu*}(r_{j,t} \mid a^t, r_{i,t}),$$

we have that the difference

$$\mathbb{E}^{\mu^*}\left[u_i(r_{i,t},a_{j,t}) \mid h_i^t\right] - \mathbb{E}^{\mu}\left[u_i(r_{i,t},a_{j,t}) \mid r_i^t, a^t, r_{i,t}\right]$$

is small for small p_t.

Further, if p_τ is small for each $\tau \geq t+1$, then since ω_τ is independent of ω_t with $t \leq \tau - 1$, regardless of $(\bar{\omega}_\tau(a^\tau))_{\tau=1}^t$, player i believes that $\bar{\omega}_{i,\tau}(a^\tau) = \bar{\omega}_{j,\tau}(a^\tau) = R$ with high probability for $\tau \geq t+1$ on the equilibrium path. Since the distribution of the recommendation given μ^* is the same as that of μ given a^τ and $\bar{\omega}_{i,\tau}(a^\tau) = \bar{\omega}_{j,\tau}(a^\tau) = R$, we have that

$$\mathbb{E}^{\mu^*}\left[(1-\delta)\sum_{\tau=t+1}^{\infty}\delta^{\tau-t-1}u_i(r_{i,\tau},a_{j,\tau}) \mid h_i^t, a_{i,t} = r_{i,t}\right] - \mathbb{E}^{\mu}\left[(1-\delta)\sum_{\tau=t+1}^{\infty}\delta^{\tau-t-1}u_i(r_{i,\tau},a_{j,\tau}) \mid r_i^t, a^t, r_{i,t}\right]$$

is small for small p_τ with $\tau \geq t+1$.

Hence, (1) and (2) imply that, there exists $\bar{p}_t > 0$ such that, if $p_\tau \leq \bar{p}_t$ for each $\tau \geq t$, then the claims of the lemma hold. Hence, if we take $p_t \leq \min_{\tau \leq t} \bar{p}_\tau$, then the claims hold. ■

We fix $\{p_t\}_{t=1}^\infty$ so that Lemma 1 holds. This fully pins down μ^* with mediated perfect monitoring.

Construction with Perfect Monitoring with Cheap Talk

Given μ^* with mediated perfect monitoring, we define the equilibrium strategy with perfect monitoring with cheap talk such that the equilibrium action distribution is the same as μ^*. We must pin down the following four objects: at the beginning of the game, what message m_i^{mediator} player i receives from the mediator; what message $m_{i,t}^{\text{1st}}$ player i sends at the beginning of period t; what action $a_{i,t}$ player i takes in period t; and what message $m_{i,t}^{\text{2nd}}$ player i sends at the end of period t.

Intuitive Argument

As in μ^*, at the beginning of the game, for each i, t, and a^t, the mediator draws $r_{i,t}^{\text{punish}}(a^t)$ according to $\alpha_i^{\varepsilon_t}$. In addition, with $p_t > 0$ pinned down in Lemma 1, she draws $\omega_t \in \{R, P\}^2$ and $r_t(a^t)$ as in μ^* for each t and a^t. She then defines $\bar{\omega}_t(a^t)$ from a^t, $r_t(a^t)$, and ω_t as in μ^*.

Intuitively, the mediator sends all the information about

$$\left(\left(\bar{\omega}_t(a^t), r_t\left(a^t\right), r_{1,t}^{\text{punish}}\left(a^t\right), r_{2,t}^{\text{punish}}\left(a^t\right)\right)_{a^t \in A^{t-1}}\right)_{t=1}^{\infty}$$

through the initial messages $(m_1^{\text{mediator}}, m_2^{\text{mediator}})$. In particular, the mediator directly sends $\left((r_{i,t}^{\text{punish}}(a^t))_{a^t \in A^{t-1}}\right)_{t=1}^{\infty}$ to player i as a part of m_i^{mediator}. Hence, we focus on how we replicate the role of the mediator in μ^* of sending $(\bar{\omega}_t(a^t), r_t(a^t))$ in each period, depending on realized history a^t.

The key features to establish are (i) player i does not know the instructions for the other player, (ii) before player i reaches period t, player i does not know her own recommendations for periods $\tau \geq t$ (otherwise, player i would obtain more information than the original equilibrium μ^* and thus might want to deviate), and (iii) no player wants to deviate (in particular, if player i deviates in actions or cheap talk, then the strategy of player j is as if the state were $\bar{\omega}_{j,t} = P$ in μ^*, for a sufficiently long time with a sufficiently high probability).

The properties (i) and (ii) are achieved by the same mechanism as in Theorem 9 of Heller, Solan and Tomala (2012, henceforth HST). In particular, without loss, let $A_i = \{1_i, ..., n_i\}$ be player i's action set. We can view $r_{i,t}(a^t)$ as an element of $\{1, ..., n_i\}$. The mediator at the beginning of the game draws $r_t(a^t)$ for each a^t.

Instead of sending $r_{i,t}(a^t)$ directly to player i, the mediator encodes $r_{i,t}(a^t)$ as follows: For a sufficiently large $N^t \in \mathbb{Z}$ to be determined, we define $p^t = N^t n_i n_j$. This p^t corresponds to p_h in HST. Let $\mathbb{Z}_{p^t} \equiv \{1, ..., p^t\}$. The mediator draws $x^j_{i,t}(a^t)$ uniformly and independently from $\mathbb{Z}_{p^t}$ for each i, t, and a^t. Given them, she defines

$$y^i_{i,t}(a^t) \equiv x^j_{i,t}(a^t) + r_{i,t}(a^t) \pmod{n_i}. \tag{5}$$

Intuitively, $y^i_{i,t}(a^t)$ is the "encoded instruction" of $r_{i,t}(a^t)$, and to obtain $r_{i,t}(a^t)$ from $y^i_{i,t}(a^t)$, player i needs to know $x^j_{i,t}(a^t)$. The mediator gives $\left((y^i_i(a^t))_{a^t \in A^{t-1}}\right)_{t=1}^{\infty}$ to player i as a part of m_i^{mediator}. At the same time, she gives $\left(\left(x^j_{i,t}(a^t)\right)_{a^t \in A^{t-1}}\right)_{t=1}^{\infty}$ to player j as a part of m_j^{mediator}. At the beginning of period t, player j sends $x^j_{i,t}(a^t)$ by cheap talk as a part of $m^{\text{1st}}_{j,t}$, based on the realized action a^t, so that player i does not know $r_{i,t}(a^t)$ until period t. (Throughout the proof, the superscript of a variable represents who is informed about the variable, and the subscript represents whose recommendation the variable is about.)

In order to incentivize player j to tell the truth, the equilibrium should embed a mechanism that punishes player i if she tells a lie. In HST, this is done as follows: The mediator draws $\alpha^i_{i,t}(a^t)$ and $\beta^i_{i,t}(a^t)$ uniformly and independently from $\mathbb{Z}_{p^t}$, and defines

$$u^j_{i,t}(a^t) \equiv \alpha^i_{i,t}(a^t) \times x^j_{i,t}(a^t) + \beta^i_{i,t}(a^t) \pmod{p^t}. \tag{6}$$

The mediator gives $x^j_{i,t}(a^t)$ and $u^j_{i,t}(a^t)$ to player j while she gives $\alpha^i_{i,t}(a^t)$ and $\beta^i_{i,t}(a^t)$ to player i. In period t, player j is supposed to send $x^j_{i,t}(a^t)$ and $u^j_{i,t}(a^t)$ to player i. If player i receives $x^j_{i,t}(a^t)$ and $u^j_{i,t}(a^t)$ with

$$u^j_{i,t}(a^t) \neq \alpha^i_{i,t}(a^t) \times x^j_{i,t}(a^t) + \beta^i_{i,t}(a^t) \pmod{p^t}, \tag{7}$$

then player i interprets that player j has deviated. For sufficiently large N^t, since player j does not know $\alpha^i_{i,t}(a^t)$ and $\beta^i_{i,t}(a^t)$, if player j tells a lie about $x^j_{i,t}(a^t)$, then with a high probability, player j creates a situation where (7) holds.

Since HST considers Nash equilibrium, they let player i minimax player j forever after (7) holds. On the other hand, since we consider sequential equilibrium, as in the proof of Lemma 2 in the text, we will create a coordination mechanism such that, if player j tells a lie, then with high probability player i minimaxes player j for a long time and player i assigns probability zero to the event that player i punishes player j.

To this end, we consider the following coordination: First, if and only if $\bar{\omega}_{i,t}(a^t) = R$, the mediator defines $u^j_{i,t}(a^t)$ as (6). Otherwise, $u^j_{i,t}(a^t)$ is randomly drawn. That is,

$$u^j_{i,t}(a^t) \equiv \begin{cases} \alpha^i_{i,t}(a^t) \times x^j_{i,t}(a^t) + \beta^i_{i,t}(a^t) \pmod{p^t} & \text{if } \bar{\omega}_{i,t}(a^t) = R, \\ \text{uniformly distributed over } \mathbb{Z}_{p^t} & \text{if } \bar{\omega}_{i,t}(a^t) = P. \end{cases} \tag{8}$$

Since both $\bar{\omega}_{i,t}(a^t) = R$ and $\bar{\omega}_{i,t}(a^t) = P$ happen with a positive probability, player i after receiving $u^j_{i,t}(a^t)$ with $u^j_{i,t}(a^t) \neq \alpha^i_{i,t}(a^t) \times x^j_{i,t}(a^t) + \beta^i_{i,t}(a^t) \pmod{p^t}$ interprets that $\bar{\omega}_{i,t}(a^t) = P$. For notational convenience, let $\hat{\omega}_{i,t}(a^t) \in \{R, P\}$ be player i's interpretation of $\bar{\omega}_{i,t}(a^t)$. After $\hat{\omega}_{i,t}(a^t) = P$, she takes period-$t$ action according to $r^{\text{punish}}_{i,t}(a^t)$. Given this inference, if player j tells a lie about $u^j_{i,t}(a^t)$ with $\bar{\omega}_{i,t}(a^t) = R$, then with a high probability, she induces a situation with $u^j_{i,t}(a^t) \neq \alpha^i_{i,t}(a^t) \times x^j_{i,t}(a^t) + \beta^i_{i,t}(a^t) \pmod{p^t}$, and player i punishes player j in period t (without noticing player j's deviation).

Second, switching to $r^{\text{punish}}_{i,t}(a^t)$ for period t only may not suffice, if player j believes that player i's action distribution given $\bar{\omega}_{i,t}(a^t) = R$ is close to the minimax strategy. Hence, we ensure that, once player j deviates, player i takes $r^{\text{punish}}_{i,\tau}(a^\tau)$ for a sufficiently long time.

To this end, we change the mechanism so that player j does not always know $u^j_{i,t}(a^t)$. Instead, the mediator draws p^t independent random variables $v^j_{i,t}(n, a^t)$ with $n = 1, ..., p^t$ uniformly from $\mathbb{Z}_{p^t}$. In addition, she draws $n^i_{i,t}(a^t)$ uniformly from $\mathbb{Z}_{p^t}$. The mediator defines $u^j_{i,t}(n, a^t)$ for each $n = 1, ..., p^t$ as follows:

$$u^j_{i,t}(n, a^t) = \begin{cases} u^j_{i,t}(a^t) & \text{if } n = n^i_{i,t}(a^t), \\ v^j_{i,t}(n, a^t) & \text{if otherwise,} \end{cases}$$

that is, $u^j_{i,t}(n, a^t)$ corresponds to $u^j_{i,t}(a^t)$ with (8) only if $n = n^i_{i,t}(a^t)$. For other n, $u^j_{i,t}(n, a^t)$ is completely random.

The mediator sends $n^i_{i,t}(a^t)$ to player i, and sends $\{u^j_{i,t}(n, a^t)\}_{n \in \mathbb{Z}_{p^t}}$ to player j. In addition, the mediator sends $n^j_{i,t}(a^t)$ to player j, where

$$n^j_{i,t}(a^t) = \begin{cases} n^i_{i,t}(a^t) & \text{if } \omega_{i,t-1}(a^{t-1}) = P, \\ \text{uniformly distributed over } \mathbb{Z}_{p^t} & \text{if } \omega_{i,t-1}(a^{t-1}) = R \end{cases}$$

is equal to $n^i_{i,t}(a^t)$ if and only if last-period $\bar{\omega}_{i,t-1}(a^{t-1})$ is equal to P.

In period t, player j is asked to send $x^j_{i,t}(a^t)$ and $u^j_{i,t}(n, a^t)$ with $n = n^i_{i,t}(a^t)$, that is, send $x^j_{i,t}(a^t)$ and $u^j_{i,t}(a^t)$. If and only if player j's messages $\hat{x}^j_{i,t}(a^t)$ and $\hat{u}^j_{i,t}(a^t)$ satisfy

$$\hat{u}^j_{i,t}(a^t) = \alpha^i_{i,t}(a^t) \times \hat{x}^j_{i,t}(a^t) + \beta^i_{i,t}(a^t) \pmod{p^t},$$

player i interprets $\hat{\omega}_{i,t}(a^t) = R$. If player i has $\hat{\omega}_{i,t}(a^t) = R$, then player i knows that player j needs to know $n^i_{i,t+1}(a^{t+1})$ to send the correct $u^j_{i,t+1}(n, a^{t+1})$ in the next period. Hence, she sends $n^i_{i,t+1}(a^{t+1})$ to player j. If player i has $\hat{\omega}_{i,t}(a^t) = P$, then she believes that player j knows $n^i_{i,t+1}(a^{t+1})$ and does not send $n^i_{i,t+1}(a^{t+1})$.

Given this coordination, once player j creates a situation with $\bar{\omega}_{i,t}(a^t) = R$ but $\hat{\omega}_{i,t}(a^t) = P$, then player j cannot receive $n^i_{i,t+1}(a^{t+1})$. Without knowing $n^i_{i,t+1}(a^{t+1})$, with a large N^{t+1}, with a high probability, player j cannot know which $u^j_{i,t+1}(n, a^{t+1})$ she should send. Then, again, she will create a situation with

$$\hat{u}^j_{i,t+1}(a^{t+1}) \neq \alpha^i_{i,t+1}(a^{t+1}) \times \hat{x}^j_{i,t}(a^{t+1}) + \beta^i_{i,t}(a^{t+1}) \pmod{p^{t+1}},$$

that is, $\hat{\omega}_{i,t+1}(a^{t+1}) = P$. Recursively, player i has $\hat{\omega}_{i,\tau}(a^\tau) = P$ for a long time with a high probability if player j tells a lie.

Finally, if player j takes a deviant action in period t, then the mediator has drawn $\bar{\omega}_{i,\tau}(a^\tau) = P$ for each $\tau \geq t+1$ for a^τ corresponding to the realized history. With $\bar{\omega}_{i,\tau}(a^\tau) = P$, in order to avoid $\hat{\omega}_{i,\tau}(a^\tau) = P$, player j needs to create a situation

$$\hat{u}^j_{i,\tau}(a^\tau) = \alpha^i_{i,\tau}(a^\tau) \times \hat{x}^j_{i,\tau}(a^\tau) + \beta^i_{i,\tau}(a^\tau) \pmod{p^\tau}$$

without knowing $\alpha^i_{i,\tau}(a^\tau)$ and $\beta^i_{i,\tau}(a^\tau)$ while the mediator's message does not tell her what is $\alpha^i_{i,t}(a^t) \times x^j_{i,t}(a^t) + \beta^i_{i,t}(a^t) \pmod{p^\tau}$ by (8). Hence, for sufficiently large N^τ, player j cannot avoid $\hat{\omega}_{i,\tau}(a^\tau) = P$ with a nonnegligible probability. Hence, player j will be minmaxed from the next period with a high probability.

The above argument in total shows that, if player j deviates, whether in communication or action, then she will be minmaxed for sufficiently long time. Lemma 1 ensures that player j does not want to tell a lie or take a deviant action.

Formal Construction

Let us formalize the above construction: As in μ^*, at the beginning of the game, for each i, t, and a^t, the mediator draws $r^{\text{punish}}_{i,t}(a^t)$ according to $\alpha^{\varepsilon_t}_i$; then she draws $\omega_t \in \{R, P\}^2$ and $r_t(a^t)$ for each t and a^t; and then she defines $\bar{\omega}_t(a^t)$ from a^t, $r_t(a^t)$, and ω_t as in μ^*. For each t and a^t, she draws $x^j_{i,t}(a^t)$ uniformly and independently from $\mathbb{Z}_{p^t}$. Given them, she defines

$$y^i_{i,t}(a^t) \equiv x^j_{i,t}(a^t) + r_{i,t}(a^t) \pmod{n_i},$$

so that (5) holds.

The mediator draws $\alpha^i_{i,t}(a^t)$, $\beta^i_{i,t}(a^t)$, $\tilde{u}^j_{i,t}(a^t)$, $v^j_{i,t}(n, a^t)$ for each $n \in \mathbb{Z}_{p^t}$, $n^i_{i,t}(a^t)$, and $\tilde{n}^j_{i,t}(a^t)$ from the uniform distribution over $\mathbb{Z}_{p^t}$ independently for each player i, each period t, and each a^t.

As in (8), the mediator defines

$$u^j_{i,t}(a^t) \equiv \begin{cases} \alpha^i_{i,t}(a^t) \times x^j_{i,t}(a^t) + \beta^i_{i,t}(a^t) \pmod{p^t} & \text{if } \bar{\omega}_{i,t}(a^t) = R, \\ \tilde{u}^j_{i,t}(a^t) & \text{if } \bar{\omega}_{i,t}(a^t) = P. \end{cases}$$

In addition, the mediator defines

$$u^j_{i,t}(n, a^t) = \begin{cases} u^j_{i,t}(a^t) & \text{if } n = n^i_{i,t}(a^t), \\ v^j_{i,t}(n, a^t) & \text{if otherwise} \end{cases}$$

and

$$n^j_{i,t}(a^t) = \begin{cases} n^i_{i,t}(a^t) & \text{if } t = 1 \text{ or } \omega_{i,t-1}(a^{t-1}) = P, \\ \tilde{n}^j_{i,t}(a^t) & \text{if } t \neq 1 \text{ and } \omega_{i,t-1}(a^{t-1}) = R, \end{cases}$$

as explained above.

Let us now define the equilibrium:

1. At the beginning of the game, the mediator sends

$$m_i^{\text{mediator}} = \left(\left(\begin{array}{c} y_{i,t}^i(a^t), \alpha_{i,t}^i(a^t), \beta_{i,t}^i(a^t), r_{i,t}^{\text{punish}}(a^t), \\ n_{i,t}^i(a^t), n_{j,t}^i(a^t), \left(u_{j,t}^i(n,a^t)\right)_{n \in \mathbb{Z}_{p^t}}, x_{j,t}^i(a^t) \end{array} \right)_{a^t \in A^{t-1}} \right)_{t=1}^{\infty}$$

   to each player i.

2. In each period t, the stage game proceeds as follows: In equilibrium,

$$m_{j,t}^{\text{1st}} = \begin{cases} u_{i,t}^j(m_{i,t-1}^{\text{2nd}}, a^t), x_{i,t}^j(a^t) & \text{if } t \neq 1 \text{ and } m_{i,t-1}^{\text{2nd}} \neq \{\text{babble}\}, \\ u_{i,t}^j(n_{i,t}^j(a^t), a^t), x_{i,t}^j(a^t) & \text{if } t = 1 \text{ or } m_{i,t-1}^{\text{2nd}} = \{\text{babble}\} \end{cases} \tag{9}$$

   and

$$m_{j,t}^{\text{2nd}} = \begin{cases} n_{j,t+1}^j(a^{t+1}) & \text{if } \hat{\omega}_{j,t}(a^t) = R, \\ \{\text{babble}\} & \text{if } \hat{\omega}_{j,t}(a^t) = P. \end{cases}$$

   Note that, since $m_{j,t}^{\text{2nd}}$ is sent at the end of period t, the players know $a^{t+1} = (a_1, ..., a_t)$.

   (a) Given player i's history $(m_i^{\text{mediator}}, (m_\tau^{\text{1st}}, a_\tau, m_\tau^{\text{2nd}})_{\tau=1}^{t-1})$, each player i sends the first message $m_{i,t}^{\text{1st}}$ simultaneously. If player i herself has not yet deviated, then

$$m_{i,t}^{\text{1st}} = \begin{cases} u_{j,t}^i(m_{j,t-1}^{\text{2nd}}, a^t), x_{j,t}^i(a^t) & \text{if } t \neq 1 \text{ and } m_{j,t-1}^{\text{2nd}} \neq \{\text{babble}\}, \\ u_{j,t}^i(n_{j,t}^i(a^t), a^t), x_{j,t}^i(a^t) & \text{if } t = 1 \text{ or } m_{j,t-1}^{\text{2nd}} = \{\text{babble}\}. \end{cases}$$

   Let $m_{i,t}^{\text{1st}}(u)$ be the first element of $m_{i,t}^{\text{1st}}$ (that is, either $u_{j,t}^i(m_{j,t-1}^{\text{2nd}}, a^t)$ or $u_{j,t}^i(n_{j,t}^i(a^t), a^t)$ on equilibrium); and let $m_{i,t}^{\text{1st}}(x)$ be the second element ($x_{j,t}^i(a^t)$ on equilibrium). As a result, the profile of the messages m_t^{1st} becomes common knowledge.
   If

$$m_{j,t}^{\text{1st}}(u) \neq \alpha_{i,t}^i(a^t) \times m_{j,t}^{\text{1st}}(x) + \beta_{i,t}^i(a^t) \pmod{p^t}, \tag{10}$$

   then player i interprets $\hat{\omega}_{i,t}(a^t) = P$. Otherwise, $\hat{\omega}_{i,t}(a^t) = R$.

   (b) Given player i's history $(m_i^{\text{mediator}}, (m_\tau^{\text{1st}}, a_\tau, m_\tau^{\text{2nd}})_{\tau=1}^{t-1}, m_t^{\text{1st}})$, each player i takes action $a_{i,t}$ simultaneously. If player i herself has not yet deviated, then player i takes $a_{i,t} = r_{i,t}$ with

$$r_{i,t} = \begin{cases} y_{i,t}^i(a^t) - m_{j,t}^{\text{1st}}(x) \pmod{n_i} & \text{if } \hat{\omega}_{i,t}(a^t) = R, \\ r_{i,t}^{\text{punish}}(a^t) & \text{if } \hat{\omega}_{i,t}(a^t) = P. \end{cases} \tag{11}$$

   Recall that $y_{i,t}^i(a^t) \equiv x_{i,t}^j(a^t) + r_{i,t}(a^t) \pmod{n_i}$ by (5). By (9), therefore, player i takes $r_{i,t}^i(a^t)$ if $\bar{\omega}_{i,t}(a^t) = R$ and $r_{i,t}^{\text{punish}}(a^t)$ if $\bar{\omega}_{i,t}(a^t) = P$ on the equilibrium path, as in μ^*.

   (c) Given player i's history $(m_i^{\text{mediator}}, (m_\tau^{\text{1st}}, a_\tau, m_\tau^{\text{2nd}})_{\tau=1}^{t-1}, m_t^{\text{1st}}, a_t)$, each player i sends the second message $m_{i,t}^{\text{2nd}}$ simultaneously. If player i herself has not yet deviated, then

$$m_{i,t}^{\text{2nd}} = \begin{cases} n_{i,t+1}^i(a^{t+1}) & \text{if } \hat{\omega}_{i,t}(a^t) = R, \\ \{\text{babble}\} & \text{if } \hat{\omega}_{i,t}(a^t) = P. \end{cases}$$

As a result, the profile of the messages m_t^{2nd} becomes common knowledge. Note that $\bar{\omega}_t(a^t)$ becomes common knowledge as well on equilibrium path.

Incentive Compatibility

The above equilibrium has full support: Since $\bar{\omega}_t(a^t)$, and $r_t(a^t)$ have full support, $(m_1^{\text{mediator}}, m_2^{\text{mediator}})$ have full support as well. Hence, we are left to verify player i's incentive not to deviate from the equilibrium strategy, given that player i believes that player j has not yet deviated after any history of player i.

Suppose that player i followed the equilibrium strategy until the end of period $t-1$. First, consider player i's incentive to tell the truth about $m_{i,t}^{\text{1st}}$. In equilibrium, player i sends

$$m_{i,t}^{\text{1st}} = \begin{cases} u_{j,t}^i(m_{j,t-1}^{\text{2nd}}, a^t), x_{j,t}^i(a^t) & \text{if } m_{j,t-1}^{\text{2nd}} \neq \{\text{babble}\}, \\ u_{j,t}^i(n_{j,t}^i(a^t), a^t), x_{j,t}^i(a^t) & \text{if } m_{j,t-1}^{\text{2nd}} = \{\text{babble}\}. \end{cases}$$

The random variables possessed by player i are independent of those possessed by player j given $(m_\tau^{\text{1st}}, a_\tau, m_\tau^{\text{2nd}})_{\tau=1}^{t-1}$, except that (i) $u_{i,t}^j(a^t) = \alpha_{i,t}^i(a^t) \times x_{i,t}^j(a^t) + \beta_{i,t}^i(a^t) \pmod{p^t}$ if $\bar{\omega}_{i,t}(a^t) = R$, (ii) $u_{j,t}^i(a^t) = \alpha_{j,t}^j(a^t) \times x_{j,t}^i(a^t) + \beta_{j,t}^j(a^t) \pmod{p^t}$ if $\bar{\omega}_{j,t}(a^t) = R$, (iii) $n_{i,\tau}^j(a^\tau) = n_{i,\tau}^i(a^\tau)$ if $\omega_{i,\tau-1}(a^{\tau-1}) = P$ while $n_{i,\tau}^j(a^\tau) = \tilde{n}_{i,\tau}^i(a^\tau)$ if $\omega_{i,\tau-1}(a^{\tau-1}) = R$, and (iv) $n_{j,\tau}^i(a^\tau) = n_{j,\tau}^j(a^\tau)$ if $\omega_{j,\tau-1}(a^{\tau-1}) = P$ while $n_{j,\tau}^i(a^\tau) = \tilde{n}_{j,\tau}^j(a^\tau)$ if $\omega_{j,\tau-1}(a^{\tau-1}) = R$. Since $\alpha_{i,t}^i(a^t)$, $\beta_{i,t}^i(a^t)$, $\tilde{u}_{i,t}^j(a^t)$, $v_{i,t}^j(n, a^t)$ $n_{i,t}^i(a^t)$, and $\tilde{n}_{i,t}^j(a^t)$ are uniform and independent, player i cannot learn $\bar{\omega}_{i,\tau}(a^\tau)$, $r_{i,\tau}(a^\tau)$, or $r_{j,\tau}(a^\tau)$ with $\tau \geq t$. Hence, player i believes at the time when she sends $m_{i,t}^{\text{1st}}$ that her equilibrium value is equal to

$$(1-\delta)\mathbb{E}^{\mu^*}\left[u_i(a_t) \mid h_i^t\right] + \delta \mathbb{E}^{\mu^*}\left[(1-\delta)\sum_{\tau=t+1}^{\infty} \delta^{\tau-t-1} u_i(a_t) \mid h_i^t\right],$$

where h_i^t is as if player i observed $\left(r_{i,t}^{\text{punish}}(a^t)\right)_{a^t \in A^{t-1} t=1}^{\infty}$, a^t, $(\bar{\omega}_\tau(a^\tau))_{\tau=1}^{t-1}$, and $r_{i,t}(a^t)$, and believed that $r_\tau(a^\tau) = a_\tau$ for each $\tau = 1, ..., t-1$ with μ^* with mediated perfect monitoring.

On the other hand, for each $e > 0$, for a sufficiently large N^t, if player i tells a lie in at least one element $m_{i,t}^{\text{1st}}$, then with probability $1-e$, player i creates a situation

$$m_{i,t}^{\text{1st}}(u) \neq \alpha_{j,t}^j(a^t) \times m_{i,t}^{\text{1st}}(x) + \beta_{j,t}^j(a^t) \pmod{p^t}.$$

Hence, (10) (with indices i and j reversed) implies that $\hat{\omega}_{j,t}(a^t) = P$.

Moreover, if player i creates a situation with $\hat{\omega}_{j,t}(a^t) = P$, then player j will send $m_{j,t}^{\text{2nd}} = \{\text{babble}\}$ instead of $n_{j,t+1}^j(a^{t+1})$. Unless $\bar{\omega}_{j,t}(a^t) = P$, since $n_{j,t+1}^j(a^{t+1})$ is independent of player i's variables, player i believes that $n_{j,t+1}^j(a^{t+1})$ is distributed uniformly over $\mathbb{Z}_{p^{t+1}}$. Hence, for each $e > 0$, for sufficiently large N^t, if $\hat{\omega}_{j,t}(a^t) = R$, then player i will send $m_{i,t+1}^{\text{1st}}$ with

$$m_{i,t+1}^{\text{1st}}(u) \neq \alpha_{j,t+1}^j(a^{t+1}) \times m_{i,t+1}^{\text{1st}}(x) + \beta_{j,t+1}^j(a^{t+1}) \pmod{p^{t+1}}$$

with probability $1-e$. Then, by (10) (with indices i and j reversed), player j will have $\hat{\omega}_{j,t+1}(a^{t+1}) = P$.

Recursively, if $\bar{\omega}_{j,\tau}(a^\tau) = R$ for each $\tau = t, .., t + T_t - 1$, then player i will induce

$\hat{\omega}_{j,\tau}(a^\tau) = P$ for each $\tau = t, .., t + T_t - 1$ with a high probability. Hence, for $\varepsilon_t > 0$ and T_t fixed in (1) and (2), for sufficiently large $\bar{N}^t$, if $N^\tau \geq \bar{N}^t$ for each $\tau \geq t$, then player i will be punished for the subsequent T_t periods regardless of player i's continuation strategy with probability no less than $1 - \varepsilon_t - \sum_{\tau=t}^{t+T_t-1} p_\tau$. ($\sum_{\tau=t}^{t+T_t-1} p_\tau$ represents the maximum probability of having $\bar{\omega}_{i,\tau}(a^\tau) = P$ for some τ for subsequent T_t periods.) (4) implies that telling a lie gives strictly lower payoff than the equilibrium payoff. Therefore, it is optimal to tell the truth about $m_{i,t}^{\text{1st}}$. (In (4), we have shown interim incentive compatibility after knowing $\bar{\omega}_{i,t}(a^t)$ and $r_{i,t}$ while here, we consider h_i^t before $\bar{\omega}_{i,t}(a^t)$ and $r_{i,t}$. Taking the expectation with respect to $\bar{\omega}_{i,t}(a^t)$ and $r_{i,t}$, (4) ensures incentive compatibility before knowing $\bar{\omega}_{i,t}(a^t)$ and $r_{i,t}$.)

Second, consider player i's incentive to take the action $a_{i,t} = r_{i,t}$ according to (11) if player i follows the equilibrium strategy until she sends $m_{i,t}^{\text{1st}}$. If she follows the equilibrium strategy, then player i believes at the time when she takes an action that her equilibrium value is equal to

$$(1-\delta)\mathbb{E}^{\mu^*}\left[u_i(a_t) \mid h_i^t\right] + \delta\mathbb{E}^{\mu^*}\left[(1-\delta)\sum_{\tau=t+1}^{\infty}\delta^{\tau-t-1}u_i(a_t) \mid h_i^t\right],$$

where h_i^t is as if player i observed $\left(r_{i,t}^{\text{punish}}(a^t)\right)_{a^t \in A^{t-1} t=1}^{\infty}$, a^t, $(\bar{\omega}_\tau(a^\tau))_{\tau=1}^{t-1}$, $\bar{\omega}_{i,t}(a^t)$, and $r_{i,t}$, and believed that $r_\tau(a^\tau) = a_\tau$ for each $\tau = 1, ..., t-1$ with μ^* with mediated perfect monitoring. (Compared to the time when player i sends $m_{i,t}^{\text{1st}}$, player i now knows $\bar{\omega}_{i,t}(a^t)$ and $r_{i,t}$ on equilibrium path.)

If player i deviates from $a_{i,t}$, then $\bar{\omega}_{j,\tau}(a^\tau) = P$ by definition for each $\tau \geq t+1$ and a^τ that is compatible with a^t (that is, $a^\tau = (a^t, a_t, ..., a_{\tau-1})$ for some $a_t, ..., a_{\tau-1}$). To avoid being minmaxed in period τ, player i needs to induce $\hat{\omega}_{j,\tau}(a^\tau) = R$ although $\bar{\omega}_{j,\tau}(a^\tau) = P$. Given $\bar{\omega}_{j,\tau}(a^\tau) = P$, since $\alpha_{i,t}^i(a^t)$, $\beta_{i,t}^i(a^t)$, $\tilde{u}_{i,t}^j(a^t)$, $v_{i,t}^j(n, a^t)$ $n_{i,t}^i(a^t)$, and $\tilde{n}_{i,t}^j(a^t)$ are uniform and independent (conditional on the other variables), regardless of player i's continuation strategy, by (10) (with indices i and j reversed), player i will send $m_{i,\tau}^{\text{1st}}$ with

$$m_{i,\tau}^{\text{1st}}(u) \neq \alpha_{j,\tau}^j(a^\tau) \times m_{i,\tau}^{\text{1st}}(x) + \beta_{j,\tau}^j(a^\tau) \pmod{p^\tau}$$

with a high probability.

Hence, for sufficiently large $\bar{N}^t$, if $N^\tau \geq \bar{N}^t$ for each $\tau \geq t$, then player i will be punished for the next T_t periods regardless of player i's continuation strategy with probability no less than $1 - \varepsilon_t$. By (3), deviating from $r_{i,t}$ gives a strictly lower payoff than her equilibrium payoff. Therefore, it is optimal to take $a_{i,t} = r_{i,t}$.

Finally, consider player i's incentive to tell the truth about $m_{i,t}^{\text{2nd}}$. Regardless of $m_{i,t}^{\text{2nd}}$, player j's actions do not change. Hence, we are left to show that telling a lie does not improve player i's deviation gain by giving player i more information.

On the equilibrium path, player i knows $\bar{\omega}_{i,t}(a^t)$. If player i tells the truth, then $m_{i,t}^{\text{2nd}} = n_{i,t+1}^i(a^{t+1}) \neq \{\text{babble}\}$ if and only if $\bar{\omega}_{i,t}(a^t) = R$. Moreover, player j sends

$$m_{j,t+1}^{\text{1st}} = \begin{cases} u_{i,t+1}^j(m_{i,t}^{\text{2nd}}, a^{t+1}), x_{i,t+1}^j(a^{t+1}) & \text{if } \bar{\omega}_{i,t}(a^t) = R, \\ u_{i,t+1}^j(n_{i,t+1}^j(a^{t+1}), a^{t+1}), x_{i,t+1}^j(a^{t+1}) & \text{if } \bar{\omega}_{i,t}(a^t) = P. \end{cases}$$

Since $n^j_{i,t+1}(a^{t+1}) = n^i_{i,t+1}(a^{t+1})$ if $\bar{\omega}_{i,t}(a^t) = P$, in total, if player i tells the truth, then player i knows $u^i_{j,t+1}(m^i_{i,t+1}(a^{t+1}), a^{t+1})$ and $x^i_{j,t+1}(a^{t+1})$. This is sufficient information to infer $\bar{\omega}_{i,t+1}(a^{t+1})$ and $r_{i,t+1}(a^{t+1})$ correctly.

If she tells a lie, then player j's messages are changed to

$$m^{\text{1st}}_{j,t+1} = \begin{cases} u^j_{i,t+1}(m^{\text{2nd}}_{i,t}, a^{t+1}), x^j_{i,t+1}(a^{t+1}) & \text{if } m^{\text{2nd}}_{i,t} \neq \{\text{babble}\}, \\ u^j_{i,t+1}(n^j_{i,t+1}(a^{t+1}), a^{t+1}), x^j_{i,t+1}(a^{t+1}) & \text{if } m^{\text{2nd}}_{i,t} = \{\text{babble}\}. \end{cases}$$

Since $\alpha^i_{i,t+1}(a^{t+1})$, $\beta^i_{i,t+1}(a^{t+1})$, $\tilde{u}^j_{i,t+1}(a^{t+1})$, $v^j_{i,t+1}(n, a^{t+1})$ $n^i_{i,t+1}(a^{t+1})$, and $\tilde{n}^j_{i,t+1}(a^{t+1})$ are uniform and independent conditional on $\bar{\omega}_{i,t+1}(a^{t+1})$ and $r_{i,t+1}(a^{t+1})$, $u^j_{i,t+1}(n, a^{t+1})$ and $x^j_{i,t+1}(a^{t+1})$ are not informative about player j's recommendation from period $t+1$ on or player i's recommendation from period $t+2$ on, given that player i knows $\bar{\omega}_{i,t+1}(a^{t+1})$ and $r_{i,t+1}(a^{t+1})$. Since telling the truth informs player i of $\bar{\omega}_{i,t+1}(a^{t+1})$ and $r_{i,t+1}(a^{t+1})$, there is no gain from telling a lie.