Supplementary Material

Supplement to “Bounding equilibrium payoffs in repeated games with private monitoring”  

TAKUO SUGAYA  
Graduate School of Business, Stanford University

ALEXANDER WOLITZKY  
Department of Economics, MIT

Proof of Proposition 2

We prove that \( \tilde{E}_{\text{talk}}(\delta, p) = \tilde{E}_{\text{med}}(\delta) \). 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 \( \tilde{v} \in \tilde{E}_{\text{med}}(\delta) \) has a corresponding \( \tilde{\mu} \) 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:

(i) At the beginning of the game, each player \( i \) receives a message \( m_{i,1}^{\text{mediator}} \) from the mediator.

(ii) In each period \( t \), the stage game proceeds as follows:

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

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

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

We call this environment perfect monitoring with cheap talk.

Takuo Sugaya: tsugaya@stanford.edu  
Alexander Wolitzky: wolitzky@mit.edu

Copyright © 2017 The Authors. Theoretical Economics. The Econometric Society. Licensed under the Creative Commons Attribution-NonCommercial License 3.0. Available at http://econtheory.org.  
DOI: 10.3982/TE2270
To this end, from $\check{\mu}$, we first create a strict full-support equilibrium $\mu$ with mediated perfect monitoring that yields payoffs close to $\check{\nu}$. We then move from $\mu$ to a similar equilibrium $\mu^*$, which will be easier to transform into an equilibrium with perfect monitoring with cheap talk. Finally, from $\mu^*$, we create an equilibrium with perfect monitoring with cheap talk with the same on-path action distribution.

Construction and properties of $\mu$

In this subsection, we consider mediated perfect monitoring throughout. Since $W^* \neq \emptyset$, by Lemma 2 in the main text, there exists a strict full-support equilibrium $\mu^{\text{strict}}$ 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, $R_\emptyset$ and $R_{\text{perturb}}$, with probabilities $1 - \eta$ and $\eta$, respectively. In state $R_\emptyset$, the mediator’s recommendation is determined as follows: If no player has deviated up to period $t$, the mediator recommends $r_t$ according to $\hat{\mu}(h^*_m)$; if only player $i$ has deviated, the mediator recommends $r^*_i$ to player $j$ according to $\alpha^*_j$, and recommends some best response to $\alpha^*_j$ to player $i$. Multiple deviations are treated as in the proof of Lemma 1. In contrast, in state $R_{\text{perturb}}$, the mediator follows the equilibrium $\mu^{\text{strict}}$. Let $\mu$ denote this strategy of the mediator. From now on, we fix $\eta > 0$ arbitrarily.

With mediated perfect monitoring, since $\mu^{\text{strict}}$ has full support, player $i$ believes that the mediator’s state is $R_{\text{perturb}}$ with positive probability after any history. Therefore, by perfect monitoring incentive compatibility and the fact that $\mu^{\text{strict}}$ 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 $h^t_{m+1}$, we have

$$
(1 - \delta)\mathbb{E}[u_i(r_t) \mid h^t_m, r_t] + \delta \mathbb{E}\left[(1 - \delta)\sum_{t=T_1}^{\infty} \delta^{t-T_1} u_i(h^*_m) \mid h^t_m, r_t, h^t_{m+1}\right] > \max_{a_i \in A_i} (1 - \delta)\mathbb{E}[u_i(a_i, r^*_t) \mid h^t_m, r_t, h^t_{m+1}] + (\delta - \delta^{T_t})\{1 - \varepsilon_t\} \max_{a_i \in A} u_i(\hat{a}_i, \alpha^*_j) + \varepsilon_t \max_{a_i \in A} u_i(a) + \delta^{T_t} \max_{a_i \in A} u_i(a).
$$

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 $\alpha^*_j$ is the minmax strategy and $\alpha^{\varepsilon_t}_j$ is a full-support perturbation of $\alpha^*_j$.) Then player $i$ has a strict incentive not to deviate from any recommendation in period $t$ on equilibrium path. Equivalently, since $\mu$ is a 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

$$
(1 - \delta)\mathbb{E}[u_i(r_t) \mid h^t_m, r_t] + \delta \mathbb{E}\left[(1 - \delta)\sum_{t=T_1}^{\infty} \delta^{t-T_1} u_i(h^*_m) \mid h^t_m\right] > (1 - \delta^{T_t})\{1 - \varepsilon_t\} \max_{a_i \in A} u_i(\hat{a}_i, \alpha^{\varepsilon_t}_j) + \varepsilon_t \max_{a_i \in A} u_i(a) + \delta^{T_t} \max_{a_i \in A} u_i(a).
$$

(S2)
That is, if a deviation is punished with probability $1 - \epsilon_t$ for $T_t$ periods including the current period, then player $i$ believes that the deviation is strictly unprofitable.\footnote{If the current on-path recommendation schedule $\Pr^\mu(r_{j,t} | h^\mu_{m}, r_{i,t})$ is very close to $a^*_j$, then (S1) may be more restrictive than (S1).}

For each $t$, we fix $\epsilon_t > 0$ and $T_t < \infty$ with (S1) and (S2). Without loss, we can take $\epsilon_t$ decreasing: $\epsilon_t \geq \epsilon_{t+1}$ for each $t$.

### Construction and properties of $\mu^*$

In this subsection, we again consider mediated perfect monitoring. We further modify $\mu$ and create the following mediator’s strategy $\mu^*$: Fix a fully mixed $\hat{\mu} \in \Delta(A)$ with $u(\hat{\mu}) \in \hat{W}^*$. 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 $\alpha^t_i$. 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 S1. Moreover, given $\omega_t = (\omega_{1,t}, \omega_{2,t})$, the mediator chooses $r_i(a^t)$ for each $a^t$ as follows: If $\omega_{1,t} = \omega_{2,t} = R$, then she draws $r_i(a^t)$ according to $\mu(a^t)(r)$ if $\omega_{1,t} = \omega_{2,t} = R$ for each $\tau \leq t - 1$; and draws $r_i(a^t)$ according to $\hat{\mu}(r)$ if there exists $\tau \leq t - 1$ with $\omega_{1,\tau} = P$ or $\omega_{2,\tau} = P$. If $\omega_{i,t} = R$ and $\omega_{j,t} = P$, then she draws $r_{i,t}(a^t)$ from $\Pr^\mu(r_{j,t} | r_{j,t}^{\text{punish}}(a^t))$ while she draws $r_{j,t}(a^t)$ randomly from $\sum_{a_j \in A_j} \frac{a_j}{|A_j|}$. 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 $\mu$ has full support, $\mu^*$ is well defined.

As will be seen, we will take $p_t$ sufficiently small. In addition, recall that $\eta > 0$ (the perturbation of $\hat{\mu}$ to $\mu$) 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 $\mu^*$. 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 $\omega_t$, $r_{i,t}^{\text{punish}}(a^t)$, and $r_i(a^t)$ for each $i$, $t$, and $a^t$. Given them, the mediator sends messages to the players as follows:

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

(ii) In each period $t$, the stage game proceeds as follows:

(a) The mediator decides $\bar{\omega}_{i,t}(a^t) \in \{R, P\}$ as follows: if there is no unilateral deviator (defined below), then the mediator sets $\bar{\omega}_i(a^t) = \omega_i$. 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 $\bar{\omega}_{i,t}(a^t) = P$, otherwise player $i$ deviates.

\footnote{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 $t$.}
if \( \tilde{\omega}_{i,t}(a') = P \). For notational convenience, let

\[
    r_{i,t} = \begin{cases} 
        r_i(a') & \text{if } \tilde{\omega}_{i,t}(a') = R, \\
        r_{i,\text{punish}}(a') & \text{if } \tilde{\omega}_{i,t}(a') = P.
    \end{cases}
\]

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 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 \( \tau \) and player \( j \) does not deviate).

Note that \( \mu^* \) is close to \( \mu \) on the equilibrium path for sufficiently small \( p_t \). Hence, on-path strict incentive compatibility for player \( i \) follows from (S1). Moreover, the incentive compatibility condition analogous to (S2) also holds.

**Lemma S1.** 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 \( h_{i,t} \equiv \left( (\tilde{\omega}_{\tau}(a_{\tau}))_{\tau=1}^t, (\tilde{\omega}_{\tau}(a_{\tau}))_{\tau=1}^t, (r_{i,\tau})_{\tau=1}^t \right) \), if player \( i \) herself has not yet deviated, we have the following two inequalities:

(i) If a deviation is punished by \( \alpha^{e_i}_j \) for the next period \( T_i \) periods with probability

\[
    1 - \varepsilon_t - \sum_{\tau=t}^{t+T_i-1} p_{\tau},
\]

then it is strictly unprofitable:

\[
    (1 - \delta) \mathbb{E}^{\mu^*}[u_i(r_{i,t}, a_{j,t}) | h_{i,t}^r] + \delta \mathbb{E}^{\mu^*} \left[ (1 - \delta) \sum_{\tau=t+1}^{\infty} \delta^{\tau-t-1} u_i(r_{i,\tau}, a_{j,\tau}) | h_{i,t}^r, a_{i,t} = r_{i,t} \right] - \max_{a_i \in A} (1 - \delta) \mathbb{E}^{\mu^*}[u_i(a_i, a_{j,t}) | h_{i,t}^r] + (\delta - \delta^{T_i}) \left( 1 - \varepsilon_t - \sum_{\tau=t}^{t+T_i-1} p_{\tau} \right) \max_{a_i} u_i(\tilde{a}_i, \alpha^{e_i}_j)
\]

(ii) If a deviation is punished by \( \alpha^{e_i}_j \) from the current period with probability \( 1 - \varepsilon_t - \sum_{\tau=t}^{t+T_i-1} p_{\tau} \), then it is strictly unprofitable:

\[
    (1 - \delta) \mathbb{E}^{\mu^*}[u_i(r_{i,t}, a_{j,t}) | h_{i,t}^r] + \delta \mathbb{E}^{\mu^*} \left[ (1 - \delta) \sum_{\tau=t+1}^{\infty} \delta^{\tau-t-1} u_i(r_{i,\tau}, a_{j,\tau}) | h_{i,t}^r, a_{i,t} = r_{i,t} \right] - \max_{a \in A} (1 - \delta) \mathbb{E}^{\mu^*}[u_i(a, a_{j,t}) | h_{i,t}^r] + \delta^{T_i} \max_{a \in A} u_i(a). \]
we have that the difference

\[ (1 - \delta^T) \left\{ \left(1 - \epsilon_t - \sum_{\tau = t}^{t+T_t-1} p_{\tau} \right) \max_{\hat{a}_t} u_i(\hat{a}_t, \alpha^{i}_{\tau}) + \left(\epsilon_t + \sum_{\tau = t}^{t+T_t-1} p_{\tau} \right) \max_{a \in A} u_i(a) \right\} + \delta^T \max_{a \in A} u_i(a). \]

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

**Proof.** Given \( \hat{\mu} \), since \( u(\hat{\mu}) \in \hat{W}^* \), for sufficiently small \( \epsilon_t > 0 \), we have

\[
(1 - \delta) E^{\hat{\mu}} \left[ u_i(r_t) \mid h^t_m, r_{i,t} \right] + \delta u_i(\hat{\mu}) > \max_{a_i \in A_i} (1 - \delta) E^{\hat{\mu}} \left[ u_i(a_i, r_{-i,t}) \mid h^t_m, r_{i,t} \right] + (\delta - \delta^T) \left\{ (1 - \epsilon_t) \max_{\hat{a}_t} u_i(\hat{a}_t, \alpha^{i}_{\tau}) + \epsilon_t \max_{a \in A} u_i(a) \right\} + \delta^T \max_{a \in A} u_i(a)
\]

and

\[
(1 - \delta) E^{\hat{\mu}} \left[ u_i(r_t) \mid h^t_m, r_{i,t} \right] + \delta u_i(\hat{\mu}) > (1 - \delta^T) \left\{ (1 - \epsilon_t) \max_{\hat{a}_t} u_i(\hat{a}_t, \alpha^{i}_{\tau}) + \epsilon_t \max_{a \in A} u_i(a) \right\} + \delta^T \max_{a \in A} u_i(a).
\]

Hence (S1) and (S2) hold with \( \mu \) replaced with \( \hat{\mu} \).

Since \( \mu^* \) has full support on the equilibrium path, a player \( i \) who has not yet deviated always believes that player \( j \) has not deviated. Hence, \( E^{\mu^*} \) 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 \( (\tilde{\omega}_{j,t}(a^t))_{\tau = 1}^{t-1} \) and \( \tilde{\omega}_{i,t}(a') \) (which are equal to \( (\omega_{\tau})_{\tau = 1}^{t-1} \) and \( \omega_{i,t} \) on path), player \( i \) believes that \( \tilde{\omega}_{j,t}(a') \) is equal to \( \omega_{j,t} \) and \( \omega_{j,t} \) is equal to \( R \) with a high probability, unless player \( i \) has deviated. Since

\[
Pr^\mu^*(r_{j,t} \mid \tilde{\omega}_{i,t}(a'), \tilde{\omega}_{j,t}(a') = R, h^t_j) = Pr^\mu^*(r_{j,t} \mid a', r_{i,t}),
\]

we have that the difference

\[
E^{\mu^*}[u_i(r_{i,t}, a_{j,t}) \mid h^t_i] - E^\mu[u_i(r_{i,t}, a_{j,t}) \mid r_{i,t}, a', r_{i,t}]
\]

is small for small \( p_t \).

Further, if \( p_{\tau} \) is small for each \( \tau \geq t + 1 \), then since \( \omega_{\tau} \) is independent of \( \omega_{t} \) with \( t \leq \tau - 1 \), regardless of \( (\tilde{\omega}_{j,t}(a^t))_{\tau = 1}^{t-1} \), player \( i \) believes that \( \tilde{\omega}_{i,t}(a^t) = \tilde{\omega}_{j,t}(a^t) = R \) with high probability for \( \tau \geq t + 1 \) on the equilibrium path. Since the distribution of the recommendation given \( \mu^* \) is the same as that of \( \mu \) (or \( \hat{\mu} \) given \( a' \) and \( (\tilde{\omega}_{i,t}(a^t), \tilde{\omega}_{j,t}(a^t)) \) = \( (R, R) \) for each \( \tau \leq t - 1 \) (or \( (\tilde{\omega}_{i,t}(a^t), \tilde{\omega}_{j,t}(a^t)) \) = \( (R, R) \) for some \( \tau \leq t - 1 \), respectively), we have that

Supplementary Material

Bounding equilibrium payoffs 5
1. for each \(h_i^t\) with \(\tilde{\omega}_{i,t}(a^t) = R\) and \((\tilde{\omega}_{i,t}(a^\tau), \tilde{\omega}_{j,t}(a^\tau)) = (R, R)\) for each \(\tau \leq t - 1\),

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

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

is small for small \(p_{\tau}\) with \(\tau \geq t + 1\); and

2. for each \(h_i^t\) with \(\tilde{\omega}_{i,t}(a^t) = R\) and \((\tilde{\omega}_{i,t}(a^\tau), \tilde{\omega}_{j,t}(a^\tau)) \neq (R, R)\) for some \(\tau \leq t - 1\),

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

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

is small for small \(p_{\tau}\) with \(\tau \geq t + 1\).

Hence, \((S1)\) and \((S2)\) (and the same inequalities with \(\mu^*\) replaced with \(\mu_{\tilde{\alpha}}\)) 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 \(p_t = \min_{\tau \leq t} \bar{p}_\tau\), then the claims hold. \(\square\)

We fix \(\{p_t\}_{t=1}^{\infty}\) so that Lemma S1 holds. This fully pins down \(\mu^*\) with mediated perfect monitoring.

**Construction with perfect monitoring with cheap talk**

Given \(\mu^*\) 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 \(\mu^*\). We must pin down the following four objects: at the beginning of the game, what message \(m_{1,1}\) player \(i\) receives from the mediator; what message \(m_{1,2}\) player \(i\) sends at the beginning of period \(t\); what action \(a_{i,t}\) player \(i\) takes in period \(t\); and what message \(m_{2,2}\) player \(i\) sends at the end of period \(t\).

**Intuitive argument** As in \(\mu^*\), at the beginning of the game, for each \(i, t, a^t\), the mediator draws \(r_{i,t}^{\text{punish}}(a^t)\) according to \(a_t^{\text{at}}\). In addition, with \(p_t > 0\) pinned down in Lemma S1, she draws \(\omega_t \in \{R, P\}^2\) and \(r_t(a^t)\) as in \(\mu^*\) for each \(t\) and \(a^t\). She then defines \(\tilde{\omega}_t(a^t)\) from \(a^t, r_t(a^t)\), and \(\omega_t\) as in \(\mu^*\).

Intuitively, the mediator sends all the information about

\[
((\tilde{\omega}_t(a^t), r_t(a^t), r_{1,t}^{\text{punish}}(a^t), r_{2,t}^{\text{punish}}(a^t)), \omega_t | a_t \in A^{t-1})_{t=1}^{\infty}
\]

through the initial messages \((m_{1,1}^{\text{mediator}}, m_{2,1}^{\text{mediator}})\). In particular, the mediator directly sends \((r_{1,t}^{\text{punish}}(a^t))_{a^t \in A^{t-1}}^{\infty}\) to player \(i\) as a part of \(m_{1,1}^{\text{mediator}}\). Hence, we focus on how
we replicate the role of the mediator in $\mu^*$ of sending $(\tilde{\omega}_t(a'), r_t(a'))$ in each period, depending on realized history $a'$.

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 $\mu^*$ 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 $\tilde{\omega}_{j, t} = P$ in $\mu^*$, 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 et al. (2012, henceforth HST). In particular, without loss, let $A_t = \{1, \ldots, n_i\}$ be player $i$’s action set. We can view $r_{i,t}(a')$ as an element of $\{1, \ldots, n_i\}$. The mediator at the beginning of the game draws $r_t(a')$ for each $a'$.

Instead of sending $r_{i,t}(a')$ directly to player $i$, the mediator encodes $r_{i,t}(a')$ 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 $p_b$ in HST. Let $\mathbb{Z}_{p^t} \equiv \{1, \ldots, p^t\}$. The mediator draws $x_{i,t}(a')$ uniformly and independently from $\mathbb{Z}_{p^t}$ for each $i, t$, and $a'$. Given them, she defines

$$y_{i,t}(a') \equiv x_{i,t}(a') + r_{i,t}(a') \mod n_i.$$

(S5)

Intuitively, $y_{i,t}(a')$ is the “encoded instruction” of $r_{i,t}(a')$, and to obtain $r_{i,t}(a')$ from $y_{i,t}(a')$, player $i$ needs to know $x_{i,t}(a')$. The mediator gives $((y_{i,t}(a'))_{a' \in A^t})_t \to \infty$ player $i$ as a part of $m^\text{mediator}_i$. At the same time, she gives $((x_{i,t}(a'))_{a' \in A^t})_t \to \infty$ player $j$ as a part of $m^\text{mediator}_j$. At the beginning of period $t$, player $j$ sends $x_{i,t}(a')$ by cheap talk as a part of $m^\text{mediator}_{ij}$. Based on the realized action $a'$, so that player $i$ does not know $r_{i,t}(a')$ 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.)

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 $\alpha_{i,t}(a')$ and $\beta_{i,t}(a')$ uniformly and independently from $\mathbb{Z}_{p^t}$, and defines

$$u_{i,t}(a') \equiv \alpha_{i,t}(a') \times x_{i,t}(a') + \beta_{i,t}(a') \mod p^t.$$  

(S6)

The mediator gives $x_{i,t}(a')$ and $u_{i,t}(a')$ to player $j$ while she gives $\alpha_{i,t}(a')$ and $\beta_{i,t}(a')$ to player $i$. In period $t$, player $j$ is supposed to send $x_{i,t}(a')$ and $u_{i,t}(a')$ to player $i$. If player $i$ receives $x_{i,t}(a')$ and $u_{i,t}(a')$ with

$$u_{i,t}(a') \neq \alpha_{i,t}(a') \times x_{i,t}(a') + \beta_{i,t}(a') \mod p^t,$$

(S7)

then player $i$ interprets that player $j$ has deviated. For sufficiently large $N^t$, since player $j$ does not know $\alpha_{i,t}(a')$ and $\beta_{i,t}(a')$, if player $j$ tells a lie about $x_{i,t}(a')$, then with a high probability, player $j$ creates a situation where (S7) holds.
Since HST considers Nash equilibrium, they let player \( i \) minimax player \( j \) forever after (S7) holds. However, since we consider sequential equilibrium, as in the proof of Lemma 2, 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 0 to the event that player \( i \) punishes player \( j \).

To this end, we consider the following coordination: First, if and only if \( \bar{\omega}_{i,t}(a') = R \), the mediator defines \( u^i_{t,t}(a') \) as (S6). Otherwise, \( u^i_{t,t}(a') \) is randomly drawn. That is,

\[
\begin{align*}
    u^i_{t,t}(a') & = \begin{cases} 
    \alpha^i_{t,t}(a') \times x^i_{t,t}(a') + \beta^i_{t,t}(a') \pmod{p'} & \text{if } \bar{\omega}_{i,t}(a') = R, \\
    \text{uniformly distributed over } \mathbb{Z}_{p'} & \text{if } \bar{\omega}_{i,t}(a') = P.
\end{cases} \tag{S8}
\end{align*}
\]

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

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

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

\[
    u^i_{t,t}(n, a') = \begin{cases} 
    u^i_{t,t}(a') & \text{if } n = n^i_{t,t}(a'), \\
    u^i_{t,t}(n, a') & \text{if otherwise},
\end{cases}
\]

that is, \( u^i_{t,t}(n, a') \) corresponds to \( u^i_{t,t}(a') \) with (S8) only if \( n = n^i_{t,t}(a') \). For other \( n \), \( u^i_{t,t}(n, a') \) is completely random.

The mediator sends \( n^i_{t,t}(a') \) to player \( i \), and sends \( \{ u^i_{t,t}(n, a') \}_{n \in \mathbb{Z}_{p'}} \) to player \( j \). In addition, the mediator sends \( n^i_{t,t}(a') \) to player \( j \), where

\[
    n^i_{t,t}(a') = \begin{cases} 
    n^i_{t,t}(a') & \text{if } \omega_{i,t-1}(a'^{-1}) = P, \\
    \text{uniformly distributed over } \mathbb{Z}_{p'} & \text{if } \omega_{i,t-1}(a'^{-1}) = R
\end{cases}
\]

is equal to \( n^i_{t,t}(a') \) if and only if last-period \( \omega_{i,t-1}(a'^{-1}) \) is equal to \( P \).
Supplementary Material

Bounding equilibrium payoffs

9

In period $t$, player $j$ is asked to send $x_{i,t}^j(a^t)$ and $u_{i,t}^j(n, a^t)$ with $n = n_{i,t}^j(a^t)$, that is, send $x_{i,t}^j(a^t)$ and $u_{i,t}^j(a^t)$. If and only if player $j$’s messages $\hat{x}_{i,t}^j(a^t)$ and $\hat{u}_{i,t}^j(a^t)$ satisfy
\[ \hat{u}_{i,t}^j(a^t) = \alpha_{i,t}^j(a^t) \times \hat{x}_{i,t}^j(a^t) + \beta_{i,t}^j(a^t) \pmod{p^t}, \]
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 $n_{i,t+1}^j(a^{t+1})$ to send the correct $u_{i,t+1}^j(n, a^{t+1})$ in the next period. Hence, she sends $n_{i,t+1}^j(a^{t+1})$ to player $j$. If player $i$ has $\hat{\omega}_{i,t}(a^t) = P$, then she believes that player $j$ knows $n_{i,t+1}^j(a^{t+1})$ and does not send $n_{i,t+1}^j(a^{t+1})$.

Given this coordination, once player $j$ creates a situation with $\hat{\omega}_{i,t}(a^t) = R$ but $\hat{\omega}_{i,t}(a^t) = P$, then player $j$ cannot receive $n_{i,t+1}^j(a^{t+1})$. Without knowing $n_{i,t+1}^j(a^{t+1})$, with a large $N^t$, with a high probability, player $j$ cannot know which $u_{i,t+1}^j(n, a^{t+1})$ she should send. Then, again, she will create a situation with
\[ \hat{u}_{i,t+1}^j(a^{t+1}) \neq \alpha_{i,t+1}^j(a^{t+1}) \times \hat{x}_{i,t+1}^j(a^{t+1}) + \beta_{i,t+1}^j(a^{t+1}) \pmod{p^{t+1}}, \]
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 $\hat{\omega}_{i,\tau}(a^{\tau}) = P$ for each $\tau \geq t$ for $a^{\tau}$ corresponding to the realized history. With $\hat{\omega}_{i,\tau}(a^{\tau}) = P$, so as to avoid $\hat{\omega}_{i,\tau}(a^{\tau}) = P$, player $j$ needs to create a situation
\[ \hat{u}_{i,\tau}^j(a^{\tau}) = \alpha_{i,\tau}^j(a^{\tau}) \times \hat{x}_{i,\tau}^j(a^{\tau}) + \beta_{i,\tau}^j(a^{\tau}) \pmod{p^{\tau}} \]
without knowing $\alpha_{i,\tau}^j(a^{\tau})$ and $\beta_{i,\tau}^j(a^{\tau})$ while the mediator’s message does not tell her what is $\alpha_{i,\tau}^j(a^{\tau}) \times x_{i,\tau}^j(a^{\tau}) + \beta_{i,\tau}^j(a^{\tau}) \pmod{p^{\tau}}$ by (S8). Hence, for sufficiently large $N^\tau$, 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 a sufficiently long time. Lemma S1 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 $\mu^*$, 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 $\alpha_{i,t}^{\varepsilon(t)}$; then she draws $\omega_t \in \{R, P\}^2$ and $r_t(a^t)$ for each $t$ and $a^t$; and then she defines $\hat{\omega}_t(a^t)$ from $a^t$, $r_t(a^t)$, and $\omega_t$ as in $\mu^*$. For each $t$ and $a^t$, she draws $x_{i,t}^j(a^t)$ uniformly and independently from $\mathbb{Z}_{p^t}$. Given them, she defines
\[ y_{i,t}^j(a^t) \equiv x_{i,t}^j(a^t) + r_{i,t}(a^t) \pmod{n_t}, \]
so that (S5) holds.

The mediator draws $\alpha_{i,t}^j(a^t)$, $\beta_{i,t}^j(a^t)$, $\hat{\alpha}_{i,t}^j(a^t)$, $\hat{\beta}_{i,t}^j(a^t)$, $v_{i,t}^j(n, a^t)$ for each $n \in \mathbb{Z}_{p^t}$, $n_{i,t}^j(a^t)$, and $\hat{n}_{i,t}^j(a^t)$ from the uniform distribution over $\mathbb{Z}_{p^t}$ independently for each player $i$, each period $t$, and each $a^t$. 
As in (S8), the mediator defines

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

In addition, the mediator defines

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

and

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

as explained above.

Let us now define the equilibrium:

(i) At the beginning of the game, the mediator sends

$$m^\text{mediator}_i \equiv \begin{pmatrix} 
y^j_{i,t}(a^t), \alpha^j_{i,t}(a^t), \beta^j_{i,t}(a^t), r^\text{punish}_{i,t}(a^t), 
n^j_{i,t}(a^t), n^j_{j,t}(a^t), (u^j_{i,t}(n, a^t))_{n \in \mathbb{Z}_{p^t}}, x^j_{i,t}(a^t) 
\end{pmatrix}_{a^t \in A^{t-1}}$$

to each player $i$.

(ii) In each period $t$, the stage game proceeds as follows: In equilibrium,

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

and

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

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

(a) Given player $i$’s history $(m^\text{mediator}_i, m^\text{1st}_1, a_1, m^\text{2nd}_{t-1}, a_{t-1})$, each player $i$ sends the first message $m^\text{1st}_{i,t}$ simultaneously. If player $i$ herself has not yet deviated, then

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

Let $m^\text{1st}_{i,t}(u)$ be the first element of $m^\text{1st}_{i,t}$ (that is, either $u^j_{i,t}(m^\text{2nd}_{j,t-1}(a^t))$ or $u^j_{i,t}(n^j_{j,t}(a^t), a^t)$ on equilibrium), and let $m^\text{1st}_{i,t}(x)$ be the second element $(x^j_{i,t}(a^t)$ on equilibrium). As a result, the profile of the messages $m^\text{1st}_{i,t}$ becomes common knowledge.
If
\[ m^1_{i,t}(u) \neq \alpha^i_{i,t}(a^i) \times m^1_{j,t}(x) + \beta^i_{i,t}(a^i) \pmod{p^t}, \] (S10)
then player \( i \) interprets \( \hat{\omega}_{i,t}(a^i) = P \). Otherwise, \( \hat{\omega}_{i,t}(a^i) = R \).

(b) Given player \( i \)'s history \((m^1_{i,1}, m^1_{i,2}, m^1_{i,3}, \ldots, m^1_{i,t})\), 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
\[
r_{i,t} = \begin{cases} 
    y^i_{i,t}(a^i) - m^1_{j,t}(x) \pmod{n_i} & \text{if } \hat{\omega}_{i,t}(a^i) = R, \\
    r^\text{punish}_{i,t}(a^i) & \text{if } \hat{\omega}_{i,t}(a^i) = P.
\end{cases} \quad \text{(S11)}
\]

Recall that \( y^i_{i,t}(a^i) \equiv x^i_{i,t}(a^i) + r_{i,t}(a^i) \pmod{n_i} \) by (S5). By (S9), therefore, player \( i \) takes \( r^\text{punish}_{i,t}(a^i) \) if \( \hat{\omega}_{i,t}(a^i) = R \) and \( r^\text{punish}_{i,t}(a^i) \) if \( \hat{\omega}_{i,t}(a^i) = P \) on the equilibrium path, as in \( \mu^* \).

(c) Given player \( i \)'s history \((m^1_{i,1}, m^1_{i,2}, m^1_{i,3}, \ldots, m^1_{i,t}, a_t)\), each player \( i \) sends the second message \( m^2_{i,t} \) simultaneously. If player \( i \) herself has not yet deviated, then
\[
m^2_{i,t} = \begin{cases} 
    n^2_{i,t+1}(a^i + 1) & \text{if } \hat{\omega}_{i,t}(a^i) = R, \\
    \{\text{babble}\} & \text{if } \hat{\omega}_{i,t}(a^i) = P.
\end{cases}
\]

As a result, the profile of the messages \( m^2_{i,t} \) becomes common knowledge. Note that \( \hat{\omega}_{i,t}(a^i) \) becomes common knowledge as well on equilibrium path.

**Incentive compatibility**

The above equilibrium has full support: Since \( \hat{\omega}_i(a^i) \) and \( r_i(a^i) \) have full support, \((m^1_{i,1}, m^2_{i,1}, a_t, m^2_{i,2}, \ldots)\) 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 \( m^1_{i,t-1} \). In equilibrium, player \( i \) sends
\[ m^1_{i,t-1} = \begin{cases} 
    u^1_{i,t}(m^2_{i,t-1}, a^i), x^i_{i,t}(a^i) & \text{if } m^2_{i,t-1} \neq \{\text{babble}\}, \\
    u^1_{i,t}(n^j_{i,t}(a^j), a^i), x^i_{i,t}(a^i) & \text{if } m^2_{i,t-1} = \{\text{babble}\}.
\end{cases} \]

The random variables possessed by player \( i \) are independent of those possessed by player \( j \) given \((m^1_{i,1}, a_t, m^2_{i,2}, \ldots)\), except that (i) \( u^1_{i,t}(a^i) = \alpha^i_{i,t}(a^i) \times x^i_{i,t}(a^i) + \beta^i_{i,t}(a^i) \pmod{p^t} \) if \( \hat{\omega}_{i,t}(a^i) = R \), (ii) \( u^1_{i,t}(a^i) = \alpha^j_{i,t}(a^j) \times x^i_{i,t}(a^i) + \beta^j_{i,t}(a^i) \pmod{p^t} \) if \( \hat{\omega}_{i,t}(a^i) = R \), (iii) \( n^j_{i,t}(a^i) = n^j_{i,t}(a^i) \) if \( \omega_{i,t-1}(a^{i-1}) = P \) while \( n^j_{i,t}(a^i) \) if \( \omega_{i,t-1}(a^{i-1}) = R \), and (iv) \( n^j_{i,t}(a^i) = n^j_{i,t}(a^i) \) if \( \omega_{j,t-1}(a^{j-1}) = P \) while \( n^j_{i,t}(a^i) \) if \( \omega_{j,t-1}(a^{j-1}) = R \).
Since \( \alpha^j_i(a') \), \( \beta^j_i(a') \), \( \bar{u}^j_i(a') \), \( v^j_i(n, a') \), \( m^j_{i,t}(a') \), and \( n^j_{i,t}(a') \) are uniform and independent, player \( i \) cannot learn \( \hat{\bar{\omega}}_{i,t}(a') \), \( r_i(a^\tau) \), or \( r_j(a^\tau) \) with \( \tau \geq t \). Hence, player \( i \) believes at the time when she sends \( m^1_{i,t} \) that her equilibrium value is equal to

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

where \( h^j_i \) is as if player \( i \) observed \( (r^\text{punish}_{i,t}(a'))_{a' \in A^t}^{\infty}, a', (\hat{\omega}_\tau(a^\tau))_{\tau = 1}^{\tau-1}, \) and \( r_{i,t}(a') \), and believed that \( r_\tau(a^\tau) = a_\tau \) for each \( \tau = 1, \ldots, t - 1 \) with \( \mu^* \) with mediated perfect monitoring.

Alternatively, for each \( \epsilon > 0 \), for a sufficiently large \( N^t \), if player \( i \) tells a lie in at least one element \( m^1_{i,t} \), then with probability \( 1 - \epsilon \), player \( i \) creates a situation

\[
m^1_{i,t}(u) \neq \alpha^j_{i,t}(a') \times m^1_{i,t}(x) + \beta^j_{i,t}(a') \mod p^t.
\]

Hence, \( \text{(S10)} \) (with indices \( i \) and \( j \) reversed) implies that \( \hat{\omega}_{i,t}(a') = P \).

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

\[
m^1_{i,t+1}(u) \neq \alpha^j_{i,t+1}(a^{t+1}) \times m^1_{i,t+1}(x) + \beta^j_{i,t+1}(a^{t+1}) \mod p^{t+1}
\]

with probability \( 1 - \epsilon \). Then, by \( \text{(S10)} \) (with indices \( i \) and \( j \) reversed), player \( j \) will have \( \hat{\omega}_{j,t+1}(a^{t+1}) = P \).

Recursively, if \( \hat{\omega}_{j,t}(a^\tau) = R, \) for each \( \tau = t, \ldots, t + T - 1 \), then player \( i \) will induce \( \hat{\omega}_{j,\tau}(a^\tau) = P \) for each \( \tau = t, \ldots, t + T - 1 \) with a high probability. Hence, for \( \epsilon_t > 0 \) and \( T \) fixed in \( \text{(S1)} \) and \( \text{(S2)} \), for sufficiently large \( N^t \), if \( N^\tau \geq N^t \) for each \( \tau \geq t \), then player \( i \) will be punished for the subsequent \( T \) periods regardless of player \( i \)'s continuation strategy with probability no less than \( 1 - \epsilon_t - \sum_{\tau=t}^{t+T-1} p_\tau \). \( \sum_{\tau=t}^{t+T-1} p_\tau \) represents the maximum probability of having \( \hat{\omega}_{i,t}(a^\tau) = P \) for some \( \tau \) for subsequent \( T \) periods.) Equation \( \text{(S4)} \) implies that telling a lie gives a strictly lower payoff than the equilibrium payoff. Therefore, it is optimal to tell the truth about \( m^1_{i,t} \). In \( \text{(S4)} \), we have shown interim incentive compatibility after knowing \( \hat{\bar{\omega}}_{i,t}(a^\tau) \) and \( r_{i,t} \), while here we consider \( h^j_i \) before \( \hat{\bar{\omega}}_{i,t}(a^\tau) \) and \( r_{i,t} \). Taking the expectation with respect to \( \hat{\bar{\omega}}_{i,t}(a^\tau) \) and \( r_{i,t}, \) \( \text{(S4)} \) ensures incentive compatibility before knowing \( \hat{\bar{\omega}}_{i,t}(a^\tau) \) and \( r_{i,t} \).

Second, consider player \( i \)'s incentive to take the action \( a_{i,t} = r_{i,t} \) according to \( \text{(S11)} \) if player \( i \) follows the equilibrium strategy until she sends \( m^1_{i,t} \). 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

\[
(1 - \delta) \mathbb{E}^\mu^*[u_i(a_t) \mid h^j_i] + \delta \mathbb{E}^\mu^* \left[ (1 - \delta) \sum_{\tau=t+1}^\infty \delta^{\tau-t-1} u_i(a_t) \mid h^j_i \right],
\]
where $h_t^i$ is as if player $i$ observed $(r_t^{\text{punish}}(a^t))_{a^t \in A^t-1}^{t=1} a^t$, $(\tilde{w}_t(\pi_t^i(t)))_{t=1}^{\infty}$, \(\tilde{w}_{i,t}(a^t)\), and $r_{i,t}$, and believed that $r_t(\pi_t^i) = a_r$ for each $\tau = 1, \ldots, t-1$ with $\mu^*$ with mediated perfect monitoring. (Compared to the time when player $i$ sends $m_{i,t}^{1\text{st}}$, player $i$ now knows $\tilde{w}_{i,t}(a^t)$ and $r_{i,t}$ on the equilibrium path.)

If player $i$ deviates from $a_{i,t}$, then $\tilde{w}_{i,t}(a^t) = P$ by definition for each $\tau \geq t+1$ and $a^t$ that is compatible with $a^t$ (that is, $a^t = (a^t, a_t, \ldots, a_{t-1})$ for some $a_t, \ldots, a_{t-1}$). To avoid being minmaxed in period $\tau$, player $i$ needs to induce $\tilde{w}_{i,t}(a^t) = R$ although $\tilde{w}_{i,t}(a^t) = P$. Given $\tilde{w}_{i,t}(a^t) = P$, since $a_{t,t}(a^t)$, $\beta_{t,t}(a^t)$, $\tilde{u}_{t,t}(a^t)$, $\nu_{t,t}(n,a^t)$ $n_{t,t}(a^t)$, and $\tilde{n}_{t,t}(a^t)$ are uniform and independent (conditional on the other variables), regardless of player $i$’s continuation strategy, by (S10) (with indices $i$ and $j$ reversed), player $i$ will send $m_{i,t}^{1\text{st}}$ with

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

with a high probability.

Hence, for sufficiently large $\tilde{N}$, if $N^t \geq \tilde{N}$ for each $\tau \geq t$, then player $i$ will be punished for the next $T_i$ periods regardless of player $i$’s continuation strategy with probability no less than $1 - \epsilon_i$. By (S3), 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 $m_{i,t}^{2\text{nd}}$. Regardless of $m_{i,t}^{2\text{nd}}$, 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 $\tilde{w}_{i,t}(a^t)$. If player $i$ tells the truth, then $m_{i,t}^{2\text{nd}} = n_{i,t+1}(a^t+1) \neq \text{(babble)}$ if and only if $\tilde{w}_{i,t}(a^t) = R$. Moreover, player $j$ sends

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

Since $n_{j,t+1}^i(a^t+1) = n_{i,t+1}^j(a^t+1)$ if $\tilde{w}_{i,t}(a^t) = P$, in total, if player $i$ tells the truth, then player $i$ knows $u_{j,t+1}(m_{i,t}^{2\text{nd}}(a^t+1), a^t+1)$ and $x_{j,t+1}^i(a^t+1)$. This is sufficient information to infer $\tilde{w}_{i,t+1}(a^t+1)$ and $r_{i,t+1}(a^t+1)$ correctly.

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

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

Since $\alpha_{i,t+1}(a^t+1)$, $\beta_{i,t+1}(a^t+1)$, $\tilde{u}_{i,t+1}(a^t+1)$, $\nu_{i,t+1}(n,a^t+1)$ $n_{i,t+1}(a^t+1)$, and $\tilde{n}_{i,t+1}(a^t+1)$ are uniform and independent conditional on $\tilde{w}_{i,t+1}(a^t+1)$ and $r_{i,t+1}(a^t+1)$, $u_{i,t+1}(n,a^t+1)$ and $x_{j,t+1}^i(a^t+1)$ 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 $\tilde{w}_{i,t+1}(a^t+1)$ and $r_{i,t+1}(a^t+1)$. Since telling the truth informs player $i$ of $\tilde{w}_{i,t+1}(a^t+1)$ and $r_{i,t+1}(a^t+1)$, there is no gain from telling a lie.

Co-editor George J. Mailath handled this manuscript.

Manuscript received 26 August, 2015; final version accepted 25 April, 2016; available online 25 April, 2016.