The Characterization of the Limit Equilibrium Payoff Set with a Mediator and General Monitoring

Takuo Sugaya[*†]
Stanford Graduate School of Business

May 7, 2014

Abstract

We characterize an upper bound of the set of sequential equilibrium payoffs in repeated games with a mediator and general (private) monitoring

We then show that if each player can observe her own realized payoff, then the upper bound is equal to the limit of the sequential equilibrium payoff set as a discount factor goes to one.

We also provide sufficient conditions to dispense with the mediator.

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

Keywords: repeated game, folk theorem, private monitoring

[*] tsugaya@gsb.stanford.edu
[†] I thank ??? for useful comments. The remaining errors are my own responsibility.

1 Introduction

Given that Fudenberg and Maskin (1986), Fudenberg, Levine, and Maskin (1994), Hörner and Olszewski (2006), and Sugaya (2014a, 2014b) prove folk theorems in an increasingly general conditions, a natural question to arise is what equilibrium payoff is supportable if one of sufficient conditions for the folk theorem is not satisfied.

The sufficient conditions for the folk theorem mainly consist of the following two assumptions: A discount factor converges to one, and identifiability conditions are satisfied. More specifically, Sugaya (2014a, 2014b) assume the following two identifiability conditions. For the two-player game, he assumes individual identifiability: If player j deviates, then player $i \neq j$ can statistically detect the deviation. For the game with more than two players, he assumes pairwise identifiability: For players i, j, and n (all different), player i can statistically identify whether player j or player n is more likely to have deviated in order to minimax the actual deviator.

Hence, there are two possible ways in which the sufficient conditions for the folk theorem are not satisfied. First, the discount factor is not converging to one; and second, the identifiability conditions are not satisfied. In this paper, we focus on the latter case. See Sugaya and Wolitzky (2014) for the analysis in the former case.

Specifically, we consider the case in which the discount factor is still converging to one, but the monitoring structure of the game is fully general. Our ultimate goal is to obtain the full characterization of the limit sequential equilibrium payoff set.

However, the analysis of the general game is very hard, partially because the available correlation device becomes richer and richer as time proceeds. Note that, with private monitoring with correlated signals, players obtain endogenous correlation devices through their own private signals, and the set of possible endogenous correlation devices increases as periods go on. At the same time, to create a specific correlation device from signals, we may need to incentivize players to take some specific action profile. The incentive to take the action profile depends on their belief about the other players histories, which in turn depend on the endogenous correlation.

To avoid the complication coming from this endogenous correlation, we assume the availability of the mediator, who can stand in and create arbitrary correlation. Note that, even with the mediator, with general monitoring, the equilibrium may not have the recursive structure. In each period, a player does not know the history of the mediator or other players. Hence, the player needs to calculate the beliefs over all the possible histories of the mediator and other players. Since the structure of the information set for each player gets more and more complicated as time goes on, the full characterization of the equilibrium payoff set with the mediator is hard to obtain.

Nonetheless, we will obtain an upper bound for the equilibrium payoff set. The key property to derive this upper bound is that, in the repeated game with the mediator, the set of equilibrium payoffs should be no smaller than that of ex ante expected continuation payoffs. To see this, take an equilibrium and the continuation payoff in the equilibrium from period 2 after each possible history. In addition, calculate the ex ante expected continuation payoff, using the distribution of histories of period 1 evaluated at the beginning of the game. Then, the mediator can create another equilibrium whose equilibrium payoff from period 1 is equal to the ex ante expected continuation payoff from period 2 in the original equilibrium as follows. The mediator draws a "fictitious history" according to the distribution of histories of period 1 evaluated at the beginning of the original equilibrium. The mediator starts the new equilibrium as if it started from period 2 of the original equilibrium, using the fictitious history. If the original equilibrium is sequentially rational, then so is the new one. Therefore, the set of ex ante equilibrium payoff should be no smaller than that of ex ante expected continuation payoff.[1]

The remaining question is how tight this characterized upper bound is. To answer this question, we show that, if each player can observe her realized own payoffs, then this characterized upper bound coincides with the limit of the sequential equilibrium payoff set as the discount factor goes to one. That is, suppose that the distribution of realized payoffs

[1]Note that it is *not* true that the set of *interim* continuation payoffs after a realization of a history of period 1 should be included in the set of equilibrium payoffs, since the equilibrium lacks the recursive structure.

is determined by the action profile taken by the players. If each player can observe the realized own payoffs, then we show that the characterized upper bound is tight.

To obtain the intuition, let us consider the two channels through which some player i's deviation affects the equilibrium. The first channel is to affect how player i monitors the other players. For example, if player i deviates from the action which generates informative signals about the other players' deviations to the one which does not, then player i and the mediator may lose the ability to discourage the other players from deviating. However, the mediator can discourage player i from all the deviations which affect player i's monitoring about the other players as follows. The mediator mixes recommendations for the other players, and then asks player i to infer which actions are recommended to the other players. If player i fails to infer the other players' actions, then the mediator punishes player i. By doing so, the mediator can discourage player i from deviations which affect the monitoring of the other players. In the context of static mechanism design, Rahman (2012) establishes the mechanism immune to the deviations of the monitor by the same reasoning.

The second channel is to affect the other players' instantaneous utilities. However, if the other players can observe their own realized payoffs, then as long as player i's deviation affects the other players' instantaneous utilities, the other players can statistically detect such a deviation through realized payoffs. Therefore, the mediator can discourage player i from deviating if it affects the other players' instantaneous utilities.

In total, the mediator can discourage player i from deviating if it affects the other players' incentives. In other words, all the undetectable deviations are "harmless," and so we can just let player i "deviate" if the original strategy is not incentive compatible.

Finally, one may wonder if we can dispense with the mediator. In the context of the static game with incomplete information, Gerardi (2004) establishes the method to replace the mediation with pre-play communication among players if there are at least five players. By seeing each player's history in each period as the player's "type" in Gerardi's mechanism, we can apply his method to replace the mediation with pre-play communication among players if there are at least five players.

The rest of the paper is organized as follows: Section 2 introduces the model, and Section 3 derives an upper bound. Section 4 establishes that, if each player can observe her own realized payoffs, then this upper bound coincides with the limit sequential equilibrium payoff set as the discount factor goes to one. Section 5 discusses the case in which some player may not observe her own realized payoffs. Finally, Section 6 proves that we can replace the mediation with pre-play communication among players if there are at least five players.

2 Model

2.1 A Stage Game

A stage game is given by $\{I, (A_i, Y_i, u_i)_{i \in I}, q\}$. Here, $I = \{1, ..., N\}$ is the set of players, and A_i is a finite set of player i's actions. For a subset of players $J \subset I$, let $-J$ denote the players not included in J: $-J \equiv I \setminus J$. Player i cannot observe players $-i$'s actions. Instead, she observes a private signal $y_i \in Y_i$ with a finite Y_i. The signal profile $y \equiv (y_i)_{i \in I}$ is distributed according to a joint conditional distribution function $q(y|a)$ given action profile $a \equiv (a_i)_{i \in I} \in A \equiv \prod_{i \in I} A_i$. As usual, let $q_i(y_i|a)$ be the marginal distribution of y_i; and let

$$q_i(Y_i|a) \equiv (q_i(y_i|a))_{y_i \in Y_i}$$

be the vector expression of the marginal distribution.

Given the action profile $a \in A$, player i's ex ante utility function is given by $u_i(a)$. The vector $u(a) = (u_i(a))_{i \in I}$ is the utility function profile.

If player i observes her realized own payoffs, then such information should be seen as a private signal and included in y_i. That is, if player i observes her realized own payoff $\tilde{u}_i \in \tilde{U}_i$, then we see $(y_i, \tilde{u}_i)$ as player i's private signal. Here, $\tilde{U}_i$ is the finite set of possible realizations of player i's payoffs. Hence, the assumption that player i observes realized own payoffs is equivalent to the assumption that player i's realized payoff is a deterministic function $\tilde{u}_i(a_i, y_i)$ of a_i and y_i. The ex ante utility function is determined by $u_i(a) =$

$\sum_{y_i \in y_i} q_i(y_i|a)\tilde{u}_i(a_i, y_i)$.

As usual, we can extend the argument of the utility function to the correlated strategy: $u_i(\mu)$ and $u(\mu)$ are the expected values of $u_i(a)$ and $u(a)$ when a is distributed according to $\mu \in \Delta(A)$, respectively. Similarly, $q(y|\mu)$ and $q_i(y_i|\mu)$ are the distributions of a signal profile y and player i's signal y_i when a is distributed according to $\mu \in \Delta(A)$, respectively. Again, let $q_i(Y_i|\mu) \equiv (q_i(y_i|\mu))_{y_i \in Y_i}$ denote the vector expression of $q_i(y_i|\mu)$.

2.2 A Repeated Game with a Mediator

We consider the infinite repetition of the stage game defined above, where a mediator has been added. See Section 6 for sufficient conditions to dispense with the mediator. The mediator can privately communicate with the players. With the mediator, Forges (1986) shows that without loss, we can consider the following canonical communication equilibrium: In each period t,

1. The mediator draws a recommendation profile $r_t \in A$ according to $\mu_t \in \Delta(A)$.

2. The mediator sends a recommendation $r_{i,t} \in A_i$ to each player i. Let $\mu_{i,t}$ denote the marginal distribution of $r_{i,t}$. Since player i cannot observe the recommendations to players $-i$, player i believes that $r_{-i,t}$ is distributed according to $\mu_{-i,t}|_{r_i}$, where $\mu_{-i,t}|_{r_{i,t}}$ is the conditional distribution of $r_{-i,t}$ given that r_t is distributed according to μ_t and $r_{i,t}$ is realized.

3. Given $r_{i,t}$, each player i takes $a_{i,t} \in A_i$. On equilibrium path, each player is obedient: $a_{i,t} = r_{i,t}$.

4. Given that players $-i$ follow the recommendations, the private signal profile $y_t = (y_{i,t})_{i \in I}$ is distributed according to the joint conditional distribution function $q(y_t|a_{i,t}, r_{-i,t})$.

5. Each player i observes her own private signal $y_{i,t}$. Since player i cannot observe players $-i$'s signals, player i believes that $y_{-i,t}$ is distributed according to $q(y_{-i,t}|r_{i,t}, y_{i,t}, \mu_{-i,t}|_{r_{i,t}})$,

where $q(y_{-i,t}|r_{i,t}, y_{i,t}, \mu_{-i,t}|_{r_{i,t}})$ is the conditional distribution of $y_{-i,t}$ given that y_t is distributed according to $q(y_t|r_t)$, $r_{-i,t}$ is distributed according to $\mu_{-i}|_{r_i}$, and $y_{i,t}$ is realized.

6. Each player i reports $m_{i,t} \in Y_i$ privately to the mediator. Here, $m_{i,t}$ represents what signal player i observes in period t according to player i's message. On equilibrium path, each player is truthful: $m_{i,t} = y_{i,t}$.

In period t, the mediator's history is $h_m^t \equiv (r_\tau, m_\tau)_{\tau=1}^{t-1}$ with $h_m^1 = \{\emptyset\}$, and her strategy is a mapping from her history to a mixture over recommendation profiles: $\sigma_m : h_m^t \mapsto \mu_t \in \Delta(A)$ for each $t \geq 1$.

To define player i's strategy in the repeated game, it will be useful to define the set of player i's stage-game strategies. Let $S_i \ni s_i$ denote the set of player i's pure strategies such that (i) for each $r_i \in A_i$, player i takes a recommendation-contingent action $s_i(r_i) \in A_i$, and that (ii) for each $r_i \in A_i$, $a_i \in A_i$, and $y_i \in Y_i$, after being recommended r_i, taking a_i, and observing y_i, player i sends a recommendation-, action-, and signal-contingent message $s_i(r_i, a_i, y_i) \in Y_i$.

Let $\Sigma_i \ni \sigma_i$ denote the set of player i's mixed strategies. Let $\sigma_i(r_i)(a_i)$ be the probability that player i takes $a_i \in A_i$ after recommended r_i; let $\sigma_i(r_i, a_i, y_i)(m_i)$ be the probability that player i sends $m_i \in Y_i$ after recommended r_i, taking a_i, and observing y_i; and let $\Sigma_i|_{r_i} \ni \sigma_i|_{r_i}$ denote the set of player i's mixed continuation strategies after player i is recommended to take r_i. In addition, let $u_i(\sigma_i, \mu)$ denote player i's ex ante payoff when the mediator recommends r according to μ and player i unilaterally deviates from the faithful strategy to σ_i; and let

$$\Pr(m|\sigma_i, r) \equiv \sum_{a_i} \sigma_i(r_i)(a_i) \sum_{y} q(y|a_i, r_{-i}) \sigma_i(r_i, a_i, y_i)(m_i) \mathbf{1}_{\{m_{-i}=y_{-i}\}}$$

be the probability that the mediator receives the message profile m after the recommendation profile r when player i unilaterally deviates to σ_i.

Especially, let s_i^* denote the "faithful" strategy such that player i obeys the recommendation and tells the truth about the signal in equilibrium: $s_i^*(r_i) = r_i$ for all $r_i \in \text{supp}(\mu_i)$

and $s_i^*(r_i, a_i, y_i) = y_i$ after $a_i = r_i$.

Given the definition of Σ_i, we define player i's strategy in the repeated game after player i's history $h_i^t \equiv (r_{i,\tau}, a_{i,\tau}, y_{i,\tau}, m_\tau)_{\tau=1}^{t-1}$ with $h_m^1 = \{\emptyset\}$ as a mapping from her history to a mixture over her stage-game strategies: $\sigma_i : h_i^t \mapsto \sigma_i(h_i^t) \in \Sigma_i$ for each $t \geq 1$. Note that $\sigma_i(h_i^t)$ specifies the following: For each possible recommendation $r_{i,t}$, the strategy $\sigma_i(h_i^t)$ specifies the mixture of player i's actions $\alpha_{i,t} \in \Delta(A_i)$. Then, given $r_{i,t}$ and $a_{i,t}$, after observing her private signal $y_{i,t}$, the strategy $\sigma_i(h_i^t)$ specifies the mixture of player i's messages $\rho_i \in \Delta(Y_i)$. Let $\sigma_i(h_i^t, r_{i,t}) \in \Delta(A_i)$ denote the mixture of player i's actions given $r_{i,t}$; and let $\sigma_i(h_i^t, r_{i,t}, a_{i,t}, y_{i,t}) \in \Delta(Y_i)$ denote the mixture of player i's messages given $r_{i,t}$, $a_{i,t}$, and $y_{i,t}$. In addition, let $\sigma_i(h_i^t, r_{i,t})(a_{i,t})$ and $\sigma_i(h_i^t, r_{i,t}, a_{i,t}, y_{i,t})(m_{i,t})$ be the probability of taking $a_{i,t}$ and sending $m_{i,t}$, respectively.

Let $E_{\text{seq}}(\delta)$ be the set of sequential equilibrium payoffs with a common discount factor δ. We will first characterize an upper bound of $E_{\text{seq}}(\delta)$: We characterize Q such that $E_{\text{seq}}(\delta) \subset Q$ for each δ. Then, we will provide a sufficient condition with which the limit sequential equilibrium payoff set coincides with this upper bound: $\lim_{\delta \to 1} E_{\text{seq}}(\delta) = Q$. Finally, we comment on the cases in which the sufficient condition does not hold.

3 Upper Bound

3.1 Characterization of the Upper Bound

We characterize an upper bound of $E_{\text{seq}}(\delta)$, denoted by Q, by the linear algorithm. To this end, it is useful to consider the set of Pareto weights, denoted by $\Lambda \equiv \{\lambda \in \mathbb{R}^N : \|\lambda\| = 1\}$. Throughout the paper, we use the Euclidean norm. For notational convenience, let $|\lambda|$ be the vector such that each ith element is $|\lambda_i|$; and let e_i be the vector whose ith component is one and other components are zero.

With a mediator, the set of sequential equilibrium payoffs is convex. Hence, we can characterize an upper bound of $E_{\text{seq}}(\delta)$ by the intersection of half-spaces, one half-space for one Pareto weight $\lambda \in \Lambda$. Since the half-space for λ is expressed by $H(\lambda) \equiv \{v \in \mathbb{R}^N :$

$\lambda \cdot v \leq h(\lambda)\}$, we want to characterize the "height" of the half-space $h(\lambda) \in \mathbb{R}$. As will be seen, the height $h(\lambda)$ is the minimum of $k(\lambda)$ and $l(\lambda)$ to be determined.

We first define $k(\lambda)$. To this end, we derive $k(\mu, \lambda)$ for each $\mu \in \Delta(A)$, and define $k(\lambda) \equiv \sup_{\mu \in \Delta(A)} k(\mu, \lambda)$. Intuitively, $k(\mu, \lambda)$ is the height of the half-space for λ when we try to implement the equilibrium strategy μ.

We now define $k(\mu, \lambda)$ by a linear programming problem for given $\mu \in \Delta(A)$ and $\lambda \in \Lambda$:

$$k(\mu, \lambda) \equiv \max_{v \in \mathbb{R}^N, x: R \times Y \to \mathbb{R}^N} \lambda \cdot v \tag{1}$$

subject to the following constraints:

1. [incentive compatibility] For each $i \in I$, the faithful strategy is optimal when the change in the continuation payoff is given by $x(r, m)$: For all $s_i \in S_i$,

$$\mathbb{E}\left[u_i(r) + x_i(r, y) | s_i^*, \mu\right] \geq \mathbb{E}\left[u_i(a_i, r_{-i}) + x_i(r, m_i, y_{-i}) | s_i, \mu\right].$$

   For the simple notation, we omit s_i^* from conditioning whenever it is not confusing.

2. [promise keeping] v is the value of the repeated game: $v = u(\mu) + \mathbb{E}\left[x(r, y) | \mu\right]$.

3. [self generation] In expectation, in equilibrium, the score $\lambda \cdot x(r, y)$ is non positive: $\mathbb{E}\left[x(r, y) | \mu\right] \leq 0$.

If there is no (v, x) satisfying the constraints 1-3, then we set $k(\mu, \lambda) = -\infty$. Given $k(\mu, \lambda)$, we define

$$k(\lambda) \equiv \sup_{\mu \in \Delta(A)} k(\mu, \lambda).$$

Since a static correlated equilibrium $\mu \in \Delta(A)$ and $x(r, m) = 0$ for all $r \in A$ and $y \in Y$ satisfy all the constraints, $k(\lambda)$ is well defined.

We second define $l(\lambda)$. Intuitively, while $k(\lambda)$ represents the restriction of the height on the half-space on equilibrium path, $l(\lambda)$ represents the restriction of the height on the half-space coming from punishment (off equilibrium path).

To define $l(\lambda)$, we derive $l(\tau, \lambda)$ for each $\tau \in \Delta(A)$, and define $l(\lambda) \equiv \sup_{\tau \in \Delta(A)} l(\tau, \lambda)$. Intuitively, $l(\tau, \lambda)$ is the punishment payoff when τ is the punishment strategy.

We now define $l(\tau, \lambda)$ by a linear programming problem for given $\tau \in \Delta(A)$ and $\lambda \in \Lambda$:

$$l(\tau, \lambda) \equiv \max_{v \in \mathbb{R}^N, x: R \times Y \to \mathbb{R}^N} \lambda \cdot v \tag{2}$$

subject to the following constraints:

1. [effective punishment] For each i and $s_i \in S_i$, the punishment strategy is not profitable:

$$v_i \geq \mathbb{E}\left[u_i(a_i, r_{-i}) + x_i(r, m) \mid s_i, \tau\right].$$

2. [self generation after unilateral deviations] For each i and $s_i \in S_i$, in expectation, the score $\lambda \cdot x(r, y)$ is non positive:

$$\mathbb{E}\left[\lambda \cdot x(r, m) \mid s_i, \tau\right] \leq 0.$$

If there is no (v, x) satisfying the constraints 1-2, then we set $l(\tau, \lambda) = -\infty$. If there exists x to satisfy the constraints 1-2 for each v, then we set $l(\tau, \lambda) = +\infty$.

Given $l(\tau, \lambda)$, we define

$$l(\lambda) \equiv \sup_{\tau \in \Delta(A)} l(\tau, \lambda).$$

Since a static correlated equilibrium $\tau \in \Delta(A)$ and $x(r, m) = 0$ for all $r \in A$ and $y \in Y$ satisfy all the constraints, $l(\lambda)$ is not $-\infty$.

These algorithms characterize the limit equilibrium payoff set, using exteremum half-spaces:

$$H(\lambda) \equiv \left\{v \in \mathbb{R}^N : \lambda \cdot v \leq \min\left\{k(\lambda), l(\lambda)\right\}\right\},$$

and

$$Q \equiv \bigcap_{\lambda \in \Lambda} H(\lambda).$$

We will prove that Q gives an upper bound of the equilibrium payoff set:

Theorem 1 *Q is an upper bound of the equilibrium payoff set: $E_{\text{seq}}(\delta) \subseteq Q$ for each $\delta \in [0,1)$.*

Proof. See Section 3.5. ■

3.2 Dual of the Characterization

Before proving Theorem 1, in order to obtain the intuition of the theorem, it is instructive to consider the dual problem of the linear programming to determine $k(\lambda)$ and $l(\lambda)$.

The next proposition provides the dual for the program to determine $k(\lambda)$:

Proposition 1 *The score $k(\lambda)$ defined in (1) is equal to $\max_{\mu \in \Delta(A)} \lambda \cdot u(\mu)$ such that, for each $i \in I$ and $\sigma_i \in \Sigma_i$ such that*

$$\Pr(m|\sigma_i, r) = \Pr(m|r) \text{ for each } r \in \operatorname{supp}(\mu) \text{ and } m \in Y, \tag{3}$$

we have

$$u_i(\sigma_i, \mu) > u_i(\mu). \tag{4}$$

Proof. See Section 7.1. ■

Rahman (2012) obtains the same condition in the context of a static mechanism design. When $\sigma_i \in \Sigma_i$ satisfies Condition (3), we say that a deviation σ_i is $\operatorname{supp}(\mu)$-*undetectable* since the distribution of the message profile is the same between σ_i and the faithful strategy.

Let

$$\mathcal{M} \equiv \left\{ \mu \in \Delta(A) : \begin{array}{c} \text{for each } i, \sigma_i \text{ with } \Pr(m|r) = \Pr(m|\sigma_i, r) \text{ for all } r \in \operatorname{supp}(\mu) \text{ and } m \in Y, \\ \text{we have } u_i(\sigma_i, \mu) \leq u_i(\mu) \end{array} \right\} \tag{5}$$

be the set of $\mu \in \Delta(A)$ satisfying the condition in Proposition 1. Let $u(\mathcal{M}) \equiv \{u(\mu)\}_{\mu \in \mathcal{M}}$ be the set of payoffs implemented by the mediator's strategy in $\mathcal{M}$.

Proposition 1 ensures that

$$\bigcap_{\lambda \in \Lambda} \left\{ v \in \mathbb{R}^N : \lambda \cdot v \leq k(\lambda) \right\} = u(\mathcal{M}). \tag{6}$$

The next proposition provides the dual for the program to determine $l(\lambda)$. To this end, it is useful to define the set of "similar deviations" for each subset of players $J \subset I$, as Renault and Tomala (2004).

Let $J \subset I$ be an arbitrary subset of players. The set of similar deviations is defined as follows:

$$SD(J) \equiv \left\{ \sigma^J = (\sigma_i)_{i \in J} \in \prod\nolimits_{j \in J} \Sigma_j : \begin{array}{c} \forall i \in J, \forall j \in J, \\ \Pr(m|\sigma_i, r) = \Pr(m|\sigma_j, r) \\ \text{for all } r \in A \text{ and } m \in Y \end{array} \right\}. \tag{7}$$

That is, if the players in J takes σ^J, then the mediator cannot distinguish who in J is more likely to have deviated.

Consider the following situation: Given the Pareto weight $\lambda \in \Lambda$, the mediator wants to minimax the Pareto weighted sum of the utilities of the players in $J = \text{supp}\,(\lambda)$, by recommending a "punishment" strategy τ. Each player in $J = \text{supp}\,(\lambda)$ is free to deviate in $SD(J)$, that is, as long as the mediator cannot tell who in J is guilty. Then, we say that

$$\min_{\tau \in \Delta(A)} \max_{\sigma \in SD(\text{supp}(\lambda))} \sum_{r \in A} \tau(r) \sum_{i \in N} |\lambda_i| \, u_i(r, \sigma_i)$$

is the harshest punishment payoff. The following proposition ensures that $l(\lambda)$ is related to the harshest punishment payoff if all the players' Pareto weights are non-positive.

Proposition 2 *The score $l(\lambda)$ defined in (2) is equal to the following:*

1. *If $\lambda_i \leq 0$ for each i, then $l(\lambda)$ is equal to the (negative) harshest punishment payoff:*

$$-\min_{\tau \in \Delta(A)} \max_{\sigma \in SD(\text{supp}(\lambda))} \sum_{r \in A} \tau(r) \sum_{i \in N} |\lambda_i| \, u_i(r, \sigma_i). \tag{8}$$

2. Otherwise, $l(\lambda) \geq k(\lambda)$ *and so* $H(\lambda) \equiv \left\{v \in \mathbb{R}^N : \lambda \cdot v \leq k(\lambda)\right\}$.

Proof. See Section 7.2. ■

Given (6) and Proposition 2, we can write

$$\begin{aligned} Q &= \bigcap_{\lambda \in \Lambda} \left\{v \in \mathbb{R}^N : \lambda \cdot v \leq \min\left\{k(\lambda), l(\lambda)\right\}\right\} \\ &= u(\mathcal{M}) \cap \left\{ v \in \mathbb{R}^N : \begin{array}{c} \text{for each } \lambda \in \Lambda \text{ with } \lambda_i \leq 0 \text{ for each } i, \\ \lambda \cdot v \leq -\min_{\tau \in \Delta(A)} \max_{\sigma \in SD(\text{supp}(\lambda))} \sum_{r \in A} \tau(r) \sum_{i \in N} |\lambda_i| \, u_i(r, \sigma_i) \end{array} \right\} \\ &= u(\mathcal{M}) \cap \left\{ v \in \mathbb{R}^N : \begin{array}{c} \text{for each } \lambda \in \Lambda, \\ |\lambda| \cdot v \geq \min_{\tau \in \Delta(A)} \max_{\sigma \in SD(\text{supp}(\lambda))} \sum_{r \in A} \tau(r) \sum_{i \in N} |\lambda_i| \, u_i(r, \sigma_i) \end{array} \right\} \\ &= u(\mathcal{M}) \cap JD. \end{aligned} \tag{9}$$

Here, for simple notation, we write

$$JD \equiv \left\{ v \in \mathbb{R}^N : \begin{array}{c} \text{for each } \lambda \in \Lambda, \\ |\lambda| \cdot v \geq \min_{\tau \in \Delta(A)} \max_{\sigma \in SD(\text{supp}(\lambda))} \sum_{r \in A} \tau(r) \sum_{i \in N} |\lambda_i| \, u_i(r, \sigma_i) \end{array} \right\}.$$

3.3 Relationship with the Repeated Games without Discounting

Renault and Tomala (2004) offer the characterization of the equilibrium payoff set in the repeated game without discounting, denoted by $E(1)$. It is instructive to compare their characterization with ours.

First, they define the undetectable deviations. Player i's deviation σ_i is undetectable if, for any $r \in A$, the distribution of m is the same:

$$\Pr(m|r) = \Pr(m|\sigma_i, r) \text{ for all } r \in A \text{ and } m \in Y.$$

Second, they define the set of the mediator's recommendations, denoted by $\mathcal{P}$, such that, for each player i and undetectable deviation σ_i, the deviation σ_i is not profitable:

$$\mathcal{P} \equiv \left\{ \mu \in \Delta(A) : \begin{array}{c} \text{for each } i, \sigma_i \text{ with } \Pr(m|r) = \Pr(m|\sigma_i, r) \text{ for all } r \in A \text{ and } m \in Y, \\ \text{we have } u_i(\sigma_i, \mu) \leq u_i(\mu). \end{array} \right\} \tag{10}$$

Let $u(\mathcal{P}) \equiv \{u(\mu)\}_{\mu \in \mathcal{P}}$ be the set of payoffs implemented by the mediator's strategy in $\mathcal{P}$.

Finally, they show that

$$E(1) = u(\mathcal{P}) \cap JD.$$

Comparing (9) and $E(1)$, we have the following two remarks.

First, our upper bound Q has the term $u(\mathcal{M})$ while $E(1)$ has the term $u(\mathcal{P})$. Comparing (5) and (10), the difference is that $\mathcal{M}$ requires that all the profitable deviations should be detected by some $r \in \operatorname{supp}(\mu)$ while $\mathcal{P}$ requires that all the profitable deviations should be detected by some $r \in A$.

This difference comes from the fact that, with discounting, players cannot take suboptimal actions with a positive probability while without discounting, players are allowed to take suboptimal actions infrequently. Hence, if a profitable deviation can be monitored only by a suboptimal action, then $u(\mathcal{P})$ is strictly larger than $u(\mathcal{M})$.

To see this difference, consider the following two-player game with the utility function given by the following utility matrix:

	L	R	L'
U	$2,2$	$0,3$	$0,2$
D	$0,0$	$0,0$	$0,1$

Player 1 has two possible signals $y_1 \in \{l, r\}$, which distinguishes whether player 2 takes "L or L'" or R only if player 1 takes D. That is,

$$\begin{aligned} q_1(l|a_1, a_2) &= 1 \text{ if } a_1 = D \text{ and } a_2 \in \{L, L'\}, \\ q_1(l|a_1, a_2) &= 0 \text{ otherwise.} \end{aligned}$$

Player 2 can perfectly observe player 1's action.

Note that $(U, L) \in \mathcal{P}$ since (i) player 1 takes a static best response and (ii) any profitable deviation of player 2 is detected when player 1 takes D. Further, since $(0, 1)$ is a Nash equilibrium payoff, $u(U, L) \in JD$. Therefore, we have $u(U, L) \in E(1)$.

On the other hand, there is no $\mu \in Q$ such that $u_1(\mu) + u_2(\mu) > 3$. In particular, we have $(U, L) \notin \mathcal{M}$. To see why, observe that, in order to have $u_1(\mu) + u_2(\mu) > 3$, we should have $\mu_2(L) > 0$. For each μ with $\mu_2(L) > 0$, unless $\mu_1|_L(D) > 0$, there exists σ_2 with $\Pr(m|r, \sigma_2) = \Pr(m|r)$ for each $r \in \text{supp}(\mu)$ and $u_2(\sigma_2, \mu) > u_2(\mu)$. Hence, we should have $\mu_1|_L(D) > 0$. Then, L' strictly dominates L in terms of payoffs and L and L' give the same distributions of player 1's signals. Hence, player 2 does not follow L. This is a contradiction.

Second, we share the same component JD in the characterization of Q and $E(1)$. Renault and Tomala (2004) obtain this restriction JD in the context of repeated games without discounting as follows. For each $\lambda \in \Lambda$, they consider a two-player repeated game between the mediator and a "deviator" with N states (State 1, 2, ..., N) unknown to the the mediator. State i corresponds to the event that player i is a unilateral deviator in the original game. The deviator maximizes $\sum_{i \in N} |\lambda_i| \, u_i(r, \sigma_i)$ while the mediator minimizes $\sum_{i \in N} |\lambda_i| \, u_i(r, \sigma_i)$, inferring which state she is in (that is, who is the deviator in the original game). Using Blackwell's approachability theorem, they show that the equilibrium payoff should be in

$$\bigcap_{\lambda \in \Lambda} \left\{ v \in \mathbb{R}^N : |\lambda| \cdot v \geq \min_{\tau \in \Delta(A)} \max_{\sigma \in SD(\text{supp}(\lambda))} \sum_{r \in A} \tau(r) \sum_{i \in N} |\lambda_i| \, u_i(r, \sigma_i) \right\}.$$

Therefore, we can interpret that Proposition 2 establishes that our characterization of $l(\lambda)$ using the movement of the continuation payoff $x(r, m)$ is equivalent to their characterization using approachability.

3.4 Punishment τ

The reason why we share the same term JD between the cases with and without discounting is that we do not impose any condition on τ. Intuitively, τ is the recommendation schedule which is used after the mediator statistically identifies some player's deviation. Since we consider the sequential equilibrium, we should impose that following τ is sequential rational (except for a deviator). On the other hand, in the literature of repeated games without discounting, it is common to consider Nash equilibrium, which does not require sequential rationality.

There are two reasons why requirement of sequential rationality is subtle. The first reason is that, with imperfect monitoring, a subset of players may not realize that the deviation has happened and the recommendation schedule is switched to τ.

For example, consider the following game: There are four players, 1, 2, 3, and 4. Player 4 is a monitor for players 1 and 2: Her utility is always equal to zero for each action profile, but she can perfectly monitor the actions by players 1 and 2. She does not have any information about player 3's action. That is, $u_4(a) = 0$ for all $a \in A$ and $y_4 \in A_1 \times A_2$ with $q_4(y_4|a) = 1$ if $y_4 = (a_1, a_2)$.

The other players do not have any information: $Y_i = \{\emptyset\}$ for all $i = 1, 2, 3$, and their utility function is independent of a_4 and given by the following utility matrix:

	l	r		l	r
U	$2,2,1$	$2,2,1$	U	$1,4,1$	$0,2,-1$
D	$2,2,1$	$2,2,1$	D	$2,0,-1$	$0,0,-1$
	L			R	

Player 1 picks U or D, player 2 picks l or r, and player 3 picks L or R.

Suppose that we required that τ should be incentive compatible. Since player 3 can guarantee the payoff of 1 by taking L forever, player 3 would have to receive the payoff of 1 with τ. Hence, τ would have to be (a_1, a_2, L) for some (a_1, a_2) or (U, l, R). If τ satisfies this condition, then player 1 could guarantee the payoff of 2 by taking D. Hence, if τ were incentive compatible, then player 1 could not receive the payoff of 1 in equilibrium.

Nonetheless, we will show that we can support (U, l, R) with probability one for each period in equilibrium. Hence, $(1, 4, 1) \in E_{\text{seq}}(\delta)$.

Suppose that the mediator recommends (U, l, R) unless player 4 has reported that $a_1 \neq U$ (player 1 deviated). Once $a_1 \neq U$ has been reported, then the mediator recommends (D, r, R) forever.

It is straightforward to see no player has an incentive to deviate from (U, l, R). In addition, since players 1 and 2 are best responding to each other in (D, r, R), we are left to verify player 3's incentive to deviate from (D, r, R).

When R is recommended, if player 3 knew that player 1 deviated and players 1 and 2 would take (D, r), then player 3 would have the incentive to deviate to L. However, since player 3 cannot monitor the actions of players 1 and 2, she never realizes player 1's deviation. Hence, she believes that they are on equilibrium path and the other players will take (U, l). Hence, taking R is optimal.

The key is that, once punishment $\tau = (D, r, R)$ becomes common knowledge, then such a punishment is not implementable, but the mediator has the way to trigger such a punishment without player 3 realizing it.

The second reason is related to the freedom of the belief system in sequential equilibrium with mediation. The freedom of the belief system depends on whether we see the mediator as a "machine" who cannot tremble or as a "player" who can. In the former interpretation, if player i is recommended an action that is not in the support of on-path recommendation, then player i needs to believe that some player has deviated. On the other hand, in the latter interpretation, player i can believe that this is a result of the mediator's tremble, rather than some player's deviation.

Which interpretation is more appropriate depends on how we interpret the role of the mediator. If we see the mediator as a theoretical device to incorporate a possibility that players may access to additional correlated signals unknown to the economist, then the machine interpretation is more appropriate since Nature does not tremble. On the other hand, as will be seen in Section 6, we see the mediator as a result of pre-play communication among players, then the player interpretation is more appropriate.

To see how this difference affects the condition for τ, let us consider the following game: There are two players, 1 and 2. Their utility function is given by the following utility matrix:

	L	R	L'	R'
U	$5,5$	$0,6$	$5,5$	$0,6$
M	$1,0$	$1,0$	$1,1$	$1,1$
D	$1,6$	$1,6$	$2,6$	$2,6$

Player 1 has two signals l and r with

$$\begin{aligned} q_1(l|a) &= 1 \text{ if } a_2 \in \{L, L'\}, \\ q_1(r|a) &= 1 \text{ if } a_2 \in \{R, R'\}. \end{aligned}$$

On the other hand, player 2 does not have any information: $Y_2 = \{\emptyset\}$.

This payoff matrix satisfied the following properties: First, to support the equilibrium in which the summation of the payoffs is no less than 10, we need to support (U, L) or (U, L').

Second, to implement (U, L) or (U, L'), we need to punish player 2 if she takes "R rather than L" or "R' rather than L'." As long as player 1 does not take M, player 2 can guarantee the payoff of 6 by playing R or R', the mediator needs to recommend M. Intuitively, τ should have (M, a_2) for some $a_2 \in A_2$ with a high probability.

In turn, suppose that player 2 knows that M is recommended to player 1 with a positive probability. Then, player 2 does not take L or R since (i) L (or R) is strictly dominated by L' (or R') in terms of player 2's payoff and (ii) the distribution of the signals is the same.

Given that player 2 does not take L or R, player 1 does not take M since (i) M is strictly dominated by D in terms of player 1's payoff and (ii) the distribution of the signals is the same.

The above discussion implies the following three claims. First, M cannot be in the support of the recommendation on the equilibrium path. Second, if we employ the machine interpretation of the mediator, then the mediator cannot incentivize player 1 to take M to punish player 2 off equilibrium path. To see why, suppose that there exists such an off-equilibrium path. Then, when player 1 is recommended M without her own deviation, player 1 knows that player 2 has deviated. Hence, player 1 knows that player 2 will not take L or R. Therefore, player 1 does not want to take M. In other words, there is no τ which is incentive compatible for player 1 and ensures that player 2's payoff is less than 6. Finally, given this, player 2's equilibrium payoff is no less than 6. Therefore, with the machine interpretation of the mediator, the summation of the equilibrium payoff is no more than 8.

On the other hand, if we go for the player interpretation, there exists an equilibrium in which the summation of the equilibrium payoff is arbitrarily close to 10. The mediator has 11 states R and (P, n) with $n = 1, ..., 10$. In each state, the mediator recommends actions and transits states as follows:

1. In state R, she recommends (U, L) and (U, R) with probability $1 - \varepsilon$ and ε with a small ε. If player 1 reports $m_1 = 1$ after $r_2 = L$, then go to state $(P, 1)$. Otherwise, stay at R.

2. In state (P, n) with $n = 1, ..., n$, she recommends (M, R) with probability one. The next state will be $(P, n + 1)$ if $n < 10$ and R if $n = 10$.

Let us specify the belief system of the players. At the point of receiving the recommendation in period t, player 1 always believes that the state is R. In particular, if $r_{1,t}$ is M or D, then she believes that the mediator perturbs her recommendation to (M, R) or (D, R) in state R.

On the other hand, player 2 believes that, unless she has deviated in periods $t-10, ..., t-1$ when the state was R, she believes that the state is R. If she deviated in period $t - n$ with $n \leq 10$ when the state was R, then she believes that the state is (P, n).

After observing $y_{1,t}$, player 1 keep her belief about the state at the beginning of period t. In particular, if player 1 believes that the state is R, recommended M or D (and so believes that $r_{2,t} = R$) and observes $y_{1,t} = l$, then she believes that the mediator recommended L or L' to player 2 as a result of trembles. Since player 2's signal is trivial ($Y_2 = \{\emptyset\}$), there is no belief update from $y_{2,t}$.

It is clear that there exists a sequence of the perturbations satisfying these criteria, so the constructed assessment is consistent.

Each player i follows recommendation except that player 2 who believes that the state is (P, n) takes L' (a static best response to M).

It is straightforward to check the sequential rationality: Since player 1 believes that the state is always R and $r_{1,t}$ is the static best response to $r_{2,t}$, it is optimal for her to follow the recommendation. When player 2 believes that they are in state R, if she is recommended L, then she is indifferent between L and L', and strictly prefers L and L' to R and R' since the latter induces $(P, 1)$. If she is recommended R, then since this is a static best response and the state transition is independent of player 2's action, it is optimal for player 2 to follow the recommendation. When player 2 believes that they are in state (P, n), since L' is a static best response and the state transition is independent of player 2's action, it is optimal for player 2 to take L'.

Finally, this equilibrium achieves the summation of the payoff of $(1 - \varepsilon) 10 + \varepsilon 6$.

The key to incentivize player 1 to take M to punish player 2 after player 2's deviation is that, even if player 1 is recommended M (the recommendation in state (P, n)) or observes $y_{1,t} = l$ (recall that player 2 takes L' in state (P, n) while player 1 after recommended M but before observing $y_{1,t}$ believes that player 2 is recommended R and follows R), player 1 keeps believing that these events are because of the mediator's trembles rather than player 2's deviation.

The above two reasons imply that a necessary and sufficient condition for the punishment strategy τ to be sequentially rational seems complicated and may depend on the fine details of the equilibrium strategy and belief system. We leave this question for the future research.

3.5 Proof of Theorem 1

We first prove that, for each $\lambda \in \Lambda$, we have $\sup_{v \in E_{\text{seq}}(\delta)} \lambda \cdot v \leq k(\lambda)$ for each $\delta \in [0,1)$.

For each $\delta < 1$, take a sequence of equilibrium strategies and payoffs $\{\sigma_m^n, (\sigma_i^n)_{i \in I}, v^n\}_{n=1}^{\infty}$ with $v^n \in E_{\text{seq}}(\delta)$ for each n such that the equilibrium payoff v^n converges to v with $\lambda \cdot v = \sup_{v' \in E_{\text{seq}}(\delta)} \lambda \cdot v'$; and let $w^n(r, m, \delta)$ be the continuation payoff from period 2 when r is recommended and m is reported.

The incentive compatibility in period 1 requires that, for each n, for each i and $s_i \in S_i$, we have

$$\begin{aligned} v_i^n &= \mathbb{E}\left[(1-\delta)u_i(r) + \delta w_i^n(r, y, \delta) | \mu^n\right] \\ &\geq \mathbb{E}\left[(1-\delta)u_i(a_i, r_{-i}) + \delta w_i^n(r, m_i, y_{-i}, \delta) | s_i, \mu^n\right]. \end{aligned} \tag{11}$$

Here, μ^n is the recommendation strategy for the mediator in period 1.

In addition, for each n, the expected value of $w^n(r, y, \delta)$ cannot be higher than the supremum of $\lambda \cdot v'$ with $v' \in E_{\text{seq}}(\delta)$:

$$\lambda \cdot \mathbb{E}\left[w^n(r, y, \delta) | \mu^n\right] \leq \lambda \cdot v = \sup_{v' \in E_{\text{seq}}(\delta)} \lambda \cdot v'. \tag{12}$$

To see why, suppose that (12) does not hold. Fix σ_m^n, the original equilibrium strategy of the mediator. Now the mediator constructs a new equilibrium as follows: First, the mediator draws "fictitious history" r, m according to

$$\Pr(r, m | \sigma_m^n) = \mu^n(r) q(m|r).$$

In the new equilibrium, in each period t after a history h_m^t, the mediator recommends actions as if the current history were (r, m, h_m^t) in the original equilibrium, that is, as if she recommended r in the first period, the report were m in the first period, and then her history from period 2 to period $t+1$ were h_m^t, in the original equilibrium.

If the original strategy is equilibrium, then so is the new one. If (12) is not true, then the ex ante value $\hat{v}^n$ of this new equilibrium satisfies

$$\lambda \cdot \hat{v}^n = \lambda \cdot \mathbb{E}\left[w^n(r, y, \delta)|\mu^n\right] > \sup_{v' \in E_{\text{seq}}(\delta)} \lambda \cdot v',$$

which is a contradiction.

Therefore, from (11) and (12), taking a subsequence if necessary, there exist μ and $w(r, y, \delta)$ such that for each i and $s_i \in S_i$,

$$\begin{aligned} v_i &= \mathbb{E}\left[(1-\delta)u_i(r) + \delta w_i(r, y, \delta)|\mu\right] \\ &\geq \mathbb{E}\left[(1-\delta)u_i(a_i, r_{-i}) + \delta w_i(r, m_i, y_{-i}, \delta)|s_i, \mu\right], \end{aligned}$$

and

$$\lambda \cdot \mathbb{E}\left[w(r, y, \delta)|\mu\right] \leq \lambda \cdot v.$$

When we define

$$x(r, m) = \frac{\delta}{1-\delta}(v - w(r, m, \delta)),$$

these μ and $x(r, m)$ satisfy [incentive compatibility], [promise keeping], and [self generation].

Hence, given Proposition 1, we have established $E_{\text{seq}}(\delta) \in u(\mathcal{M})$. Hence, we are left to show $E_{\text{seq}}(\delta) \in JD$.

Since Renault and Tomala (2004) show that $E(1) \subset JD$, it suffices to show that $E_{\text{seq}}(\delta) \subset E(1)$. With the mediator (more weakly, with public randomization device), the set of equilibrium payoffs is monotonic with respect to the discount factor. To see this, take any $\delta < 1$ and $\delta' \leq 1$ with $\delta' > \delta$ (with $\delta' = 1$, we refer to the repeated game without discounting). Suppose that, in the repeated game with δ', in each period t, the mediator renews the game with probability $\frac{\delta'-\delta}{\delta'}$. Then, each equilibrium in the repeated game with δ is an equilibrium in the repeated game with δ'. Therefore, the theorem is proven.

4 Observable Own Realized Payoffs

Since we have shown that $E_{\text{seq}}(\delta) \subseteq Q$, we now want to offer a sufficient condition for Q to be a tight upper bound in the limit: $\lim_{\delta \to 1} E_{\text{seq}}(\delta) \subseteq Q$. In particular, we will show that, if each player can observe her realized own payoffs, then we have $\lim_{\delta \to 1} E_{\text{seq}}(\delta) \subseteq Q$, as long as the dimension of Q is equal to the number of players.

As noted in Section 2.1, the assumption that player i observes her realized own payoffs is equivalent to the assumption that player i's realized payoffs are a deterministic function of a_i and y_i. Therefore, we have the following theorem:

Theorem 2 *If, for each i, there exists $\tilde{u}_i(a_i, y_i)$ such that $u_i(a) = \sum_{y_i \in Y_i} q_i(y_i|a)\tilde{u}_i(a_i, y_i)$ for each $a \in A$, then we have $\lim_{\delta \to 1} E_{\text{seq}}(\delta) \subseteq Q$ if $\dim(Q) = N$.*

Proof. Follows from Theorem 3 and Proposition 3 below. ■

To see why the observable-realized-payoff assumption is sufficient, fix $\lambda \in \Lambda$ arbitrarily and let $\mu \in \Delta(A)$ be the recommendation which achieves the score $k(\lambda)$.

We first slightly perturbs μ to $\mu^{\text{full}} = (1 - \eta)\,\mu + \eta \sum_{a \in A} \frac{1}{|A|} a$ with a small η so that μ^{full} has full support. With a sufficiently small η, the perturbed recommendation μ^{full} achieves $\lambda \cdot u(\mu^{\text{full}})$ close to $k(\lambda)$.

Second, consider the two potential channels through which player i's $\text{supp}(\mu^{\text{full}})$-undetectable deviation may affect the equilibrium construction. Note that now $\text{supp}(\mu^{\text{full}}) = A$.

The first channel is to affect players $-i$'s payoffs. Importantly, with the observable-realized-payoff assumption, we can ignore this channel. If player i's deviation is $\text{supp}(\mu^{\text{full}})$-undetectable, then it will not affect the other players' ex ante payoffs for each $a_{-i} \in A_{-i}$. Otherwise, the deviation would change the distribution of players $-i$'s realized payoffs, which means players $-i$ could detect player i's deviation after some $a_{-i} \in A_{-i}$. Hence, player i's $\text{supp}(\mu^{\text{full}})$-undetectable deviation is "harmless" to the other players in terms of payoffs.

The second channel is to affect the monitoring of players $-i$'s deviations. We can ensure that, if player i's deviation is not $\text{supp}(\mu^{\text{full}})$-detectable, then such a deviation does not affect the monitoring of players $-i$'s actions. Since player i's deviation does not change the distribution of the messages for each recommendation profile after each $r \in \text{supp}(\mu^{\text{full}}) = A$, player i's report of y_i after the deviation can monitor players $-i$'s deviations as well as player i's faithful strategy. This intuition is the same as the one observed in Rahman (2012).

Finally, consider how to discourage player i's $\text{supp}(\mu^{\text{full}})$-detectable deviations. Intuitively, we achieve this incentive by reducing the continuation payoff properly after statistically detecting player i's deviation. The remaining issue is whether there exists a sequentially rational punishment τ^* to achieve the low (punishment) payoff. To this end, we take τ that achieves the minimax value in (8). By the restriction of Q by $l(\lambda)$, this τ can achieve the punishment payoff. Using the observable-realized-payoff assumption, we can construct a sequentially rational and sufficiently severe punishment from τ.

4.1 Proof of Theorem 2

The formal proof of Theorem 2 consists of the following two steps. First, we provide a lower bound of the equilibrium payoff, denoted by $\underline{Q}$, such that $\underline{Q} \subseteq \lim_{\delta \to 1} E_{\text{seq}}(\delta)$. Second, we prove that the lower bound $\underline{Q}$ coincides with the upper bound Q if each player i observes her realized own payoffs.

4.1.1 A Lower Bound $\underline{Q}$

We characterize a lower bound $\underline{Q}$ of $\lim_{\delta \to 1} E_{\text{seq}}(\delta)$ by a linear programming that is similar, but different from the one to characterize the upper bound Q.

The lower bound $\underline{Q}$ is characterized by the intersection of half-spaces:

$$\begin{aligned} \underline{Q} &\equiv \bigcap_{\lambda \in \Lambda} \underline{H}(\lambda), \\ \underline{H}(\lambda) &\equiv \left\{ v \in \mathbb{R}^N : \lambda \cdot v \leq \min\left\{ \underline{k}(\lambda), \underline{l}(\lambda) \right\} \right\} \end{aligned}$$

for $\underline{k}(\lambda)$ and $\underline{l}(\lambda)$ to be determined.

Let us now define $\underline{k}(\lambda)$. We first define $\underline{k}(\mu, \lambda)$ for each $\mu \in \Delta(A)$, and then define $\underline{k}(\lambda) \equiv \sup_{\mu \in \Delta(A)} \underline{k}(\mu, \lambda)$. Specifically, $\underline{k}(\mu, \lambda)$ is defined by the following linear programming for given $\mu \in \Delta(A)$ and $\lambda \in \Lambda$:

$$\underline{k}(\mu, \lambda) = \max_{v \in \mathbb{R}^N, x: R \times Y \to \mathbb{R}^N} \lambda \cdot v$$

subject to the following constraints:

1. [effective full support] If $\text{supp}(\lambda) \geq 2$, then for each $i \in I$ and $r_i \in \text{supp}(\mu_i)$, the conditional distribution of r_{-i} satisfies the full support:

$$\text{aff}\left(\{q_i(Y_i|r_i, r_{-i})\}_{r_{-i} \in \text{supp}(\mu_{-i}|_{r_i})}\right) = \text{aff}\left(\{q_i(Y_i|r_i, r_{-i})\}_{r_{-i} \in A_{-i}}\right).$$

2. [incentive compatibility] For each $i \in I$, the faithful strategy is optimal when the change in the continuation payoff is given by $x(r, m)$: For all $s_i \in S_i$,

$$\mathbb{E}\left[u_i(r) + x_i(r, y)|\mu\right] \geq \mathbb{E}\left[u_i(a_i, r_{-i}) + x_i(r, m_i, y_{-i})|s_i, \mu\right].$$

3. [strict incentive compatibility] Moreover, if $\text{supp}(\lambda) \geq 2$, then for each player i with $\lambda_i \neq 0$ and $r_i \in \text{supp}(\mu_i)$, each deviation $s_i|_{r_i}$ that induces a different distribution of m given some r_{-i} is strictly discouraged: For each $s_i|_{r_i} \in S_i|_{r_i}$ with

$$\Pr(m|r_i, r_{-i}) \neq \Pr(m|s_i|_{r_i}, r_{-i}) \text{ for some } r_{-i} \in A_{-i} \text{ and } m \in Y,$$

we have

$$\mathbb{E}\left[u_i(r) + x_i(r, y)|\mu, r_i\right] > \mathbb{E}\left[u_i(a_i, r_{-i}) + x_i(r, m_i, y_{-i})|\mu, r_i, s_i|_{r_i}\right].$$

4. [promise keeping] v is the value of the repeated game: $v = u(\mu) + \mathbb{E}\left[x(r, y)|\mu\right]$.

5. [self generation] In expectation, in equilibrium, the score $\lambda \cdot x(r, y)$ is non positive: $\mathbb{E}\left[x(r, y)|\mu\right] \leq 0$.

6. [ex post self generation] If $\lambda = e_i$ or $-e_i$ for some $i \in I$, then after each $r \in \text{supp}(\mu)$ and $y \in Y$, $\lambda \cdot x(r, m)$ is non-positive: $\lambda \cdot x(r, m) \leq 0$ for each $r \in \text{supp}(\mu)$ and $y \in Y$.

We offer the discussion of the new conditions after we define $\underline{l}(\lambda)$ and state the theorem.

We now define $\underline{l}(\lambda)$. As will be seen, $\underline{l}(\lambda)$ represents the incentive compatibility in the following event: First, the mediator starts with μ to achieve high $\underline{k}(\mu, \lambda)$. If there are a lot of periods with $\lambda \cdot x(r, m) > 0$, then the mediator will switch to punish the players. By [ex post self generation], such a history does not happen if $\text{supp}(\lambda) = 1$. Hence, we define $\underline{l}(\lambda) = +\infty$ if $\text{supp}(\lambda) = 1$.

For $\text{supp}(\lambda) \geq 2$, we define $\underline{l}(\lambda) \equiv \sup l(\tau, \lambda)$ subject that $\tau \in \Delta(A)$ is strictly incentive compatible (see below for the formal definition). Only difference from $l(\lambda)$ for the upper bound is that we require τ to be strictly incentive compatible. Intuitively, if τ is strictly incentive compatible, then we can incentivize the players to take τ when the mediator switches to recommending τ.

We say that τ is strictly incentive compatible if the following condition is satisfied: There exists $x^\tau(r, m)$ such that, for each i and $r_i \in \text{supp}(\tau_i)$, each deviation $s_i|_{r_i}$ that induces a different distribution of m given some r_{-i} is strictly discouraged: For each $s_i|_{r_i} \in S_i|_{r_i}$ with

$$\Pr(m|r_i, r_{-i}) \neq \Pr(m|s_i|_{r_i}, r_{-i}) \text{ for some } r_{-i} \in A_{-i} \text{ and } m \in Y,$$

we have

$$\mathbb{E}\left[u_i(r) + x_i^\tau(r, y)|\tau, r_i\right] > \mathbb{E}\left[u_i(a_i, r_{-i}) + x_i^\tau(r, m_i, y_{-i})|\tau, r_i, s_i|_{r_i}\right].$$

Note that this is analogous to [strict incentive compatibility] for $\underline{k}(\mu, \lambda)$.

Given $\underline{k}(\lambda)$ and $\underline{l}(\lambda)$, we consider the half-space

$$\underline{H}(\lambda) \equiv \left\{v \in \mathbb{R}^N : \lambda \cdot v \leq \min\left\{\underline{k}(\lambda), \underline{l}(\lambda)\right\}\right\}.$$

We will show that the intersection of these half-spaces, denoted by $\underline{Q} \equiv \bigcap_{\lambda \in \Lambda} \underline{H}(\lambda)$, is a lower bound of the limit equilibrium payoff with full dimensionality condition:

Theorem 3 *If* $\dim(\underline{Q}) = N$, *then we have* $\underline{Q} \subseteq \lim_{\delta \to 1} E_{\text{seq}}(\delta)$.

Proof. See Section 7.3. ■

4.1.2 Observable Realized Payoffs

We can show that, if each player can observe her own realized payoffs, then the new conditions for $\underline{Q}$ are innocuous, and the upper bound Q and the lower bound $\underline{Q}$ coincide. Then, from Theorems 1 and 3, we have $\underline{Q} = \lim_{\delta \to 1} E_{\text{seq}}(\delta) = Q$.

Proposition 3 *With observable-realized-payoff assumption, we have* $\underline{Q} = Q$.

Proof. Follows from Lemmas 1 and 2 ■

Comparing the algorithms for Q and $\underline{Q}$, it is sufficient to show the following two lemmas:

Lemma 1 *For each* $\varepsilon > 0$, *we have* $\underline{k}(\lambda) \geq \min\{k(\lambda), l(\lambda)\} - \varepsilon$.

Proof. See Section 7.3.6. ■

Lemma 2 *For each* $\varepsilon > 0$, *we have* $\underline{l}(\lambda) \geq \min\{k(\lambda), l(\lambda)\} - \varepsilon$.

Proof. See Section 7.3.7. ■

To obtain the intuition for Lemmas 1 and 2, let us examine three new conditions, [effective full support], [strict incentive compatibility], and [ex post self generation] for $\underline{k}(\mu, \lambda)$, which we impose in addition to the conditions for $k(\mu, \lambda)$. In addition, we require that τ is strictly incentive compatible. We explain these new conditions are innocuous when each player can observe her realized own payoffs.

Suppose that we have μ establishing all the requirements for $k(\mu, \lambda)$. We can achieve [effective full support] if player i has the strict incentive with μ or each player can observe her realized own payoffs by perturbing μ as follows. For notational convenience, let $q_i \in \Delta(Y_i)$ be an extreme point of $\text{co}(\{q_i(Y_i|r_i, r_{-i})\}_{r_{-i} \in A_{-i}})$; and let $R_{-i}(q_i)$ be set of players $-i$'s actions which achieves q_i.

First, let us consider the case with strict incentives. Take some $\mu_{-i}(r_i, q_i) \in \Delta(R_{-i}(q_i))$ such that players $-i$ best respond to $\left(r_i, \mu_{-i}(r_i, q_i)\right)$ if we restrict players $-i$'s deviation to $R_{-i}(q_i)$. Consider a new recommendation $\tilde{\mu}$ by the mediator such that, first, the mediator draws r_i according to μ_i, and then, given r_i, recommends r_{-i} according to $(1-\eta)\,\mu_{-i}|_{r_i} + \eta\mu_{-i}(r_i, q_i)$ with a small η.

Now we have included q_i to $\{q_i(Y_i|r_i, r_{-i})\}_{r_{-i}\in \text{supp}(\tilde{\mu}_{-i}|_{r_i})}$. In addition, players $-i$ have the incentive to follow the recommendation by the following reasons: (i) since q_i is extreme, player i can statistically distinguish whether players $-i$ take some $r_{-i} \in R_{-i}(q_i)$ or not. Hence, the mediator can always discourage players $-i$ from taking $r_{-i} \notin R_{-i}(q_i)$ by reducing $x_{-i}(r, m)$. (ii) by definition, players $-i$ best respond to $\left(r_i, \mu_{-i}(r_i, q_i)\right)$ if we restrict players $-i$'s deviation to $R_{-i}(q_i)$.

Therefore, to achieve [effective full support], we are left to check if player i keeps the incentive to follow the recommendation. With a small η, player i's payoff of taking a_i against $\tilde{\mu}_{-i}|_{r_i}$ is close to that against $\mu_{-i}|_{r_i}$. Therefore, we can achieve [effective full support] if player i has the strict incentive with μ.

Second, consider the case in which each player can observe her realized own payoffs. As in Section 4, we perturb μ to $\mu^{\text{full}} = (1-\eta)\,\mu + \eta \sum_{a\in A} \frac{1}{|A|} a$. Note that [effective full support] is satisfied with μ^{full}. By letting each player n deviate to the best strategy among supp(μ^{full})-undetectable deviations, we keeps [effective full support] since, as mentioned earlier, if players $-i$ deviated to some strategy that does not take $R_{-i}(q_i)$, then player i could detect such a deviation since q_i is extreme.

Moreover, as seen in Section 4, player n's supp(μ^{full})-undetectable deviations do not affect other players' payoffs or monitoring, we can keep [incentive compatibility] and [self generation].

We require that the incentive is strict for each deviation that changes the distribution of the messages if supp(λ) ≥ 2. To see what situation this condition excludes, suppose that we have μ establishing all the requirements other than [strict incentive compatibility].

Suppose that there exist player i, $r_i \in \mu_i$, $s_i|_{r_i}$, $\tilde{s}_i|_{r_i}$, and $\alpha \in (0,1)$ such that $\alpha s_i|_{r_i} + (1-\alpha)\tilde{s}_i|_{r_i} = r_i$ and $\alpha u_i(s_i|_{r_i}, r_{-i}) + (1-\alpha)u_i(\tilde{s}_i|_{r_i}, r_{-i}) = u_i(r_i, r_{-i})$ for each $r_{-i} \in A_{-i}$. That is, mixing $s_i|_{r_i}$ and $\tilde{s}_i|_{r_i}$ with probabilities $\alpha \in (0,1)$ and $1-\alpha$ is equivalent to r_i for player i. Hence, we cannot give a strict incentive for player i to take r_i.

Suppose that $s_i|_{r_i}$ and $\tilde{s}_i|_{r_i}$ induce the extremum points of distribution of m for each r. To achieve [strict incentive compatibility], the mediator has to recommend $s_i(r_i)$ and $\tilde{s}_i(r_i)$ (actions taken by $s_i|_{r_i}$ and $\tilde{s}_i|_{r_i}$) with probabilities α and $1-\alpha$, instead of r_i. Player i's incentive is preserved. However, this may affect players $-i$'s payoffs and create deviation incentives for players $-i$. Hence, [strict incentive compatibility] is not always achievable.

However, if each player can observe her realized own payoffs, then the mediator can recommend $s_i(r_i)$ and $\tilde{s}_i(r_i)$ without affecting players $-i$'s utilities. To see why, since $\alpha s_i|_{r_i} + (1-\alpha)\tilde{s}_i|_{r_i}$ is a supp(μ)-undetectable deviation from r_i, (perturbing μ to μ^{full} if necessary,) we can conclude that player i's payoffs are not affected:

$$u_{-i}\left(\alpha s_i|_{r_i} + (1-\alpha)\tilde{s}_i|_{r_i}, a_{-i}\right) = u_{-i}\left(r_i, a_{-i}\right) \text{ for all } a_{-i} \in A_{-i}.$$

Hence, the mediator can recommend $s_i(r_i)$ and $\tilde{s}_i(r_i)$ with probabilities α and $1-\alpha$. Since $s_i|_{r_i}$ and $\tilde{s}_i|_{r_i}$ induce the extremum points of distribution of the messages to the mediator, we can find $x_i(r,m)$ which strictly incentivizes player i to follow the recommendation.

In general, if each player can observe her realized own payoffs, then the mediator recommends a convex combination of the actions which induces the extreme distributions of m instead of r_i so that we can satisfy [strict incentive compatibility]. The same logic explains that requiring [strict incentive compatibility] for $\underline{l}(\lambda)$ is innocuous.

Third, we require that, if $\lambda = e_i$ or $-e_i$ for some $i \in I$, then $\lambda \cdot x(r,m) \leq 0$ for each $r \in \text{supp}(\mu)$ and $y \in Y$. Following Fudenberg and Levine (1998), we can show that if we required this condition for each $\lambda \in \Lambda$, that is, if $\lambda \cdot x(r,m) \leq 0$ for each $r \in \text{supp}(\mu)$ and $y \in Y$ if $\text{supp}(\lambda) = 1$ for each $\lambda \in \Lambda$, then it would be straightforward to show Theorem 3. Hence, one can interpret that our contribution is to relax ex post self generation to

$\mathbb{E}[\lambda \cdot x(r,m)|\mu] \leq 0$ for $\lambda \in \Lambda$ with $\text{supp}(\lambda) \geq 2$.

If each player n can observe her realized own payoff $\tilde{u}_n$, then this condition is innocuous. Suppose that $\lambda = e_i$ (the explanation for $\lambda = -e_i$ is symmetric and so is omitted). Since each player n observes $\tilde{u}_n$, we define $x_i(r,m) = \frac{1}{K}\sum_{n \neq i}(\tilde{u}_n - U_n)$ and $x_j(r,m) = K\tilde{u}_i + \sum_{n \neq i,j} \tilde{u}_n$ for a sufficiently large K. Here, U_n is the maximum realization of player n's ex post payoffs. Given K, the mediator recommends $\bar{a} \in A$ that maximizes $u_i(a) + \frac{1}{K}\sum_{n \neq i} u_n(a)$.

Since $\tilde{u}_n$ does not affect $x_n(r,m)$, each player has the incentive to tell the truth about $\tilde{u}_n$. In addition, the players' incentive is aligned: Player i wants to maximize $u_i(a) + \mathbb{E}[x_i(r,m)|a] = u_i(a) + \frac{1}{K}\sum_{n \neq i} u_n(a)$+constant; and player $j \neq i$ wants to maximize $u_j(a) + \mathbb{E}[x_j(r,m)|a] = K\{u_i(a) + \frac{1}{K}\sum_{n \neq i} u_n(a)\}$+constant. Hence, all the players have incentive to follow $\bar{a}$. Finally, for sufficiently large K, $u_i(\bar{a}) + \mathbb{E}[x_i(r,m)|\bar{a}]$ is sufficiently close to $\max_{a \in A} u_i(a)$. Hence, $\lambda \cdot v$ can be arbitrarily close to $\max_{a \in A} \lambda \cdot u(a)$.

5 Unobservable Realized Payoffs

We have shown that, if each player can observe her own realized payoffs, then we have $\underline{Q} = \lim_{\delta \to 1} E_{\text{seq}}(\delta) = Q$. The following question naturally arises: If there is a player who cannot observe her own realized payoffs, then how large the gap is between Q (or $\underline{Q}$) and $\lim_{\delta \to 1} E_{\text{seq}}(\delta)$.

We now offer two examples. The first example shows that $\lim_{\delta \to 1} E_{\text{seq}}(\delta) \subsetneq Q$, and the second example shows that $\underline{Q} \subsetneq \lim_{\delta \to 1} E_{\text{seq}}(\delta)$.

5.1 An Example of $\lim_{\delta \to 1} E_{\text{seq}}(\delta) \subsetneq Q$

The following example proves that the upper bound is not always tight. Suppose that there are two players 1 and 2, and their utility functions are given by the following payoff matrix:

	$U(2)$	$D(2)$	$U_p(2)$	$U_m(2)$	$D_p(2)$	$D_m(2)$	$U_d(2)$	$D_d(2)$
$U(1)$	$6, 6$	$2, 8$	$-10, 7$	$-10, 5$	$-10, 9$	$-10, 7$	$-10, 6$	$-10, 8$
$D(1)$	$8, 2$	$0, 0$	$-10, 3$	$-10, 1$	$-10, 1$	$-10, -1$	$-10, 2$	$-10, 0$
$U_p(1)$	$7, -10$	$3, -10$	$-10, -10$	$-10, -10$	$-10, -10$	$-10, -10$	$-9, 0$	$-9, 0$
$U_m(1)$	$5, -10$	$1, -10$	$-10, -10$	$-10, -10$	$-10, -10$	$-10, -10$	$-10, 0$	$-10, 0$
$D_p(1)$	$9, -10$	$1, -10$	$-10, -10$	$-10, -10$	$-10, -10$	$-10, -10$	$-10, 0$	$-10, 0$
$D_m(1)$	$7, -10$	$-1, -10$	$-10, -10$	$-10, -10$	$-10, -10$	$-10, -10$	$-10, 0$	$-10, 0$
$U_d(1)$	$6, -10$	$2, -10$	$0, -10$	$0, -10$	$0, -10$	$0, -10$	$0, 0$	$0, 0$
$D_d(1)$	$8, -10$	$0, -10$	$0, -10$	$0, -10$	$0, -10$	$0, -10$	$0, 0$	$0, 0$

Each player has two signals, $y_i \in \{u, d\}$ and z, where y_i is private and z is public. The private signal y_i distinguishes whether player $j \neq i$ takes an action whose first character is U or D: $y_i = u$ with probability one if $a_j \in \{U(j), U_p(j), U_m(j), U_d(j)\}$; $y_i = D$ with probability one if $a_j \in \{D(j), D_p(j), D_m(j), D_d(j)\}$.

On the other hand, the distribution of z is determined as follows:

1. g with probability $\frac{1}{2}$ and b with probability $\frac{1}{2}$ if $a_i \neq U_p(i), D_p(i), U_m(i), D_m(i)$ for any i

2. g with probability 0 and b with probability 1 if $a_i = U_p(i), D_p(i)$ for some i

3. g with probability 1 and b with probability 0 if $a_i = U_m(i), D_m(i)$ for any i

Note that $U_d(i)$ and $D_d(i)$ weakly dominates $U(i)$ and $D(i)$ in terms of payoffs and information. This dominance is strict unless the opponent takes only $U(j)$ or $D(j)$.

In addition, when player i deviates from $U(i)$ to $U_p(i)$, player i can improves her instantaneous utility by one regardless of a_j while only difference in information is that the probability that g happens will be reduced by $\frac{1}{2}$.

For notational convenience, let G denote a static game with payoff matrix

	$U(2)$	$D(2)$
$U(1)$	$6,6$	$2,8$
$D(1)$	$8,2$	$0,0$

Let $U(G)$ be the feasible payoff set in G, and $A(G)$ be the action profile set of G.

We can show that although $v = (6,6) \in Q$, there does not exist a sequential equilibrium to support $(\frac{16}{3}, \frac{16}{3})$ for any $\delta < 1$.

First, let us prove that $v = (6,6) \in Q$. For each λ, for each $a \in A(G)$, we have $k(\lambda) \geq \lambda \cdot u(a)$. To see why, consider $\mu = a \in A(G)$ and the following $x(r,m)$: For each i,

1. If $m_j = (y_j, z)$ with $y_j = U$ (or D) although $a_i = D(i)$ (or $U(i)$), then $x_i(r,m) = -8$.
2. Otherwise,
   (a) If $r_i = D(i)$, then $x_i(r,m) = 0$.
   (b) If $r_i = U(i)$, for $m_j = (y_j, z)$, we have $x_i(r,m) = 2$ with $z = g$ and $x_i(r,m) = -2$ with $z = b$ to discourage the deviations to $U_p(i)$ or $U_m(i)$.

Moreover, since $D_d(1), D_d(2)$ is a static Nash equilibrium, $l(\lambda) \geq \lambda \cdot u(D_d(1), D_d(2))$.

Therefore, we have $(6,6) \in Q$.

Now let us prove that there does not exist an equilibrium to support $v = (6,6)$. Suppose that, for each $\varepsilon > 0$, there exists $\bar{\delta}$ such that for $\delta > \bar{\delta}$, we have $\lambda \cdot \bar{v} \geq \lambda \cdot (6-\varepsilon, 6-\varepsilon)$ with $\lambda = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ and $\bar{v} \in \arg\max_{v \in E(\delta)} \lambda \cdot v$.

The following lemma is useful:

Lemma 3 *If there exists a history of the mediator after which the summation of the two players' instantaneous utilities is more than zero, then it is common knowledge among the players that the mediator recommends only $U(i)$ and $D(i)$ to each player i.*

Proof. To support the instantaneous utilities summing up to a positive number, there should exist player i to whom $U(i)$ or $D(i)$ is recommended.

Notice that $U_d(i)$ and $D_d(i)$ weakly dominates $U(i)$ and $D(i)$ in terms of payoffs and information, and that the domination is strict unless player j takes only $U(j)$ and $D(j)$. Hence, whenever $U(i)$ and $D(i)$ is recommended, player i knows that $U(j)$ and $D(j)$ is recommended to player j. Then, player j knows that $U(i)$ and $D(i)$ is recommended to player i by the symmetric reason.

Inductively, we can conclude that it is common knowledge among the players that the mediator recommends only $U(i)$ and $D(i)$ to each player i. ■

Given this lemma, without loss of generality, we can assume that in period 1, the mediator recommends only $U(i)$ and $D(i)$ to each player i. Then, we have the following three conditions: For $\bar{v} \in \arg\max_{v \in E(\delta)} \lambda \cdot v$,

1. The equilibrium payoff is decomposed as

$$\bar{v} = \mathbb{E}\left[(1-\delta)u(r) + \delta w(r,m) \mid \mu\right]. \tag{13}$$

2. As noted above, it is common knowledge that $r_i \in \{U(i), D(i)\}$ for each player i. By observing $y_i \in \{U, D\}$, player i knows r_j. Hence, on equilibrium path, the continuation play is common knowledge. Therefore, we need to have

$$\lambda \cdot w(r,m) \leq \lambda \cdot \bar{v} \tag{14}$$

for all r and m that happens in equilibrium.

On the other hand, the incentive compatibility for player i requires the following: For $r_i = U(i)$, we need to have

$$\begin{aligned} & \frac{\delta}{1-\delta}\left\{\mathbb{E}\left[w_i(r,m) \mid U(i), \mu_j|_{U(i)}\right] - \mathbb{E}\left[w_i(r,m) \mid U(i), \mu_j|_{U(i)}, a_i = U_p(i)\right]\right\} \\ \geq\ & u_i(U_p(i), \mu_j|_{U(i)}) - u_i(U(i), \mu_j|_{U(i)}). \end{aligned}$$

Since (i) $u_i(U_p(i), \mu_j|_{U(i)}) - u_i(U(i), \mu_j|_{U(i)}) = 1$ regardless of $\mu_j|_{U(i)}$ and (ii) only difference in the distribution of m is that the probability that $z = g$ happens will be reduced by $\frac{1}{2}$ after $U_p(i)$, we need to have

$$\frac{\delta}{1-\delta} \left\{\mathbb{E}\left[w_i(r,m) \mid U(i), \mu_j|_{U(i)}, z = g\right] - \mathbb{E}\left[w_i(r,m) \mid U(i), \mu_j|_{U(i)}, z = b\right]\right\} \geq 2.$$

At the same time, we also need to have

$$\begin{aligned} & \frac{\delta}{1-\delta} \left\{\mathbb{E}\left[w_i(r,m) \mid U(i), \mu_j|_{U(i)}\right] - \mathbb{E}\left[w_i(r,m) \mid U(i), \mu_j|_{U(i)}, a_i = U_m(i)\right]\right\} \\ \geq\ & u_i(U_m(i), \mu_j|_{U(i)}) - u_i(U(i), \mu_j|_{U(i)}). \end{aligned}$$

Since (i) $u_i(U_m(i), \mu_j|_{U(i)}) - u_i(U(i), \mu_j|_{U(i)}) = -1$ regardless of $\mu_j|_{U(i)}$ and (ii) only difference in the distribution of m is that the probability that $z = g$ happens will be increased by $\frac{1}{2}$ after $U_m(i)$, we need to have

$$\frac{\delta}{1-\delta} \left\{\mathbb{E}\left[w_i(r,m) \mid U(i), \mu_j|_{U(i)}, z = g\right] - \mathbb{E}\left[w_i(r,m) \mid U(i), \mu_j|_{U(i)}, z = b\right]\right\} \leq 2.$$

Hence, we have

$$\frac{\delta}{1-\delta} \left\{\mathbb{E}\left[w_i(r,m) \mid U(i), \mu_j|_{U(i)}, z = g\right] - \mathbb{E}\left[w_i(r,m) \mid U(i), \mu_j|_{U(i)}, z = b\right]\right\} = 2. \qquad (15)$$

Applying the same argument with $U(i)$, $U_p(i)$, and $U_m(i)$ replaced with $D(i)$, $D_p(i)$, and $D_m(i)$, we have

$$\frac{\delta}{1-\delta} \left\{\mathbb{E}\left[w_i(r,m) \mid D(i), \mu_j|_{U(i)}, z = g\right] - \mathbb{E}\left[w_i(r,m) \mid D(i), \mu_j|_{U(i)}, z = b\right]\right\} = 2. \qquad (16)$$

Since (15) and (16) hold for each player i, we should have

$$\frac{\delta}{1-\delta} \left\{\mathbb{E}\left[w_1(r,m) + w_2(r,m) \mid \mu, z = g\right] - \mathbb{E}\left[w_1(r,m) + w_2(r,m) \mid \mu, z = b\right]\right\} = 4.$$

Together with (14), we have

$$\frac{\delta}{1-\delta}\mathbb{E}\left[\lambda \cdot w(r,m) \mid \mu\right] \leq \frac{\delta}{1-\delta}\lambda \cdot \bar{v} - \sqrt{2}.$$

By (13) and (14), we have

$$\begin{aligned}\lambda \cdot \bar{v} &= \mathbb{E}\left[\lambda \cdot u(r) + \frac{\delta}{1-\delta}\lambda \cdot (w(r,m) - \bar{v}) \mid \mu\right] \\ &\leq \lambda \cdot (6,6) - \sqrt{2},\end{aligned}$$

which is a contradiction.

5.2 An Example of $\underline{Q} \subsetneq \lim_{\delta \to 1} E_{\text{seq}}(\delta)$

The next example shows that the lower bound is not always tight either.

There are five players. Player 5 has only one action and her utility does not depend on action profiles. That is, player 5's only role is to provide the information of signals to the mediator in order to monitor the other players.

Player $i \in \{3,4\}$ has three actions, $U(i)$, $D(i)$, and $U_m(i)$. When players 3 and 4 take $U(3)$ and $U(4)$, utilities of players 1 and 2 are given by the following payoff matrix:

	$U(2)$	$M(2)$	$D(2)$	$U_p(2)$	$U_m(2)$	$P(2)$
$U(1)$	$2,2$	$0,0$	$0,0$	$-10,\frac{3}{2}$	$-10,\frac{13}{6}$	$-10,2$
$M(1)$	$0,0$	$2,2$	$0,0$	$0,0$	$0,0$	$-10,2$
$D(1)$	$0,0$	$0,0$	$1,1$	$0,0$	$0,0$	$-10,1$
$U_p(1)$	$\frac{3}{2},-10$	$0,0$	$0,0$	$0,0$	$0,0$	$-10,\frac{3}{2}+2$
$U_m(1)$	$\frac{13}{6},-10$	$0,0$	$0,0$	$0,0$	$0,0$	$-10,\frac{13}{6}+2$
$P(1)$	$2,-10$	$2,-10$	$1,-10$	$\frac{3}{2}+2,-10$	$\frac{13}{6}+2,-10$	$0,0$

Note that, unless $U(1),U(2)$, $M(1),M(2)$, or $D(1),D(2)$ are the only recommendation pairs, taking $P(i)$ strictly dominates $U(i)$, $M(i)$, or $D(i)$.

When player 3 or 4 does not take $U(3)$ or $U(4)$, player $i \in \{1,2\}$'s payoffs are decreased by 10 unless she takes $P(i)$: For each $i \in \{1,2\}$ and $a \in A$, we have

$$
\begin{aligned}
u_i(a) &= u_i(a_1, a_2, U(3), U(4)) - 10 \text{ if } a_i \neq P(i) \text{ and } (a_3, a_4) \neq (U(3), U(4)), \\
u_i(a) &= u_i(a_1, a_2, U(3), U(4)) \text{ if } a_i = P(i) \text{ or } (a_3, a_4) \neq (U(3), U(4)).
\end{aligned}
$$

The payoffs of players 3 and 4 are determined as follows: It depends on (a_1, a_2) which of $U(3)$ and $D(3)$ is better:

$$
\begin{aligned}
u_3(a_{-3}, U(3)) &= -1 \text{ and } u_3(a_{-3}, D(3)) = 3 \\
\text{if } (a_1, a_2) &= (U(1), U(2)), (M(1), U(2)), (M(1), U(2)), (M(1), M(2)), \\
u_3(a_{-3}, U(3)) &= 1 \text{ and } u_3(a_{-3}, D(3)) = -3 \text{ otherwise.}
\end{aligned}
$$

$$
\begin{aligned}
u_4(a_{-4}, U(4)) &= 1 \text{ and } u_4(a_{-4}, D(4)) = -1 \\
\text{if } (a_1, a_2) &= (M(1), U(2)), (M(1), M(2)), (M(1), D(2)), \\
u_4(a_{-4}, U(4)) &= -1 \text{ and } u_4(a_{-4}, D(4)) = 1 \text{ otherwise.}
\end{aligned}
$$

On the other hand, for each player $i \in \{3,4\}$, $U_m(i)$ is strictly better than $U(i)$ unless each player $n \in \{1,2\}$ takes $a_n \in \{U(n), M(n), D(n)\}$ and player $n \in \{3,4\} \setminus \{i\}$ takes $a_n = U(n)$:

$$
\begin{aligned}
u_i(a_{-i}, U_m(i)) &= u_i(a_{-i}, U(i)) \\
\text{if } a_n &\in \{U(n), M(n), D(n)\} \text{ for each } n \in \{1,2\} \text{ and} \\
a_n &= U(n) \text{ for } n \in \{3,4\} \setminus \{i\}. \\
u_i(a_{-i}, U_m(i)) &= u_i(a_{-i}, U(i)) + 1 \text{ otherwise.}
\end{aligned}
$$

The signal structure is as follows: Players 1, 2, 3, and 4 do not observe any informative signals: $Y_n = \{\emptyset\}$ for $n \in \{1,2,3,4\}$. Player 5 observes a signal that is informative whether player $i \in \{1,2\}$ takes $U_m(i)$, $U_p(i)$, or something else: $y_5 \in \{g, b\}$ and

$$q_5(g|a) = \begin{cases} 1 & \text{if } a_1 = U_m(1) \text{ or } a_2 = U_m(2) \\ 0 & \text{if } a_1 = U_p(1) \text{ or } a_2 = U_p(2) \\ \frac{1}{4} & \text{otherwise} \end{cases}$$

In this game, we first show that $\underline{Q}$ does not contain v with $v_1 + v_2 \geq 3$:

Lemma 4 *In the above example, for each v with $v_1 + v_2 \geq 3$, the effective full support is not satisfied.*

Proof. Suppose we want to support the equilibrium in which the summation of the payoffs of players 1 and 2 is no less than 3: $v_1 + v_2 \geq 3$. Then, μ needs to put some probability on r with $r \in \{U(i), M(i), D(i)\}$ for each $i \in \{1,2\}$ and $r_i = U(i)$ for each $i \in \{3,4\}$. For a simple notation, let E denote the event that $r_i \in \{U(i), M(i), D(i)\}$ for each $i \in \{1,2\}$ and $r_i = U(i)$ for each $i \in \{3,4\}$.

Observe that when E happens, E becomes common knowledge. To see why, consider player $i \in \{1,2\}$'s incentive. Unless $r_j \in \{U(j), M(j), D(j)\}$ for $j = \{1,2\} \setminus \{i\}$ and $r_j = U(j)$ for each $j \in \{3,4\}$, $P(i)$ strictly dominates each $a_i \in \{U(i), M(i), D(i)\}$ in terms of payoffs and the distribution of the signals is the same. Hence, unless player i assigns the belief of one to E, player i has the incentive to deviate.

Now consider player $i \in \{3,4\}$'s incentive. Unless $r_j \in \{U(j), M(j), D(j)\}$ for $j = \{1,2\}$ and $r_j \in U(j)$ for each $j \in \{3,4\} \setminus \{i\}$, $U_m(i)$ strictly dominate $U(i)$ in terms of payoffs and the distribution of the signals is the same. Hence, unless player i assigns the belief of one to E, player i has the incentive to deviate.

Therefore, E becomes common knowledge. This contradicts to the effective full support.

■

Moreover, we can also show that the strict incentive compatibility is not satisfied either:

Lemma 5 *In the above example, for each v with $v_1+v_2 \geq 3$, the strict incentive compatibility is not satisfied.*

Proof. From the above discussion, we focus on the mediator's recommendation μ given E.

For such μ to achieve $v_1 + v_2 \geq 3$, player $i \in \{1,2\}$ should have an incentive to take $U(i)$, $M(i)$, or $D(i)$. Hence, $(U(1), U(2))$, $(M(1), M(2))$, or $(D(1), D(2))$ should be the only recommendation pairs with a positive probability, since otherwise $P(i)$ strictly dominates $U(i)$, $M(i)$, or $D(i)$.

On the other hand, player $i \in \{3,4\}$ should have an incentive to take $U(3)$. Since their actions are not monitored, $U(i)$ should be a static best response to $\mu_{-i}|_{U(i)}$ for $i \in \{3,4\}$. Hence, from player 3's incentive, we should have $(a_1, a_2) = (M(1), M(2))$ with probability no less than $\frac{1}{2}$; and from player 4's incentive, we should have $(a_1, a_2) = (M(1), M(2))$ with probability no more than $\frac{1}{4}$.

In total, μ should satisfy

$$\begin{aligned} \Pr((D(1), D(2), U(3), U(4))\,|\mu) &= p_D \geq \tfrac{1}{2}, \\ \Pr((M(1), M(2), U(3), U(4))\,|\mu) &= p_M \leq \tfrac{1}{4}, \\ \Pr((U(1), U(2), U(3), U(4))\,|\mu) &= 1 - p_D - p_M. \end{aligned} \tag{17}$$

When $r = (D(1), D(2), U(3), U(4))$ or $(M(1), M(2), U(3), U(4))$ is recommended, since each player i takes a static best response, no intertemporal incentive is necessary. On the other hand, when $r = (U(1), U(2), U(3), U(4))$ is recommended, player $i \in \{1,2\}$ can deviate from $U(i)$ to $U_m(i)$. Since the instantaneous utility gain of this deviation is $\frac{1}{6}$ while the probability of $y_5 = g$ is decreased by $\frac{1}{4}$,

$$x_i(r, g) - x_i(r, b) \geq \frac{2}{3} \text{ for each } i \in \{1,2\}.$$

If $x(r, g) - x(r, b) > \frac{2}{3}$, then it would be profitable for player i to deviate from $U(i)$ to $U_p(i)$: Since the instantaneous utility loss of this deviation is $\frac{1}{2}$ while the probability of $y_5 = g$ is increased by $\frac{3}{4}$, the total gain would be greater than $-\frac{1}{2} + \frac{3}{4} \times \frac{2}{3} > 0$. Hence, we have

$$x(r, g) - x(r, b) = \frac{2}{3} \text{ for each } i \in \{1,2\}. \tag{18}$$

Hence, incentive to follow μ is weak after recommended $U(1)$ or $U(2)$. ■

Lemmas 4 and 5 imply the following: Suppose that the mediator recommends a stationary recommendation μ with (17) for sufficiently long time (that is, T_δ with δ^{T_δ} is less than one) to maximize $v_1 + v_2$, that is, the Pareto weight is $\lambda = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0\right)$. (18) implies that players 1 and 2's continuation payoff should be increased exactly by $\frac{2}{3}$ when $y_5 = g$ happens rather than $y_5 = b$. Hence, after observing g repeatedly, the feasibility constraint will prevent us from keeping the continuation payoff increasing.

However, we can achieve the equilibrium payoff v with $v_1 + v_2 \geq 3$ by a non-stationary recommendation schedule.

Lemma 6 *For each $\varepsilon > 0$, for sufficiently large δ, there exists $v \in E_{\text{seq}}(\delta)$ with $v_1 + v_2 \geq 3 - \varepsilon$.*

Proof. See Section 7.4. ■

To see the intuition, consider the following (non-stationary) recommendation schedule: In period 1, the mediator recommends r according to μ_0 with

$$\begin{aligned}
\Pr((D(1), D(2), U(3), U(4)) \,|\mu_0) &= \frac{1}{2}, \\
\Pr((M(1), M(2), U(3), U(4)) \,|\mu_0) &= \frac{1}{4}, \\
\Pr((U(1), U(2), U(3), U(4)) \,|\mu_0) &= \frac{1}{4}.
\end{aligned}$$

In each period $t \geq 2$, if player 5 reported the signal of g last period, that is, if $m_{5,t-1} = g$, then $r = (M(1), M(2), U(3), U(4))$ with probability one. If player 5 reported the signal of b last period, that is, if $m_{5,t-1} = b$, then $r = (D(1), D(2), U(3), U(4))$ with probability $\frac{2}{3}$ and $r = (U(1), U(2), U(3), U(4)) = \frac{1}{3}$.

On equilibrium path, in each period t, each player $i \in \{1, 2, 3, 4\}$ believes that $m_{5,t-1} = g$ with probability $\frac{1}{4}$ and $m_{5,t-1} = b$ with probability $\frac{3}{4}$. Hence, each player $i \in \{1, 2, 3, 4\}$

believes that

$$
\begin{aligned}
r_t &= (D(1), D(2), U(3), U(4)) \text{ with probability } \Pr(m_{5,t-1} = b) \times \frac{2}{3} = \frac{1}{2}, \\
r_t &= (M(1), M(2), U(3), U(4)) \text{ with probability } \Pr(m_{5,t-1} = g) = \frac{1}{4}, \\
r_t &= (U(1), U(2), U(3), U(4)) \text{ with probability } \Pr(m_{5,t-1} = b) \times \frac{2}{3} = \frac{1}{3}.
\end{aligned} \tag{19}
$$

Hence, player $i \in \{3,4\}$ does not have an incentive to deviate.

Consider player $i \in \{1,2\}$'s incentive in period t. If $M(i)$ or $D(i)$ is recommended, then $M(j)$ or $D(j)$ is recommended to the opponent $j \in \{1,2\} \setminus \{i\}$, respectively. Since this is a static best response and the distribution of y_5 is independent of player i's action, player i has the incentive to follow the recommendation. If $U(i)$ is recommended, then $U(j)$ is recommended to the opponent $j \in \{1,2\} \setminus \{i\}$. There are following four possibilities:

1. If player i follows the recommendation, then the payoff in period t is 2. Since $\Pr(m_{5,t} = g) = \frac{1}{4}$, the expected payoff in period $t+1$ is $\frac{3}{2}$, and that in period $\tau \geq t+2$ is $\frac{3}{2}$.

2. If player i takes $U_p(i)$, then the payoff in period t is $\frac{3}{2}$ and, that in period $t+1$ is 2 since $y_{5,t} = g$ with probability one. Assuming that player i follows the recommendation from the next period, the expected payoff in period $\tau \geq t+2$ is $\frac{3}{2}$.

3. If player i takes $U_m(i)$, then the payoff in period t is $\frac{13}{6}$ and, that in period $t+1$ is $\frac{1}{3} \times 2 + \frac{2}{3} \times 1 = \frac{4}{3}$ since $y_{5,t} = b$ with probability one. Again, assuming that player i follows the recommendation from the next period, the expected payoff in period $\tau \geq t+2$ is $\frac{3}{2}$.

4. Player i does not take $a_i \notin \{U(i), U_p(i), U_m(i)\}$ since $U(i)$ is weakly better in terms of payoffs and the distribution of y_5 does not change.

Hence, if we neglect the discounting, the change in the instantaneous utility in period t cancels out with that in period $t+1$. Hence, player i is indifferent between $U(i)$, $U_p(i)$, and $U_m(i)$. Section 7.4 offers the formal proof.

Note that players 3 and 4 need to believe that μ_t satisfies (19) on equilibrium path to achieve $v_1 + v_2 \geq 3$. At the same time, the continuation payoff of players 1 and 2 need to be increased after $y_5 = g$ happens since otherwise they have an incentive to deviate to $U_m(i)$. The problem is that, for each μ_t satisfying (19), $v_1 + v_2$ cannot exceed 3.

The above example shows that while we keep players 3 and 4 believing μ_t satisfying (19), we can change the recommendation in a history-dependent way so that players 1 and 2 believe that, after increasing the probability of $y_5 = g$ by deviating $U_m(i)$, the continuation payoff (w_1, w_2) satisfies $w_1 + w_2 > 3$.

This example might suggest that obtaining the general characterization for the case with non-observable realized payoffs is hard, since the availability of such history-dependent recommendation strategies depends on the fine details of the belief structure.

6 Dispensability of the Mediator

When we see the mediator as a theoretical exercise to show a clean and general result, we may want to replace the mediator with pre-play communication among players.

As will be seen, if we try to replace the mediator with communication among players, then such a communication may create an incentive for a player to tell a lie in order to change the distribution of the other players' actions or obtain more precise information about the other players' actions. Since we have to detect a liar to effectively punish the liar, the result depends on the number of players.

Using pre-play communication protocol introduced by Gerardi (2004), we can show that, with at least five players, we can replace the mediator with pre-play communication among players.

Since Gerardi (2004) restricts attention to equilibria in which recommendation profile r is drawn from probabilities that are rational numbers, so do we: Let $\underline{Q}^{\text{rational}}$ be the same set as $\underline{Q}$ except that we require that $\mu(r)$ and $\tau(r)$ in definition of $\underline{Q}$ should be rational numbers for each $r \in A$.

Without mediation but with pre-play communication, in each period, the stage game consists of the following two phases:

1. [pre-play phase] We employ the pre-play communication as Gerardi (2004). Specifically, the players exchange messages multiple times. At each point, for each player $i \in I$ and each subset of players $J \subset I$, player i can send a message which is shared only among the players in the subset J.

2. [action phase] Based on the history in the pre-play phase, each player takes an action $a_{i,t}$.

Let $E_{\text{seq}}^{\text{without}}(\delta)$ be the set of sequential equilibrium payoffs in a repeated game of the stage game above. We have the following result:

Theorem 4 *If* $\dim(\underline{Q}^{\text{rational}}) = N \geq 5$, *then* $\underline{Q}^{\text{rational}} \subseteq \lim_{\delta \to 1} E_{\text{seq}}^{\text{without}}(\delta)$.

Proof. In general, Gerardi (2004) considers a one-shot game with incomplete information. Let $\Theta_i \ni \theta_i$ be player i's type space. Suppose we have constructed the equilibrium with the mediator such that each player tells the truth about θ_i to the mediator; given reported type profile θ; the mediator draws the recommendation $r \in A$ according to $\mu(\theta) \in \Delta(A)$; and each player follows the recommendation.

Given such an equilibrium with the mediator, we can construct a pre-play communication protocol without the mediator such that (i) each player does not have an incentive to tell a lie; (ii) for each realization of θ, the recommendation profile r is drawn from $\mu(\theta)$; and (iii) each player does not learn anything about θ_{-i} and r_{-i}.

Using this mechanism, we can replace the mediator in our equilibrium with pre-play communication. Arbitrarily fix an equilibrium with the mediator, $(\sigma_m, (\sigma_i)_{i \in I})$.

We now replace the mediator with pre-play communication. We see player i's history h_i^t at the beginning of period t as player i's "type." In each period t, we construct the pre-play communication protocol such that (i) each player does not have an incentive to tell a lie; (ii) for each realization of h^t, the recommendation profile r is drawn from $\sigma^m(h^t)$; and (iii) each player does not learn anything about h_{-i}^t and r_{-i}. Since h_i^t contains $(r_{i,\tau}, m_{i,\tau})_{\tau=1}^{t-1}$, we can calculate the mediator's recommendation from h^t, which is denoted by $\sigma^m(h^t)$ here. ■

7 Appendix

7.1 Proof of Proposition 1

Without loss of generality, we can assume that $\mathbb{E}[x(r,y)|\mu] = 0$ and $v = u(\mu)$. Hence, we can write $k(\mu,\lambda)$ as follows:

$$k(\mu,\lambda) \equiv \max_{\mu\in\Delta(A)} \lambda \cdot u(\mu)$$

such that there exists $x(r,y)$ such that, for each $i \in I$ and $s_i \in S_i$, we have

$$\mathbb{E}[x_i(r, m_i, y_{-i})|s_i,\mu] - \mathbb{E}[x_i(r,y)|s_i^*,\mu] \leq \mathbb{E}[u_i(r)|\mu] - \mathbb{E}[u_i(a_i, r_{-i})|s_i,\mu].$$

Following Rahman (2012), we can apply Theorem 22.1 of Rockafellar (1960) to obtain the dual: The existence of such $x(r,y)$ is equivalent to the non-existence of $(\sigma_i(s_i))_{s_i,i}$ with $\rho_i(s_i) \geq 0$ for all i and s_i such that

$$\sum_{s_i} \sigma_i(s_i)\{\Pr(r,m|s_i,\mu) - \Pr(r,m|\mu)\} = 0 \text{ for each } r \in A \text{ and } m \in Y, \quad (20)$$

$$\sum_{s_i} \sigma_i(s_i)(\mathbb{E}[u_i(r)|\mu] - \mathbb{E}[u_i(a, r_{-i})|s_i,\mu]) < 0. \quad (21)$$

If $\sum_{s_i} \sigma_i(s_i) = 0$, then since $\sigma_i(s_i) \geq 0$ for all i and s_i, we have $\sigma_i(s_i) = 0$ for all i and s_i. Hence, (21) is not satisfied. Therefore, we focus on the case with $\sum_{s_i} \sigma_i(s_i) > 0$. Without loss, we can assume that $\sum_{s_i} \sigma_i(s_i) = 1$ since otherwise we can dividing both sides by $\sum_{s_i} \sigma_i(s_i)$. Seeing $\sigma_i(s_i)$ as a probability that player i takes s_i, Conditions (20) and (21) are equivalent to the existence of player i's unilateral deviation $\sigma_i \in \Sigma_i$ such that

$$\begin{aligned} \Pr(r, m|\sigma_i, \mu) &= \Pr(r, m|\mu) \text{ for each } r \in A \text{ and } m \in Y, \\ \mathbb{E}\left[u_i(a, r_{-i})|\sigma_i, \mu\right] &> \mathbb{E}\left[u_i(r)|\mu\right]. \end{aligned}$$

Equivalently, there exists $\sigma_i \in \Sigma_i$ such that

$$\begin{aligned} \Pr(m|\sigma_i, r) &= \Pr(m|r) \text{ for each } r \in \operatorname{supp}(\mu) \text{ and } m \in Y, \\ \mathbb{E}\left[u_i(a, r_{-i})|\sigma_i, \mu\right] &> \mathbb{E}\left[u_i(r)|\mu\right]. \end{aligned}$$

Therefore, the existence of $x(r, y)$ is equivalent to non-existence of $\sigma_i \in \Sigma_i$ with (3) and (4):

$$k(\mu, \lambda) \equiv \max_{\mu \in \Delta(A)} \lambda \cdot u(\mu)$$

such that, for each $i \in I$, there does not exist $\sigma_i \in \Sigma_i$ with (3) and (4).

7.2 Proof of Proposition 2

First, we show that, if there exists player j with $\lambda_j > 0$, then $k(\lambda) \leq l(\lambda)$. To see why, take v, μ, and x such that [incentive compatibility] and [promise keeping] are satisfied: For each $i \in I$ and $\sigma_i \in \Sigma_i$,

$$v_i = \mathbb{E}\left[u_i(r) + x_i(r, y)|\mu\right] \geq \mathbb{E}\left[u_i(a_i, r_{-i}) + x_i(r, m_i, y_{-i})|\sigma_i, \mu\right]. \tag{22}$$

Let us define $\bar{x}(r, y)$ such that, for each $r \in A$ and $y \in Y$,

$$\bar{x}_i(r, y) = \begin{cases} x_i(r, y) - \frac{K}{\lambda_i} & \text{for player } i = j, \\ x_i(r, y) & \text{for player } i \neq j \end{cases}$$

with a large $K > 0$. Since $\lambda_j > 0$, this means that we reduce player j's continuation payoff. Note that $\tau = \mu$ and $\bar{x}(r, y)$ satisfies the following:

1. For each i and each strategy σ_i, since (22) holds and $\frac{K}{\lambda_j} > 0$, we have

$$v_i \geq \mathbb{E}\left[u_i(a_i, r_{-i}) + \bar{x}_i(r, m) \mid \sigma_i, \tau\right].$$

Hence, [effective punishment] is satisfied.

2. For each i and each strategy $\sigma_i \in \Sigma_i$, we have

$$\mathbb{E}\left[\lambda \cdot \bar{x}(r, m) \mid \sigma_i, \tau\right] = \mathbb{E}\left[\lambda \cdot x(r, m) \mid \sigma_i, \tau\right] - K.$$

Hence, [self generation after unilateral deviation] is satisfied for sufficiently large K.

Second, we show that, if $\lambda_i \leq 0$ for all i, then

$$l(\lambda) = \min_{\tau \in \Delta(A)} \max_{\sigma \in SD(\text{supp}(\lambda))} \sum_{r \in A} \tau(r) \sum_{i \in N} |\lambda_i| \, u_i(r, \sigma_i).$$

Recall that $l(\tau, \lambda)$ is the maximum of $\lambda \cdot v$ such that there exists $x(r, m)$ such that the following two conditions are satisfied:

1. [effective punishment] For each i and $s_i \in S_i$, the punishment strategy is not profitable:

$$v_i \geq \mathbb{E}\left[u_i(a_i, r_{-i}) + x_i(r, m) \mid s_i, \tau\right].$$

2. [self generation after unilateral deviation] For each i and $s_i \in S_i$, in expectation, the score $\lambda \cdot x(r, y)$ is non positive:

$$\mathbb{E}\left[\lambda \cdot x(r, m) \mid s_i, \tau\right] \leq 0.$$

For i with $\lambda_i = 0$, defining $x_i(r, m) = 0$ for all r and m, sufficiently large v_i satisfies [effective punishment] without affecting $\lambda \cdot v$ and [self generation after unilateral deviation]. Hence, we focus on $i \in \text{supp}(\lambda)$.

Let $x_i(A,Y) \equiv \{x_i(r,m)\}_{r,m}$ be the column vector expression of $x_i(r,m)$; and let $\Pr(A,Y|s_i,\tau) \equiv \{\Pr(r,m|s_i,\tau)\}_{r,m}$ be the column vector expression of $\Pr(r,m|s_i,\tau)$. Given this notation, the above two conditions are equivalent to the following:

1. For each $i \in \text{supp}(\lambda)$ and $s_i \in S_i$, we have

$$\begin{pmatrix} \mathbf{0} \\ \Pr(A,Y|s_i,\tau) \\ \mathbf{0} \end{pmatrix} \cdot \begin{pmatrix} x_1(A,Y) \\ \vdots \\ x_i(A,Y) \\ \vdots \\ x_N(A,Y) \end{pmatrix} \leq v_i - u_i(s_i,\tau).$$

2. For each $i \in \text{supp}(\lambda)$ and $s_i \in S_i$, we have

$$\begin{pmatrix} \lambda_1 \Pr(A,Y|s_i,\tau) \\ \vdots \\ \lambda_N \Pr(A,Y|s_i,\tau) \end{pmatrix} \begin{pmatrix} x_1(A,Y) \\ \vdots \\ x_N(A,Y) \end{pmatrix} \leq 0.$$

By Theorem 22.1 of Rockafellar (1960), the existence of such $(x_i(A,Y))_i$ is equivalent to the non-existence of non-negative $\{\sigma_i(s_i), \eta_i(s_i)\}_{i,s_i}$ such that

1.

$$\sum_{i,s_i} \sigma_i(s_i)\,(v_i - u_i(\sigma_i,\tau)) < 0.$$

2. For all $i \in \text{supp}(\lambda)$,

$$\sum_{s_i} \sigma_i(s_i) \Pr(A,Y|s_i,\tau) + \lambda_i \sum_{n \in \text{supp}(\lambda), s_n} \eta_n(s_n) \Pr(A,Y|s_n,\tau) = \mathbf{0}.$$

Dividing both sides by $|\lambda_i| \neq 0$, these two conditions are equivalent to

1.

$$\sum_{i\in \text{supp}(\lambda),s_i} \frac{\sigma_i(s_i)}{|\lambda_i|}\left(|\lambda_i|\, v_i - |\lambda_i|\, u_i(s_i,\tau)\right) < 0. \tag{23}$$

2. For all $i \in \text{supp}\,(\lambda)$,

$$\sum_{s_i} \frac{\sigma_i(s_i)}{|\lambda_i|} \Pr(A,Y|s_i,\tau) = \sum_{n\in \text{supp}(\lambda),s_n} \eta_n(s_n) \Pr(A,Y|s_n,\tau).$$

Adding the second condition with respect to $(r,m) \in A \times Y$, we have

$$\sum_{s_i} \frac{\sigma_i(s_i)}{|\lambda_i|} = \sum_{n\in \text{supp}(\lambda),s_n} \eta_n(s_n). \tag{24}$$

Note that $\sum_{n,s_n} \eta_n(s_n) = 0$ implies $\sigma_i(s_i) = 0$ for all s_i (recall that $\sigma_i(s_i) \geq 0$ for all s_i). Since $\sigma_i(s_i) = 0$ for all s_i does not satisfy (23), we should have $\sum_{n,s_n} \eta_n(s_n) > 0$. Dividing both sides $\sum_{n,s_n} \eta_n(s_n)$, without loss, we can assume that $\sum_{n,s_n} \eta_n(s_n) = 1$. Then, (24) implies $\sum_{s_i} \frac{\sigma_i(s_i)}{|\lambda_i|} = 1$ for all i.

Given $\sum_{n\in \text{supp}(\lambda),s_n} \eta_n(s_n) = 1$ and $\sum_{s_i} \frac{\sigma_i(s_i)}{|\lambda_i|} = 1$ for all $i \in \text{supp}\,(\lambda)$, the above two conditions are equivalent to

1.

$$|\lambda| \cdot v - \sum_i \sum_{s_i} \frac{\sigma(s_i)}{|\lambda_i|} |\lambda_i|\, u_i(s_i,\tau) < 0.$$

2. For all $i \in \text{supp}\,(\lambda)$,

$$\sum_{s_i} \frac{\sigma(s_i)}{|\lambda_i|} \Pr(A,Y|s_i,\tau) = \sum_{n\in \text{supp}(\lambda),s_n} \eta(s_n) \Pr(A,Y|s_n,\tau)$$

$\Leftrightarrow$

$$\sum_{s_i} \frac{\sigma(s_i)}{|\lambda_i|} \Pr(m|r,s_i) = \sum_{n\in \text{supp}(\lambda),s_n} \eta(s_n) \Pr(m|r,s_n) \text{ for all } r \in \text{supp}\,(\tau) \text{ and } m \in Y.$$

Seeing $\left(\frac{\sigma_i(s_i)}{|\lambda_i|}\right)_{s_i}$ as player i's deviation (recall that $\sum_{s_i} \frac{\sigma_i(s_i)}{|\lambda_i|} = 1$ and $\sigma_i(s_i) \geq 0$ for each $i \in \text{supp}(\lambda)$ and $s_i \in S_i$), the non-existence of $\left\{\{\sigma_i(s_i), \eta(\sigma_i)\}_{\sigma_i}\right\}_{i \in \text{supp}(\lambda)}$ is equivalent to the non-existence of $(\sigma_i)_{i \in \text{supp}(\lambda)}$ with $\sigma_i \in \Sigma_i$ for each $i \in \text{supp}(\lambda)$ such that, for each $i, j \in \text{supp}(\lambda)$, we have

$$|\lambda| \cdot v < \sum_i |\lambda_i| \, u_i(\sigma_i, \tau), \tag{25}$$

$$\Pr(m|r, \sigma_i) = \Pr(m|r, \sigma_j) \text{ for all } r \in \text{supp}(\tau) \text{ and } m \in Y. \tag{26}$$

This implies that, without loss, we can focus on the case in which τ has the full support: $\text{supp}(\lambda) = A$. Otherwise, with a small ε, consider τ_ε with $\tau_\varepsilon(a) = (1-\varepsilon)\,\tau(a) + \frac{\varepsilon}{|A|}$. For such τ_ε, for sufficiently small ε, if there does not exist $(\sigma_i)_{i \in \text{supp}(\lambda)}$ with (25) and (26), then there does not exist $(\sigma_i)_{i \in \text{supp}(\lambda)}$ with

$$|\lambda| \cdot v < \sum_i |\lambda_i| \, u_i(\sigma_i, \tau_\varepsilon),$$

$$\Pr(m|r, \sigma_i) = \Pr(m|r, \sigma_j) \text{ for all } i, j \in \text{supp}(\lambda), r \in A, \text{ and } m \in Y.$$

Therefore, $l(\tau, \lambda) \leq l(\tau_\varepsilon, \lambda)$ for sufficiently small ε. Since $l(\lambda) = \sup_{\tau \in \Delta(A)} l(\tau, \lambda)$, we have $l(\lambda) = \sup_{\tau \in \Delta(A):\text{supp}(\tau) = A} l(\tau, \lambda)$.

For τ with $\text{supp}(\tau) = A$, for each v, the existence of $x(r, m)$ with [effective punishment] and [self generation after unilateral deviation] is equivalent to non-existence of$(\sigma_i)_{i \in \text{supp}(\lambda)}$ with

$$|\lambda| \cdot v < \sum_i |\lambda_i| \, u_i(\sigma_i, \tau),$$

$$\Pr(m|r, \sigma_i) = \Pr(m|r, \sigma_j) \text{ for all } i, j \in \text{supp}(\lambda), r \in A, \text{ and } m \in Y.$$

This is equivalent to

$$|\lambda| \cdot v \geq \min_{\tau \in \Delta(A)} \max_{\sigma \in SD(\text{supp}(\lambda))} \sum_{r \in A} \tau(r) \sum_{i \in N} |\lambda_i| \, u_i(r, \sigma_i).$$

Therefore, since $\lambda_i \leq 0$ for all i, we have

$$-l(\lambda) \geq \min_{\tau \in \Delta(A)} \max_{\sigma \in SD(\mathrm{supp}(\lambda))} \sum_{r \in A} \tau(r) \sum_{i \in N} |\lambda_i| \, u_i(r, \sigma_i).$$

7.3 Proof of Theorem 3

7.3.1 Linear Programming

To prove theorem 3, we first give sufficient conditions for $\underline{Q} \subseteq \lim_{\delta \to 1} E_{\mathrm{seq}}(\delta)$, motivated by Fudenberg and Levine (1994), and Kandori and Matsushima (1998).

To prove $\underline{Q} \subseteq \lim_{\delta \to 1} E_{\mathrm{seq}}(\delta)$, it suffices to show that, for each $\lambda \in \Lambda$ and $\varepsilon > 0$, we can find T, $\bar{\delta}$, σ_m^T, v, and $w(h_m^{T+1}, \delta)$ for $\delta > \bar{\delta}$, satisfying the following three conditions: Suppose that the players play the repeated game for T periods. $\bar{\delta}$ is the lower bound of the discount factor above which we can find an equilibrium. Intuitively, $w(h_m^{T+1}, \delta)$ is a continuation payoff from periods $T+1$ on if the discount factor is δ. The mediator's strategy during the first T periods is $\sigma_m^T : h_m^t \mapsto \mu_t \in \Delta(A)$ for each $t = 1, ..., T$, while player i's strategy during the first T periods is $\sigma_i^T : h_i^t \mapsto \sigma_i \in \Sigma_i$ for each $t = 1, ..., T$. Let Σ_i^T be the set of player i's strategies. Then, σ_m^T should satisfy the following:

1. [FL: incentive compatibility] After each history h_i^t, it is optimal to be faithful: For each i, h_i^t, and $\sigma_i^T \in \Sigma_i$,

$$\begin{aligned} & \mathbb{E}\left[(1-\delta)\sum_{t=1}^{T} \delta^{t-1} u_i(a) + \delta^T w_i(h_m^{T+1}, \delta) | h_i^t, \sigma_m^T\right] \\ \geq \; & \mathbb{E}\left[(1-\delta)\sum_{t=1}^{T} \delta^{t-1} u_i(a) + \delta^T w_i(h_m^{T+1}, \delta) | h_i^t, \sigma_m^T, \sigma_i^T\right]. \end{aligned}$$

   Here, we use σ_m^T to represent the strategy profile such that all the players are faithful, and σ_m^T, σ_i^T to represent the strategy profile such that player i takes σ_i^T while all the other players are faithful.

2. [FL: promise keeping] $v \in \mathbb{R}^N$ is the equilibrium payoff with the faithful strategy profile,

which is close to the extreme point in the targeted payoff set:

$$\begin{aligned} v &= \mathbb{E}\left[(1-\delta)\sum_{t=1}^{T}\delta^{t-1}u(a)+\delta^T w(h_m^{T+1},\delta)|\sigma_m^T\right], \\ \lambda\cdot v &\geq \max_{v'\in\underline{Q}}\lambda\cdot v'-\varepsilon. \end{aligned}$$

3. [FL: self generation] For each h_m^{T+1}, we have

$$\lambda\cdot w(h_m^{T+1},\delta)\leq\lambda\cdot v$$

   and

$$\lim_{\delta\to 1}\sup\max_{i,h_m^{T+1},\delta}\frac{\left|w_i(h_m^{T+1},\delta)\right|}{1-\delta}<\infty.$$

   That is, the continuation payoff profile is below the hyperplane supporting the equilibrium payoff set after each history, and its variation is of order $1-\delta$.

The key difference from the characterization of $\underline{Q}$ is that self generation should be satisfied after all the histories h_m^{T+1}, rather than in expectation.

Lemma 7 *Suppose that, for each $\lambda\in\Lambda$ and $\varepsilon>0$, we can find T, $\bar{\delta}$, σ_m^T, v, and $w(h_m^{T+1},\delta)$ for $\delta>\bar{\delta}$ such that [FL: incentive compatibility], [FL: promise keeping], and [FL: self generation] are satisfied. Then, if $\dim(\underline{Q})=N$, we have $\underline{Q}\subseteq\lim_{\delta\to 1}E_{\text{seq}}(\delta)$.*

Proof. The same as Fudenberg and Levine (1994). ■

We now fix $\lambda\in\Lambda$ arbitrarily and define these objects T, $\bar{\delta}$, σ_m^T, v, and $w(h_m^{T+1},\delta)$.

First, consider the case with $\lambda=e_i$ or $-e_i$ for some player i. Then, by definition of $\underline{k}(\lambda)$, for each $\varepsilon>0$, there exist $\mu\in\Delta(A)$, v, and $x(r,m)$ such that

1. [incentive compatibility] For each $i\in I$, the faithful strategy is optimal when the change in the continuation payoff is given by $x(r,m)$: For all $s_i\in S_i$,

$$\mathbb{E}\left[u_i(r)+x_i(r,y)|\mu\right]\geq\mathbb{E}\left[u_i(a_i,r_{-i})+x_i(r,m_i,y_{-i})|s_i,\mu\right]. \tag{27}$$

2. [promise keeping] v is the value of the repeated game: $v = u(\mu) + \mathbb{E}[x(r,y)|\mu]$, and

$$\lambda \cdot v = \underline{k}(\lambda, \mu) \geq \underline{k}(\lambda) - \varepsilon. \tag{28}$$

3. [ex post self generation] If $\lambda = e_i$ or $-e_i$ for some $i \in I$, then after each $r \in \text{supp}(\mu)$ and $y \in Y$, $\lambda \cdot x(r,m)$ is non-positive:

$$\lambda \cdot x(r,m) \leq 0 \tag{29}$$

for each $r \in \text{supp}(\mu)$ and $y \in Y$.

Without loss, we can assume that $x(r,m) = 0$ for $r \notin \text{supp}(\mu)$.

With $T = 1$, $\bar{\delta} > 0$, $\sigma_m^T = \mu$, v, and $w(h_m^2, \delta) = v + \frac{1-\delta}{\delta} x(r,m)$, we can satisfy [FL: incentive compatibility], [FL: promise keeping], and [FL: self generation]:

1. [FL: incentive compatibility] Since

$$\begin{aligned} & \mathbb{E}\left[(1-\delta)u_i(a) + \delta w(h_m^2, \delta)|\mu, \sigma_i\right] \\ = \; & (1-\delta)\,\mathbb{E}\left[u_i(a) + x(h_m^2, \delta)|\mu, \sigma_i\right] + \delta v, \end{aligned}$$

(27) implies [FL: promise keeping].

2. [FL: promise keeping] Since

$$\begin{aligned} & \mathbb{E}\left[(1-\delta)u(a) + \delta w(h_m^2, \delta)|\mu\right] \\ = \; & (1-\delta)\,\mathbb{E}\left[u_i(a) + x(h_m^2, \delta)|\mu\right] + \delta v \\ = \; & v, \end{aligned}$$

(28) implies [FL: incentive compatibility].

3. [FL: self generation] For each r, m, we have

$$\lambda \cdot w(h_m^2, \delta) = \lambda \cdot v + \lambda \cdot \frac{1-\delta}{\delta} x(r,m)$$

and

$$\lim_{\delta \to 1} \sup \max_{i, h_m^2, \delta} \frac{|w_i(h_m^2, \delta)|}{1-\delta} = \max_{i,r,m} |x_i(r,m)|,$$

(29) implies [FL: self generation].

Hence, we will focus on the case in which there are more than one player whose Pareto weight is non-zero: $\text{supp}(\lambda) \geq 2$.

7.3.2 Mediator's Strategy

Regular Recommendation We first pick μ according to which the mediator recommends r_t with a high probability. We fix μ such that it gives a high $\underline{k}(\mu, \lambda)$: $\underline{k}(\mu, \lambda) \geq \max_{v' \in \underline{Q}} \lambda \cdot v' - \frac{\varepsilon}{5}$.

For notational convenience, let $v^\lambda \in \underline{Q}$ such that

$$\lambda \cdot v^\lambda = \max_{v' \in \underline{Q}} \lambda \cdot v' - \frac{\varepsilon}{5}.$$

Since $\dim(\underline{Q}) = N$, such v^λ exists.

Given μ, let v^μ be the value associated with μ. Then, we have

$$\lambda \cdot v^\mu = \underline{k}(\mu, \lambda) \geq \max_{v' \in \underline{Q}} \lambda \cdot v' - \frac{\varepsilon}{5} = \lambda \cdot v^\lambda. \tag{30}$$

Take $x^\mu(r,m)$ in definition of $\underline{k}(\mu, \lambda)$:

$$\mathbb{E}\left[u(r) + x^\mu(r,m) | \mu\right] = v^\mu.$$

From $x^\mu(r,m)$, we define

$$x(r, m|\mu) \equiv x^\mu(r,m) + v^\lambda - v^\mu.$$

Note that, with μ and $x(r, m|\mu)$, we have the following four conditions:

1. [incentive compatibility] is satisfied: For each i, $r_i \in \operatorname{supp}(\mu_i)$, and $s_i(r_i) \in S_i(r_i)$, we have

$$\mathbb{E}\left[u_i(r) + x_i(r, y|\mu)|\mu\right] \geq \mathbb{E}\left[u_i(a_i, r_{-i}) + x_i(r, m_i, y_{-i}|\mu)|s_i, \mu\right]. \tag{31}$$

2. [strict incentive compatibility] is satisfied: For each i, $r_i \in \operatorname{supp}(\mu_i)$, $s_i(r_i) \in S_i(r_i)$ with

$$\Pr(m|r_i, r_{-i}) \neq \Pr(m|s_i(r_i), r_{-i}) \text{ for some } r_{-i} \in A_{-i} \text{ and } m \in Y,$$

we have

$$\mathbb{E}\left[u_i(r) + x_i(r, y|\mu)|\mu, r_i\right] > \mathbb{E}\left[u_i(a_i, r_{-i}) + x_i(r, m_i, y_{-i}|\mu)|s_i, \mu, r_i\right]. \tag{32}$$

3. The value is v^λ

$$\lambda \cdot v^\lambda = \lambda \cdot \mathbb{E}\left[u(r) + x(r, m|\mu)|\mu\right] \geq \max_{v' \in \underline{Q}} \lambda \cdot v' - \frac{\varepsilon}{5}. \tag{33}$$

4. by (30), [self generation] is satisfied:

$$\mathbb{E}\left[\lambda \cdot x(r, m|\mu)|\mu\right] \leq 0. \tag{34}$$

Punishment Recommendation We second pin down the recommendation schedule τ, which the mediator uses in order to punish the players after a history statistically indicating a deviation. We fix strictly incentive compatible τ such that $l(\tau, \lambda)$ is close to $\underline{l}(\lambda)$:

$$l(\tau, \lambda) \geq \max_{v' \in \underline{Q}} \lambda \cdot v' - \frac{\varepsilon}{10}.$$

Given τ, there exist $\tilde{v}^\tau$ and $x^\tau(r,m)$ such that

$$\begin{aligned} l(\tau,\lambda) &= \lambda\cdot\tilde{v}^\tau \geq \max_{v'\in\underline{Q}} \lambda\cdot v' - \frac{\varepsilon}{10} = \max_{v'\in\underline{Q}} \lambda\cdot v' - \frac{\varepsilon}{5} + \frac{\varepsilon}{10} \\ &= \lambda\cdot v^\lambda + \frac{\varepsilon}{10} \text{ by (30).} \end{aligned} \tag{35}$$

and the following two conditions are satisfied:

1. [effective punishment] For each i and $s_i \in S_i$,

$$\tilde{v}_i^\tau \geq \mathbb{E}\left[u_i(a_i, r_{-i}) + x_i^\tau(r,m) \mid s_i, \tau\right].$$

2. [self generation after unilateral deviation] For each i and $s_i \in S_i$,

$$\mathbb{E}\left[\lambda\cdot x^\tau(r,m) \mid s_i,\tau\right] \leq 0.$$

Define v^{punish} such that

$$v_i^{\text{punish}} = v_i^\lambda - \frac{\varepsilon}{10}. \tag{36}$$

With v^{punish}, by (35), we have

$$\lambda\cdot v^{\text{punish}} \leq \lambda\cdot v^\lambda + \frac{\varepsilon}{10} \leq \lambda\cdot\tilde{v}^\tau. \tag{37}$$

We define

$$x^{\text{punish}}(r,m|\tau) \equiv x^\tau(r,m) + v^{\text{punish}} - v^\tau.$$

Note that, with τ and $x(r,m|\tau)$, we have the following three conditions:

1. The value v^λ is sufficiently large: For each i and $s_i \in S_i$,

$$v_i^\lambda > v_i^{\text{punish}} = \mathbb{E}\left[u_i(a_i,r_{-i}) + x_i^{\text{punish}}(r,m|\tau) \mid s_i,\tau\right]. \tag{38}$$

2. By (37), [self generation after unilateral deviation] is satisfied: For each i and $s_i \in S_i$,

$$\begin{aligned} & \mathbb{E}\left[\lambda \cdot x^{\text{punish}}(r, m|\tau) \mid s_i, \tau\right] \\ = \ & \mathbb{E}\left[\lambda \cdot x^{\tau}(r, m) \mid s_i, \tau\right] + \lambda \cdot v^{\text{punish}} - \lambda \cdot v^{\tau} \leq 0. \end{aligned} \tag{39}$$

At the same time, since τ satisfies [strict incentive compatibility], there exists $x(r, m|\tau)$ such that, for each i, $r_i \in \text{supp}(\tau_i)$, $s_i(r_i) \in S_i(r_i)$ with

$$\Pr(m|r_i, r_{-i}) \neq \Pr(m|s_i(r_i), r_{-i}) \text{ for some } r_{-i} \in A_{-i} \text{ and } m \in Y,$$

we have

$$\mathbb{E}\left[u_i(r) + x_i(r, m|\tau)|\tau, r_i\right] > \mathbb{E}\left[u_i(a_i, r_{-i}) + x_i(r, m_i, y_{-i}|\tau)|s_i, \tau, r_i\right]. \tag{40}$$

By adding/subtracting a constant, we make sure that

$$\lambda \cdot x(r, m|\tau) \leq 0 \text{ for all } r \in A \text{ and } m \in Y. \tag{41}$$

For notational convenience, let

$$v^{\tau} = \mathbb{E}\left[u(r) + x(r, m|\tau)|\tau\right]$$

be the value when τ is recommended and $x(r, m|\tau)$ is used.

Blocks and States We third consider the mediator's strategy. The mediator sees entire T periods as B repetitions of T/B-period blocks. That is, there are B blocks, and block b consists of periods $(b-1)T/B + 1, ..., bT/B$ for $b = 1, ..., B$. Let $t_b \equiv (b-1)T/B$ (that is, $t_b + 1$ is the first period of block b); and let $T_b \equiv \{t_b + 1, ..., t_{b+1}\}$ be the set of periods in block b.

We fix B sufficiently large so that

$$\frac{\max_{r\in A, y\in Y} \lambda \cdot x(r,y|\mu)}{B} < \frac{\varepsilon}{4}. \tag{42}$$

The mediator attaches two possible states $s(b) \in \{G, B\}$ to each block. Intuitively, $s(b) = G$ implies that [FL: self generation] is not an issue until block b while $s(b) = B$ implies that it has been an issue.

Next, we define the mediator's strategy in each block b given the state $s(b)$, and then define the state transition.

Strategy in Each Block Given the States We fourth define the mediator's strategy in each block b given the state $s(b)$. To this end, it will be useful to define $\eta > 0$ as follows: By (33), for sufficiently small $\eta > 0$, we have

$$(1 - B\eta)\lambda \cdot v^{\lambda} + B\eta \min\left\{\lambda \cdot v^{\lambda}, \lambda \cdot v^{\tau}\right\} \geq \max_{v' \in \underline{Q}} \lambda \cdot v' - \frac{\varepsilon}{4}. \tag{43}$$

In addition, by (36), for sufficiently small $\eta > 0$, we can make sure that

$$(1 - B\eta)v_i^{\mu} + B\eta \min\left\{v_i^{\mu}, v_i^{\tau}\right\} > v_i^{\text{punish}} \tag{44}$$

for all $i \in I$. We fix $\eta > 0$ such that (43) and (44) hold.

Given $\eta > 0$, in each block b, if the state is G, then the mediator first mixes the recommendation schedule μ and τ: The mediator picks μ with probability $1 - \eta$ and τ with probability η. Let $\alpha(b) \in \{\mu, \tau\} \subset \Delta(A)$ be the picked recommendation schedule in block b. Given $\alpha(b)$, the mediator recommends r_t according to $\alpha(b)$ *i.i.d.* across periods.

On the other hand, if the state is B, then the mediator recommends r_t according to τ for sure. That is, if $s(b) = B$, then $\alpha(b) = \tau$ with probability one, and the mediator recommends r_t according to $\alpha(b) = \tau$ *i.i.d.* across periods.

State Transition We fifth define the state transition $s(b)$. The initial condition is $s(1) = G$. In addition, once $s(b) = B$ happens, we have $s(b+1) = B$ for sure. Hence, we concentrate on the case with $s(b) = G$ and give the condition with which we have $s(b+1) = B$.

To this end, it will be useful to define $t_{\text{likelihood}} > 0$, $\varepsilon_{\text{likelihood}} > 0$, $K_{\text{affine}} > 0$, and $\varepsilon_{\text{affine}} > 0$ as follows. Let

$$f_i(A_i|b) \equiv (f_i(r_i|b))_{r_i \in A_i} \equiv \left(\frac{\# \{t \in T_b : r_{i,t} = r_i\}}{T/B} \right)_{r_i \in A_i}$$

be the frequency of periods in block b such that player i is recommended to take r_i. In addition, for all r_i with $f_i(r_i|b) > 0$, let

$$\begin{aligned} f_i(Y_i|r_i, b) &\equiv (f_i(y_i|r_i, b))_{y_i \in Y_i} \\ &\equiv \left(\frac{\# \{t \in T_b : r_{i,t} = r_i, y_{i,t} = y_i\}}{f_i(r_i|b)} \right)_{y_i \in Y_i} \end{aligned}$$

be the conditional frequency of periods in block b such that player i observes y_i, conditional on that player i is recommended to take r_i. For r_i with $f_i(r_i|b) = 0$, define $f_i(y_i|r_i, b) = 0$ for all $y_i \in Y_i$.

On the other hand, for all r_i with $f_i(r_i|b) > 0$, let

$$\begin{aligned} f_{-i}(A_{-i}|r_i, b) &\equiv (f_{-i}(r_{-i}|r_i, b))_{r_{-i} \in A_{-i}} \\ &\equiv \left(\frac{\# \{t \in T_b : r_t = r\}}{f_i(r_i|b)} \right)_{r_{-i} \in A_{-i}} \end{aligned}$$

be the conditional frequency of periods in block b such that the recommendation profile is r, conditional on that player i is recommended to take r_i. For r_i with $f_i(r_i|b) = 0$, define $f_i(r_{-i}|r_i, b) = 0$ for all $r_{-i} \in A_{-i}$.

In addition, let

$$q(y_i|\mu_{-i}|_{r_i}, r_i) \equiv \sum_{\tilde{r}_{-i} \in A_{-i}} \mu(\tilde{r}_{-i}|r_i) q(y_i|r_i, \tilde{r}_{-i})$$

be the conditional probability of player i's signal y_i given that the mediator uses μ and player

i is recommended to take r_i. That is, $q(y_i|\mu_{-i}|_{r_i}, r_i)$ is the ex ante mean of $f_i(y_i|r_i, b)$.

Consider player i's faithful history in block b, denoted by $h_i^b = (r_{i,t}, y_{i,t})_{t=t_b+1}^{t_b+T/B}$.[2] We will identify the set of player i's histories after which player i believes that the realized frequency of the recommendation to the other players given r_i is "erroneous."

Lemma 8 *For each $b = 1, ..., B$, suppose that the mediator uses $\alpha(b) = \mu$. There exists $t_{\text{likelihood}} > 0$ such that, for each $\varepsilon_{\text{likelihood}} > 0$, $i \in I$, and $h_i^b = (r_{i,t}, y_{i,t})_{t=t_b+1}^{t_b+T/B}$, if there exist $r_i \in \text{supp}(\mu_i)$ and $r_{-i} \in \text{supp}\left(\mu_{-i}|_{r_i}\right)$ such that*

$$\left| \sum_{y_i} f_i(y_i|r_i, b) \frac{q(y_i|r_{-i}, r_i) - q(y_i|\mu_{-i}|_{r_i}, r_i)}{q(y_i|\mu_{-i}|_{r_i}, r_i)} \right| > \varepsilon_{\text{likelihood}}, \tag{45}$$

then player i believes that the frequency of r_{-i} given r_i is far from the true conditional probability:

$$\begin{aligned} & \Pr\left(|f_{-i}(r_{-i}|r_i, b) - \mu(r_{-i}|r_i)| > t_{\text{likelihood}}\varepsilon_{\text{likelihood}} | \alpha(b) = \mu, h_i^b\right) \\ \geq\ & 1 - \exp\left(-\left(t_{\text{likelihood}}\varepsilon_{\text{likelihood}}\right)^2 \frac{T}{B}\right). \end{aligned}$$

In (45), whenever $f_i(y_i|r_i, b) = 0$, we define

$$f_i(y_i|r_i, b) \frac{q(y_i|r_{-i}, r_i) - q(y_i|\mu_{-i}|_{r_i}, r_i)}{q(y_i|\mu_{-i}|_{r_i}, r_i)} = 0$$

even if $q(y_i|\mu_{-i}|_{r_i}, r_i) = 0$. Then, (45) is well defined by the following reason: By the effective full support, we have $q(y_i|\mu_{-i}|_{r_i}, r_i) > 0$ for each y_i if $q(y_i|r_{-i}, r_i) > 0$ for some r_{-i}. For y_i with $q(y_i|r_{-i}, r_i) = 0$ for all r_{-i}, we have $f_i(y_i|r_i, b) = 0$ with probability one.

Proof. The condition (45) implies that, for each r_{-i} with $\mu(r_{-i}|r_i) > 0$, we have

$$\left| \sum_{y_i} f_i(y_i|r_i, b) \frac{\mu(r_{-i}|r_i) q(y_i|r_i, r_{-i})}{\sum_{\tilde{r}_{-i} \in A_{-i}} \mu(\tilde{r}_{-i}|r_i) q(y_i|r_i, \tilde{r}_{-i})} - \mu(r_{-i}|r_i) \right| > \min_{\tilde{r}: \mu(\tilde{r}_{-i}|\tilde{r}_i) > 0} \mu(\tilde{r}_{-i}|\tilde{r}_i) \varepsilon_{\text{likelihood}}.$$

[2] Since the history is faithful, we omit $a_{i,t} = r_{i,t}$ and $m_{i,t} = y_{i,t}$.

By Bayes rule, this is equivalent to

$$\left|\sum_{y_i} f_i(y_i|r_i,b)\Pr(r_{-i}|r_i,y_i)-\mu(r_{-i}|r_i)\right| > \min_{\tilde{r}:\mu(\tilde{r}_{-i}|\tilde{r}_i)>0}\mu(\tilde{r}_{-i}|\tilde{r}_i)\varepsilon_{\text{likelihood}}.$$

Let

$$t_{\text{likelihood}}=\frac{1}{2}\min_{\tilde{r}:\mu(\tilde{r}_{-i}|\tilde{r}_i)>0}\mu(\tilde{r}_{-i}|\tilde{r}_i).$$

Then, since $\sum_{y_i} f_i(y_i|r_i,b)\Pr(r_{-i}|r_i,y_i)$ is the conditional expectation of $f_{-i}(r_{-i}|r_i,b)$ given h_i^b, by Hoeffding's inequality, we have

$$\begin{aligned}&\Pr\left(|f_{-i}(r_{-i}|r_i,b)-\mu(r_{-i}|r_i)|>t_{\text{likelihood}}\varepsilon_{\text{likelihood}}|\alpha(b)=\mu,h_i^b\right)\\ \geq\ & 1-\exp\left(-\left(t_{\text{likelihood}}\varepsilon_{\text{likelihood}}\right)^2\frac{T}{B}\right),\end{aligned}$$

as desired. ■

We now fix $t_{\text{likelihood}}>0$. On the other hand, we will fix $\varepsilon_{\text{likelihood}}$ together with $\varepsilon_{\text{affine}}>0$.

Let us now explain the meaning of $K_{\text{affine}}>0$ and $\varepsilon_{\text{affine}}>0$. Let $h_m^b=(r_t,m_t)_{t=t_b+1}^{t_b+T/B}$ be the mediator's history in block b. For each i and $r_i\in\operatorname{supp}(\mu_i)$, let

$$H_i(Y_i|r_i)\equiv\operatorname{aff}\left(\{q_i(Y_i|r_i,a_{-i})\}_{a_{-i}\in A_{-i}}\right)$$

be the affine hull of player i's signal distributions with respect to players $-i$'s actions. [effective full support] ensures that

$$H_i(Y_i|r_i)=\operatorname{aff}\left(\{q_i(Y_i|r_i,r_{-i})\}_{r_{-i}\in\operatorname{supp}(\mu_{-i}|r_i)}\right). \tag{46}$$

Lemma 16 of Sugaya (2014a) shows that there exists $K_{\text{affine}}>0$ such that, for each $\varepsilon_{\text{affine}}>0$, there exists a random variable $Z^i_{\varepsilon_{\text{affine}}}(b,h_m^b)\in\{0,1\}$ such that if player i is faithful and so $m_{i,t}=y_i$ for all $t=t_b+1,...,t_b+T/B$, then the following three conditions are satisfied:

1. [Condition Z independence] Players $-i$ cannot control the distribution of $Z^i_{\varepsilon_{\text{affine}}}(b, h^b_m)$: Given that the mediator recommends $\alpha(b) = \mu$ and player i takes the faithful strategy σ_i^*,
$$\Pr(\{Z^i_{\varepsilon_{\text{affine}}}(b, h^b_m) = 0\} | \mu, \sigma_i^*, \sigma_{-i}^T)$$
is independent of σ_{-i}^T.

2. [Condition Z rare] $Z^i_{\varepsilon_{\text{affine}}}(b, h^b_m)$ is equal to zero with a high probability: Given $\alpha(b) = \mu$ and σ_i^*,
$$\Pr(\{Z^i_{\varepsilon_{\text{affine}}}(b, h^b_m) = 0\} | \mu, \sigma_i^*) \geq 1 - \exp\left(-\frac{\varepsilon^2_{\text{affine}}}{K_{\text{affine}}} \frac{T}{B}\right). \quad (47)$$
Here, we omit σ_{-i}^T since the distribution is independent of σ_{-i}^T from [Condition Z independence].

3. [Condition Z distance] If $f_i(Y_i|r_i, b)$ is far from $H_i(Y_i|r_i)$, then $Z^i_{\varepsilon_{\text{affine}}}(b, h^b_m) = 1$: If there exists $r_i \in \text{supp}(\mu)$ such that
$$d\left(f_i(Y_i|r_i, b), H_i(Y_i|r_i)\right) > \varepsilon_{\text{affiie}},$$
then
$$Z^i_{\varepsilon_{\text{affiie}}}(b, h^b_m) = 1.$$
In general, $d(a, A)$ is Hausdorff metric between a and A.

Intuitively, $Z^i_{\varepsilon_{\text{affine}}}(b, h^b_m) = 1$ if and only if the frequency of y_i is far from the affine hull of player i's signal distributions with respect to players $-i$'s actions. Since we take the affine hull, players $-i$ cannot control $Z^i_{\varepsilon_{\text{affine}}}(b, h^b_m)$.

We now fix $K_{\text{affine}} > 0$. Finally, let us fix $\varepsilon_{\text{likelihood}} > 0$ and $\varepsilon_{\text{affine}} > 0$ so that the following lemma holds:

Lemma 9 *For each $b = 1, ..., B$, suppose that the mediator uses $\alpha(b) = \mu$. There exist $\varepsilon_{\text{likelihood}} > 0$ and $\varepsilon_{\text{affine}} > 0$ such that, for each $i \in I$ and her faithful history $h^b_i =$*

$(r_{i,t}, y_{i,t})_{t=t_b+1}^{t_b+T/B}$, *the following claim holds: Suppose that the following three conditions are satisfied:*

1. *For each* $r_i \in \text{supp}(\mu^i)$, *the frequency of* r_i *is close to* $\mu^i(r_i)$*:*

$$\left| f_i(r_i|b) - \mu^i(r_i) \right| \leq \varepsilon_{\text{likelihood}}. \tag{48}$$

2. *For each* $r_i \in \text{supp}(\mu^i)$ *and* $r_{-i} \in \text{supp}\left(\mu_{-i}|_{r_i}\right)$, *we have*

$$\left| \sum_{y_i} f_i(y_i|r_i, b) \frac{q(y_i|r_{-i}, r_i) - q(y_i|\mu_{-i}|_{r_i}, r_i)}{q(y_i|\mu_{-i}|_{r_i}, r_i)} \right| \leq \varepsilon_{\text{likelihood}}. \tag{49}$$

3. $Z^i_{\varepsilon_{\text{affine}}}(b, h^b_m) = 0$.

Then, player i *believes that the expected value of the summation of* $x(r_t, y_t|\mu)$ *is sufficiently small:*

$$\Pr\left(\left\{ \sum_{t \in T_b} \lambda \cdot x(r_t, y_t|\mu) > \frac{\varepsilon}{4} \frac{T}{B} \right\} |\alpha(b) = \mu, h^b_i \right) \leq \exp\left(-\left(\frac{\varepsilon}{8\bar{x}}\right)^2 \frac{T}{B} \right),$$

where

$$\bar{x} = N \max_{i, r \in A, y \in Y} \left\{ \max\left\{ |x_i(r, y|\mu)|, |x_i(r, m|\tau)| \right\} \right\}. \tag{50}$$

Proof. By [Condition Z distance] for $Z^i_{\varepsilon_{\text{affine}}}(b, h^b_m)$, the fact that $Z^i_{\varepsilon_{\text{affine}}}(b, h^b_m) = 0$ implies

$$d\left(f_i(Y_i|r_i, b), H_i(Y_i|r_i)\right) \leq \varepsilon_{\text{affine}}.$$

By (46), there exist $(t(r_{-i}))_{r_{-i} \in \text{supp}(\mu_{-i}|_{r_i})} \in \mathbb{R}^{\left|\text{supp}(\mu_{-i}|_{r_i})\right|}$ and $e \in \mathbb{R}^{|Y_i|}$ such that

$$\begin{cases} f_i(Y_i|r_i, b) = q(Y_i|\mu_{-i}|_{r_i}, r_i) + \sum_{r_{-i} \in \text{supp}(\mu_{-i}|_{r_i})} t(r_{-i}) \left\{ q(Y_i|r_{-i}, r_i) - q(Y_i|\mu_{-i}|_{r_i}, r_i) \right\} + e, \\ \|e\| \leq \varepsilon_{\text{affine}}. \end{cases} \tag{51}$$

Substituting this expression of $f_i(Y_i|r_i, b)$ into (49), for sufficiently small $\varepsilon_{\text{affine}} > 0$ compared to $\varepsilon_{\text{likelihood}} > 0$, we have

$$\left|\sum_{y_i} \sum_{r_{-i} \in \text{supp}(\mu_{-i}|_{r_i})} t(r_{-i}) \left\{ q(Y_i|r_{-i}, r_i) - q(Y_i|\mu_{-i}|_{r_i}, r_i) \right\} \frac{q(y_i|r_{-i}, r_i) - q(y_i|\mu_{-i}|_{r_i}, r_i)}{q(y_i|\mu_{-i}|_{r_i}, r_i)}\right| \leq 2\varepsilon_{\text{likelihood}}.$$

Multiplying both sides with $t(r_{-i})$ and sum them up with respect to $r_{-i} \in \text{supp}(\mu_{-i}|_{r_i})$ yields

$$\begin{aligned} & \left|\sum_{y_i} \sum_{r_{-i} \in \text{supp}(\mu_{-i}|_{r_i})} \frac{\{t(r_{-i})\}^2 \left\{ q(y_i|r_{-i}, r_i) - q(y_i|\mu_{-i}|_{r_i}, r_i) \right\}^2}{q(y_i|\mu_{-i}|_{r_i}, r_i)}\right| \\ \leq \; & 2\varepsilon_{\text{likelihood}} \sum_{r_{-i} \in \text{supp}(\mu_{-i}|_{r_i})} |t(r_{-i})| \,. \end{aligned}$$

Since $\{t(r_{-i})\}_{r_{-i}}$ satisfying (51) is uniformly bounded, for sufficiently small $\varepsilon_{\text{likelihood}} > 0$, this means that

$$\sum_{r_{-i} \in \text{supp}(\mu_{-i}|_{r_i})} t(r_{-i}) \left\{ q(Y_i|r_{-i}, r_i) - q(Y_i|\mu_{-i}|_{r_i}, r_i) \right\}$$

is sufficiently close to the origin.

Since $\mathbb{E}\left[\lambda \cdot x(r_t, y_t|\mu)|\mu\right] \leq 0$ by (34), together with (48), for sufficiently small $\varepsilon_{\text{likelihood}} > 0$, we have

$$\mathbb{E}\left[\sum_{t \in T_b} \lambda \cdot x(r_t, y_t|\mu)|\alpha(b) = \mu, h_i^b\right] \leq \frac{\varepsilon}{8}\frac{T}{B}.$$

Hence, by Hoeffding's inequality, we have

$$\Pr\left(\left\{\sum_{t \in T_b} \lambda \cdot x(r_t, y_t|\mu) > \frac{\varepsilon}{4}\frac{T}{B}\right\} |\alpha(b) = \mu, h_i^b\right) \leq \exp\left(-\left(\frac{\varepsilon}{8\bar{x}}\right)^2 \frac{T}{B}\right),$$

as desired. ■

From now on, we fix $\varepsilon_{\text{likelihood}} > 0$ and $\varepsilon_{\text{affine}} > 0$ so that Lemma 9 holds.

Given $\varepsilon_{\text{likelihood}} > 0$, $\varepsilon_{\text{affine}} > 0$, and $Z^i_{\varepsilon_{\text{affine}}}(b, h_m^b)$, we define the state transition. As noted, $s(b) = B$ implies that the self generation in Section 7.3.1 has been an issue. As will be seen in the definition of the continuation payoff below, this becomes an issue only if all of the following four conditions are satisfied:

1. [Regular recommendation] The mediator recommends μ: $\alpha(b) = \mu$.

2. [Regular realization of r_i] For each player $i \in I$ and $r_i \in \text{supp}(\mu_i)$, the frequency of r_i is regular:

$$\max_{r_i \in \text{supp}(\mu_i)} |f_i(r_i|b) - \mu_i(r_i)| \leq \varepsilon_{\text{likelihood}}. \tag{52}$$

3. [Regular realization of r_{-i}] For each player $i \in I$, $r_i \in \text{supp}(\mu_i)$, and $r_{-i} \in \text{supp}\left(\mu_{-i}|_{r_i}\right)$, the frequency of r_{-i} given r_i is regular:

$$\max_{r_i \in \text{supp}(\mu_i), r_{-i} \in \text{supp}\left(\mu_{-i}|_{r_i}\right)} |f_{-i}(r_{-i}|r_i, b) - \mu(r_{-i}|r_i, b)| \leq t_{\text{likelihood}} \varepsilon_{\text{likelihood}}. \tag{53}$$

4. [Regular affine hull of $-i$] For each player $i \in I$, we have $Z^i_{\varepsilon_{\text{affine}}}(b, h^b_m) = 0$.

Hence, the state transits from $s(b) = G$ to $s(b+1) = B$ if and only if all of the four conditions are satisfied and the summation of $\lambda \cdot x(r_t, m_t|\mu)$ is erroneously high (that is, self generation is an issue):

$$\sum_{t \in T(b)} \lambda \cdot x(r_t, m_t|\mu) \geq \frac{\varepsilon}{4} \frac{T}{B}. \tag{54}$$

Continuation Payoff $w(h^{T+1}_m, \delta)$ As noted, if one of the four conditions ([Regular recommendation], [Regular realization of r_i], [Regular realization of r_{-i}], and [Regular affine hull of $-i$]) is violated, then we will make sure that self generation is not an issue. To this end, take $\bar{x} \in \mathbb{R}^N$ sufficiently large such that

$$\min_{i \in \text{supp}(\lambda)} (\lambda_i)^2 \cdot \bar{x} \geq \max_{r \in A, y \in Y} \lambda \cdot x(r, m|\mu).$$

For each $i \in I$, we define an $|I|$-dimensional vector as follows:

$$\bar{x}^i(r, m|\mu) \equiv x(r, m|\mu) - \bar{x}\lambda_i e_i$$

for all $r \in A$ and $m \in Y$. Note that, for each $i \in I$, we have

$$\lambda \cdot \bar{x}^i(r, m|\mu) \leq 0 \tag{55}$$

for all $r \in A$ and $m \in Y$.

In addition, for some arbitrarily fixed $\bar{\imath} \in \text{supp}(\lambda)$, we define

$$\bar{x}(r, m|\mu) = \bar{x}^{\bar{\imath}}(r, m|\mu)$$

so that self generation is not an issue:

$$\lambda \cdot \bar{x}(r, m|\mu) \leq 0 \tag{56}$$

for all $r \in A$ and $m \in Y$.

Since $\bar{x}$ is constant, strict incentive compatibility holds with $\bar{x}^i(r, m|\mu)$ and $\bar{x}(r, m|\mu)$.

In each block b, the movement of the continuation payoff, denoted by $x(h_m^b, b, \delta)$, is defined as follows.

If $s(b) = G$, that is, if self generation constraint has not been an issue, then $x(h_m^b, b, \delta)$ depends on which of the following two cases is true:

1. If [Regular recommendation] is not the case, [Regular recommendation of r_i] is not the case for some $i \in I$ and $r_i \in \text{supp}(\mu_i)$, or [Regular recommendation of r_{-i}] is not the case for some $i \in I$, $r_i \in \text{supp}(\mu_i)$, and $r_{-i} \in \text{supp}\left(\mu_{-i}|_{r_i}\right)$, then we define

$$x(h_m^b, b, \delta) = \begin{cases} \sum_{t \in T_b} \delta^{t-1} \bar{x}(r_t, m_t|\mu) & \text{if } \alpha(b) = \mu, \\ \sum_{t \in T_b} \delta^{t-1} x(r_t, m_t|\tau) & \text{otherwise.} \end{cases}$$

(41) and (56) make sure that $\lambda \cdot x(h_m^b, b, \delta) \leq 0$ for all h_m^b. Since the distribution of the recommendations is out of player i's control, subtracting $\lambda_i \bar{x}$ from $x(r_t, m_t|\mu)$ after irregular recommendations does not affect player i's incentive.

2. If [Regular affine of $-i$] is not the case for some player i, then take $n \in \text{supp}(\lambda)$ with $n \neq i$ and define

$$x(h_m^b, b, \delta) = \begin{cases} \sum_{t \in T_b} \delta^{t-1} \bar{x}^n(r_t, m_t|\mu) & \text{if } \alpha(b) = \mu, \\ \sum_{t \in T_b} \delta^{t-1} x(r_t, m_t|\tau) & \text{otherwise.} \end{cases}$$

(41) and (55) make sure that $\lambda \cdot x(h_m^b, b, \delta) \leq 0$ for all h_m^b. Note that we subtract $\lambda_n \bar{x}$ for player n who is not player i. Since the distribution of $Z^i_{\varepsilon_{\text{affine}}}(b, h_m^b)$ is independent of players $-i$'s actions, this does not affect player n's incentive.

3. Otherwise, the mediator recommends μ by [Regular recommendation]. The mediator gives the reward, which is the summation of $x(r, m|\mu)$:

$$x(h_m^b, b, \delta) = \sum_{t \in T_b} \delta^{t-1} x(r_t, m_t|\mu).$$

If $s(b) = B$, that is, if self generation constraint has been an issue, then we make sure that $\lambda \cdot x(h_m^b, b, \delta)$ is not a large positive number:

$$x(h_m^b, b, \delta) = \begin{cases} \sum_{t \in T_b} \delta^{t-1} x^{\text{pairwise}}(r_t, m_t|\tau) & \text{if } \lambda \cdot \sum_{t \in T_b} \delta^{t-1} x^{\text{pairwise}}(r_t, m_t|\tau) \leq \frac{\varepsilon}{4} \frac{T}{B}, \\ 0 & \text{otherwise.} \end{cases} \tag{57}$$

Given $x(h_m^b, b, \delta)$, we define the continuation payoff as follows:

$$w(h_m^{T+1}, \delta) = v + \delta^{-T}(1 - \delta) \left\{ \sum_{b=1}^{B} x(h_m^b, b, \delta) - \frac{1}{2} \lambda \varepsilon T \right\}, \tag{58}$$

where

$$v = \frac{1 - \delta}{1 - \delta^T} \mathbb{E} \left[\sum_{t=1}^{T} \delta^{t-1} u(a_t) + \sum_{b=1}^{B} x(h_m^b, b, \delta) - \frac{1}{2} \lambda \varepsilon T | \sigma_m^T \right]. \tag{59}$$

7.3.3 Promise Keeping

We now make sure that the promise keeping in Section 7.3.1 is satisfied:

$$\lambda \cdot \mathbb{E}\left[(1-\delta)\sum_{t=1}^{T}\delta^{t-1}u(a_t) + \delta^T w(h_m^{T+1},\delta)|\sigma_m^T\right] \geq \max_{v'\in \underline{Q}} \lambda \cdot v' - \varepsilon.$$

By substituting (58), this is equivalent to

$$\begin{aligned} & \lambda \cdot \mathbb{E}\left[(1-\delta)\sum_{t=1}^{T}\delta^{t-1}u(a_t) + \delta^T v + (1-\delta)\left\{\sum_{b=1}^{B} x(h_m^b, b, \delta) - \frac{1}{2}\lambda\varepsilon T\right\}|\sigma_m^T\right] \\ \geq \ & \max_{v'\in \underline{Q}} \lambda \cdot v' - \varepsilon. \end{aligned}$$

By (59), this is again equivalent to

$$\lambda \cdot \frac{1-\delta}{1-\delta^T}\mathbb{E}\left[\sum_{t=1}^{T}\delta^{t-1}u(a_t) + \sum_{b=1}^{B} x(h_m^b, b, \delta) - \frac{1}{2}\lambda\varepsilon T|\sigma_m^T\right] \geq \max_{v'\in \underline{Q}} \lambda \cdot v' - \varepsilon.$$

By the central limit theorem, on equilibrium path, $s(b+1) = B$ happens with a small probability. Since [Regular recommendation] implies that $s(b+1) = B$ only if $\alpha(b) = \mu$, we have

$$\begin{aligned} & \Pr\left(\{s(b+1) = B\}\,|s(b) = G\right) \\ \leq \ & \Pr\left(\{s(b+1) = B\}\,|\alpha(b) = \mu\right) \\ \leq \ & \Pr\left(\left\{\sum_{t\in T(b)} \lambda \cdot x(r_t, m_t|\mu) \geq \frac{\varepsilon}{4}\frac{T}{B}\right\}|\alpha(b) = \mu\right). \end{aligned}$$

By (**??**), we have

$$\mathbb{E}\left[\lambda \cdot x(r_t, m_t|\mu)|\mu\right] \leq 0.$$

Hence, by Hoeffding's inequality, we have

$$\begin{aligned}
& \Pr\left(\left\{\sum_{t\in T(b)}\lambda\cdot x(r_t,m_t)\geq\frac{\varepsilon}{4}\frac{T}{B}\right\}|\alpha(b)=\mu\right) \\
\leq\ & \exp\left(-\left(\frac{\varepsilon}{4\max_{r\in A,m\in Y}\lambda\cdot x(r,m|\mu)}\right)^2\frac{T}{B}\right).
\end{aligned}$$

Similarly, given $s(b)=G$ and $\alpha(b)=\mu$, the probability that [Regular realization of r_i], [Regular realization of r_{-i}], and [Regular affine hull of $-i$] are satisfied for all $i\in I$, $r_i\in\text{supp}\,(\mu_i)$, and $r_{-i}\in\text{supp}\left(\mu_{-i}|_{r_i}\right)$ with probability no less than

$$1-N\,|A|\left\{\exp\left(-\varepsilon^2_{\text{likelihood}}\frac{T}{B}\right)+\exp\left(-t^2_{\text{likelihood}}\varepsilon^2_{\text{likelihood}}\frac{T}{B}\right)+\exp\left(-\frac{\varepsilon^2_{\text{affine}}}{K_{\text{affine}}}\frac{T}{B}\right)\right\}.$$

Hence, for sufficiently large T, it suffices to show that

$$\frac{1-\delta}{1-\delta^T}\lambda\cdot\mathbb{E}\left[\sum_{t=1}^{T}\delta^{t-1}u(a_t)+\sum_{b=1}^{B}x(h^b_m,b,\delta)-\frac{1}{2}\lambda\varepsilon T|\mathcal{G}\right]\geq\max_{v'\in\underline{Q}}\lambda\cdot v'-\frac{4}{5}\varepsilon,$$

where $\mathcal{G}$ is the event that (i) $s(b)=G$ for all b and (ii) when $\alpha(b)=\mu$, then [Regular realization of r_i], [Regular realization of r_{-i}], and [Regular affine hull of $-i$] are satisfied for all $i\in I$, $r_i\in\text{supp}\,(\mu_i)$, and $r_{-i}\in\text{supp}\left(\mu_{-i}|_{r_i}\right)$.

Taking the limit where δ goes to one, it suffices to show that

$$\frac{1}{T}\lambda\cdot\mathbb{E}\left[\sum_{t=1}^{T}u(a_t)+\sum_{b=1}^{B}x(h^b_m,b,1)-\frac{1}{2}\lambda\varepsilon T|\mathcal{G}\right]\geq\max_{v'\in\underline{Q}}\lambda\cdot v'-\frac{3}{4}\varepsilon.$$

This is equivalent to

$$\frac{1}{T}\lambda\cdot\mathbb{E}\left[\sum_{t=1}^{T}u(a_t)+\sum_{b=1}^{B}x(h^b_m,b,1)|\mathcal{G}\right]\geq\max_{v'\in\underline{Q}}\lambda\cdot v'-\frac{1}{4}\varepsilon. \tag{60}$$

Note that we have

$$\begin{aligned} \frac{1}{T/B}\lambda\cdot\mathbb{E}\left[\sum_{t\in T(b)}\left(u(a_t)+x(r_t,m_t|\mu)\right)|\mu\right] &= v^{\mu}, \\ \frac{1}{T/B}\lambda\cdot\mathbb{E}\left[\sum_{t\in T(b)}\left(u(a_t)+x(r_t,m_t|\tau)\right)|\tau\right] &= v^{\tau} \end{aligned}$$

by definition (**??**). Since $\alpha(b)=\tau$ happens with probability η in each b with $s(b)=G$, we have

$$\frac{1}{T}\lambda\cdot\mathbb{E}\left[\sum_{t=1}^{T}u(a_t)+\sum_{b=1}^{B}x(h_m^b,b,1)|\mathcal{G}\right]\geq(1-B\eta)\lambda\cdot v^{\mu}+B\eta\min\left\{\lambda\cdot v^{\mu},\lambda\cdot v^{\tau}\right\}.$$

By (43), this is greater than $\max_{v'\in\underline{Q}}\lambda\cdot v'-\frac{1}{4}\varepsilon$, as desired from (60).

7.3.4 Self Generation

Let us now verify self generation. By definition of (58), it suffices to show that

$$\lambda\cdot\sum_{b=1}^{B}x(h_m^b,b,\delta)\leq\frac{1}{2}\lambda\varepsilon T.$$

Note that

$$\lambda\cdot x(h_m^b,b,\delta)>\frac{\varepsilon}{4}\frac{T}{B}$$

only in block b with $s(b)=G$. In addition, once $\lambda\cdot x(h_m^b,b,\delta)>\frac{\varepsilon}{4}\frac{T}{B}$ happens in block b, from block $b+1$ on, we have $s(\tilde{b})=B$ for all $\tilde{b}\geq b+1$. Moreover, with $s(\tilde{b})=B$, we have

$$\lambda\cdot x(h_m^b,\tilde{b},\delta)\leq\frac{\varepsilon}{4}\frac{T}{B}.$$

Hence, the maximum realization of $\lambda\cdot\sum_{b=1}^{B}x(h_m^b,b,\delta)$ is

$$\frac{\varepsilon}{4}\frac{T}{B}(B-1)+\frac{\max_{r\in A,y\in Y}\lambda\cdot x(r,y|\mu)T}{B}\leq\frac{\varepsilon}{2}T$$

by (42). Hence, we have

$$\lambda \cdot w(h_m^{T+1}, \delta) < 0$$

as desired.

7.3.5 Incentive Compatibility

Since we have verified [FL: promise keeping] and [FL: self generation] in Section 7.3.1, we are left to verify player i's incentive to follow the recommendation.

Each player i maximizes

$$\mathbb{E}\left[(1-\delta)\sum_{t=1}^{T}\delta^{t-1}u_i(a_t) + \delta^T w(h_m^{T+1}, \delta)|\sigma_i^T, \sigma_m^T\right].$$

Substituting (58) and ignoring a constant, player i maximizes

$$\mathbb{E}\left[\sum_{t=1}^{T}\delta^{t-1}u_i(a_t) + \sum_{b=1}^{B} x_i(h_m^b, b, \delta)|\sigma_i^T, \sigma_m^T\right].$$

Since the continuation strategy of the mediator from block b only depends on $s(b)$, let

$$v_i(s(b), b, \delta) = \max_{\sigma_i^T}\mathbb{E}\left[\sum_{t\in T(b)}\delta^{t-1}u_i(a_t) + x_i(h_m^b, b, \delta)|\sigma_i^T, \sigma_m^T, s(b)\right]$$

be player i's value from block b.

To show the optimality of the faithful strategy, it is useful to obtain the upper bound of player i's value $v_i(B, b, \delta)$ when the mediator is in state $s(b) = B$. Since the state will be $s(\tilde{b}) = B$ for all $\tilde{b} \geq b$, if player i knows that $s(b) = B$, then she maximizes the value from block b.

Note that (39) guarantees that

$$\mathbb{E}\left[\lambda \cdot x^{\text{punish}}(r, m|\tau)|\sigma_i, \tau\right] \leq 0$$

for each strategy $\sigma_i \in \Sigma_i$. Hence, with *each strategy of player* i, the ex ante probability at the beginning of block b that

$$\lambda \cdot \sum_{t \in T_b} \delta^{t-1} x^{\text{punish}}(r_t, m_t | \tau) > \frac{\varepsilon}{4} \frac{T}{B}$$

happens is with probability no more than

$$\exp\left(-\left(\frac{\varepsilon}{4\bar{x}}\right)^2 \frac{T}{B}\right)$$

by (50) and Hoeffding's inequality.

Hence, by (57), we have

$$\begin{aligned}
v_i(B,b,\delta) \;\; &\leq \;\; \left\{1 - \exp\left(-\left(\frac{\varepsilon}{4\bar{x}}\right)^2 \frac{T}{B}\right)\right\} \max_{\sigma_i} \mathbb{E}\left[\sum_{t \in T(b)} \delta^{t-1} \left\{u_i(a_t) + x_i^{\text{punish}}(r_t, m_t | \tau)\right\} | \sigma_i^T, \tau\right] \\
&\quad + \exp\left(-\left(\frac{\varepsilon}{4\bar{x}}\right)^2 \frac{T}{B}\right) \max_a u_i(a) \\
&= \;\; \left\{1 - \exp\left(-\left(\frac{\varepsilon}{4\bar{x}}\right)^2 \frac{T}{B}\right)\right\} \sum_{t \in T(b)} \delta^{t-1} \max_{\sigma_i} \mathbb{E}\left[u_i(a_i, \tau_{-i}) + x_i^{\text{punish}}(r, m_i, y_{-i}) | \sigma_i, \tau\right] \\
&\quad + \exp\left(-\left(\frac{\varepsilon}{4\bar{x}}\right)^2 \frac{T}{B}\right) \max_a u_i(a) \\
&\leq \;\; \left\{1 - \exp\left(-\left(\frac{\varepsilon}{4\bar{x}}\right)^2 \frac{T}{B}\right)\right\} \sum_{t \in T(b)} \delta^{t-1} v_i^{\text{punish}} + \exp\left(-\left(\frac{\varepsilon}{4\bar{x}}\right)^2 \frac{T}{B}\right) \max_a u_i(a)
\end{aligned}$$

by (38). For sufficiently large T, the last expression converges to

$$\sum_{t \in T(b)} \delta^{t-1} v_i^{\text{punish}},$$

and for sufficiently large δ, this expression in turn converges to $T v_i^{\text{punish}}$.

Therefore, by (44), for sufficiently large T and for sufficiently large δ, we have

$$v_i(B,b,\delta) < T\left\{(1 - B\eta) v_i^{\mu} + B\eta \min\left\{v_i^{\mu}, v_i^{\tau}\right\}\right\}. \tag{61}$$

Next, we verify that, on equilibrium path, player i believes that $s(b+1) = G$ with a high probability. Specifically, let $h_i^{\leq b} = ((r_{i,t}, y_{i,t})_{t \in T(\tilde{b})})_{\tilde{b} \leq b}$ be player i's on-path (faithful) history at the end of block b.[3] We will show that, for each $b = 1, ..., B-1$ and each on-path history $h_i^{\leq b}$, we have

$$\Pr\left(s(b+1) = G | h_i^{\leq b}, (\alpha(\tilde{b}))_{\tilde{b}=1}^{b+1}\right) \geq 1 - \frac{B}{\eta^B} \exp(-eT)$$

with

$$e = \frac{1}{B}\left((t_{\text{likelihood}} \varepsilon_{\text{likelihood}})^2 + \left(\frac{\varepsilon}{8\bar{x}}\right)^2\right).$$

Note that we also allow player i to notice what recommendation schedule $\alpha(\tilde{b})$ the mediator uses from block 1 to block $b+1$.

It suffices to show that player i believes that $s(b+1) = G$ with a high probability, conditional on $s(b) = G$:

$$\Pr\left(s(b+1) = G | s(b) = G, h_i^{\leq b}, (\alpha(\tilde{b}))_{\tilde{b}=1}^{b+1}\right) \geq 1 - \frac{1}{\eta^B} \exp(-eT).$$

In addition, since any sequence of $(\alpha(\tilde{b}))_{\tilde{b}=1}^{b+1}$ can happen with probability no less than η^B if $s(b+1) = G$, it suffices to show that

$$\Pr\left(s(b+1) = G | s(b) = G, \alpha(b), h_i^{\leq b}\right) \geq 1 - \exp(-eT).$$

To infer the next state given $s(b) = G$ and $\alpha(b)$, only the history in the current block is informative:

$$\begin{aligned} & \Pr\left(s(b+1) = G | s(b) = G, \alpha(b), h_i^{\leq b}\right) \\ = \ & \Pr\left(s(b+1) = G | s(b) = G, \alpha(b), h_i^{b}\right). \end{aligned}$$

Further, if $s(b) = G$ and $\alpha(b) = \tau$, then [Regular recommendation] does not hold and $s(b+1) = G$ for sure.

[3] Since the history is faithful, we omit $a_{i,t} = r_{i,t}$ and $m_{i,t} = y_{i,t}$.

Therefore, in total, it suffices to show that

$$\Pr\left(s(b+1)=G|s(b)=G,\mu,h_i^b\right)\geq 1-\exp(-eT).$$

We classify player i's history h_i^b into the following four categories:

1. Player i's recommended actions are irregular:

$$\max_{r_i\in\text{supp}(\mu_i)}\left|f_i(r_i|b)-\mu^i(r_i)\right|>\varepsilon_{\text{likelihood}}.$$

2. Player i believes that players $-i$'s recommended actions are irregular:

$$\max_{r_i\in\text{supp}(\mu_i),r_{-i}\in\text{supp}\left(\mu_{-i}|_{r_i}\right)}\left|\sum_{y_i}f_i(y_i|r_i,b)\frac{q(y_i|r_{-i},r_i)-q(y_i|\mu_{-i}|_{r_i},r_i)}{q(y_i|\mu_{-i}|_{r_i},r_i)}\right|>\varepsilon_{\text{likelihood}}.$$

3. $Z^i_{\varepsilon_{\text{affine}}}(b,h_m^b)=1$.

4. None of the above three conditions is satisfied.

If Condition 1 is satisfied, then [Regular realization of r_i] is not satisfied for some $r_i\in\text{supp}\left(\mu_i\right)$ and $s(b+1)=G$ for sure.

If Condition 2 is satisfied, then by Lemma 8, player i believes that [Regular realization of r_{-i}] is not satisfied for some $r_i\in\text{supp}\left(\mu^i\right)$ and $r_{-i}\in\text{supp}\left(\mu_{-i}|_{r_i}\right)$ and so $s(b+1)=G$ with probability $1-\exp\left(-\left(t_{\text{likelihood}}\varepsilon_{\text{likelihood}}\right)^2\frac{T}{B}\right)$.

If Condition 3 is satisfied, then [Regular affine hull of $-i$] is not satisfied and so $s(b+1)=G$ for sure.

If Condition 4 is satisfied, then by Lemma 9, player i believes that

$$\lambda\cdot\sum_{t\in T_b}x(r_t,y_t|\mu)\leq\frac{\varepsilon}{4}\frac{T}{B}$$

and $s(b+1)=G$ with probability $1-\exp\left(-\left(\frac{\varepsilon}{8\bar{x}}\right)^2\frac{T}{B}\right)$.

Therefore, in total, we have

$$\Pr\left(s(b+1) = G | s(b) = G, \alpha(b), h_i^b\right) \geq 1 - \exp(-eT)$$

with

$$e = \frac{1}{B}\left(\left(t_{\text{likelihood}}\varepsilon_{\text{likelihood}}\right)^2 + \left(\frac{\varepsilon}{8\bar{x}}\right)^2\right),$$

as desired.

Therefore, the following two claims are true: For sufficiently large T and sufficiently large δ,

1. [On-path belief] For each $b = 1, ..., B-1$, for each on-path history $h_i^{\leq b}$, player i believes that the next state is G:

$$\Pr\left(s(b+1) = G | h_i^{\leq b}, (\alpha(\tilde{b}))_{\tilde{b}=1}^{b+1}\right) \geq 1 - \frac{B}{\eta^B}\exp(-eT).$$

2. [On-path value] By (61), the value in the good state is better than the value in the bad state:

$$v_i(s(b), G, \delta) > v_i(s(b), B, \delta).$$

Let us now consider player i's incentive to follow the recommendation and tell the truth about the signal observations on equilibrium path by backward induction:

The last block B As seen in [On-path belief], player i believes that $s(B) = G$ with probability no less than $1 - \frac{B}{\eta^B}\exp(-eT)$.

Suppose that $\alpha(b) = \mu$. If $s(B) = G$ with probability one, for each r_i, by (32), player i has the strict incentive not to deviate so that the distribution of m is changed. Hence, for sufficiently large T, by [On-path belief], on equilibrium path, player i has the strict incentive not to deviate so that the distribution of m is changed.

On the other hand, if the distribution of m is the same between the faithful strategy and a deviation s_i, then the distribution of $x(r, m|\mu)$ is also the same. Hence, (31) implies that the faithful strategy maximizes the instantaneous utility. Hence, player i has the incentive to be faithful.

The same discussion for $\alpha(b) = \tau$ with $x(r, m|\mu)$ replaced with $x(r, m|\tau)$ establishes the incentive compatibility.

Block b given the incentive compatibility for blocks $b+1, ..., B$ Suppose that player i is faithful from block $b+1$ on on the equilibrium path. Fixing $s(b+1)$, the same discussion for block B establishes the result.

Hence, we are left to show that player i does not have an incentive to deviate to change the distribution of $s(b+1)$. From [On-path value], player i does not want to induce $s(b+1) = B$. If $Z^i_{\varepsilon_{\text{affiie}}}(b, h^b_m) = 1$ with truthtelling, then $s(b+1) = s(b)$ with the faithful strategy and player i does not have an incentive to deviate.

Suppose $Z^i_{\varepsilon_{\text{affiie}}}(b, h^b_m) = 0$ with truthtelling. From [On-path belief], with the faithful strategy, player i believes that $s(b+1) = G$ with probability no less than $1 - \frac{B}{\eta^B}\exp(-eT)$. To change the distribution of $Z^i_{\varepsilon_{\text{affiie}}}(b, h^b_m)$, player i has to deviate in a way that changes the distribution of m. By (32) and (**??**), for sufficiently large T, it is optimal for player i to be faithful.

Therefore, it is optimal for player i to be faithful, as desired.

7.3.6 Proof of Lemma 1

We first consider the case with $\text{supp}(\lambda) \geq 2$. From Proposition 1, there exists $\mu \in \mathcal{M}$ with $k(\lambda) = \lambda \cdot u(\mu)$. We first perturb μ to $\mu^{\text{full}} = (1-\eta)\mu + \eta \sum_{a \in A} \frac{\eta}{|A|}$ so that μ^{full} has full support. Note that, for each $\varepsilon > 0$, for sufficiently small $\eta > 0$, we have

$$\lambda \cdot u(\mu^{\text{full}}) > k(\lambda) - \frac{\varepsilon}{2}. \tag{62}$$

We want to construct $\hat{\mu}^{\text{full}}$ satisfying $\underline{k}(\hat{\mu}^{\text{full}}, \lambda) \geq k(\lambda) - \varepsilon$ from μ^{full}. To this end, for each i, let $\Sigma_i^{\text{undetectable}}$ be the set of undetectable deviations:

$$\Sigma_i^{\text{undetectable}} \equiv \left\{\sigma_i \in \Sigma_i : \Pr(m|\sigma_i, r) = \Pr(m|r) \text{ for all } r \in A = \text{supp}(\mu^{\text{full}})\right\}.$$

Note that $\Sigma_i^{\text{undetectable}}$ includes the faithful strategy.

Since (i) all supp(μ)-detectable deviations are supp(μ^{full})-detectable and (ii) μ^{full} and μ are close to each other, for each $\varepsilon > 0$, sufficiently small $\eta > 0$, by (5), we have

$$\max_{i \in I} \left| u_i(\mu) - \max_{\sigma_i \in \Sigma_i^{\text{undetectable}}} u_i(\sigma_i, \mu) \right| < \frac{\varepsilon}{2}. \tag{63}$$

An undetectable strategy $\sigma_i \in \Sigma_i^{\text{undetectable}}$ has two important properties. First, since the realized payoffs are observable, $\Pr(m|\sigma_i, r) = \Pr(m|r)$ for all $r \in A$ implies that the distribution of players $-i$'s realized payoffs is the same between σ_i and player i's faithful strategy for each $r \in A$. Hence, for each i and $\sigma_i \in \Sigma_i^{\text{undetectable}}$, we have

$$u_j(\sigma_i|_{r_i}, r_{-i}) = u_j(r) \tag{64}$$

for each $j \in -i$ and $r \in A$.

Second, suppose that some player j's deviation σ_j is detectable: There exist $\bar{r} \in A$ and $\bar{m} \in Y$ such that $\Pr(\bar{m}|\sigma_j, \bar{r}) \neq \Pr(\bar{m}|\bar{r})$. Since player i's supp(μ^{full})-undetectable deviation $\sigma_i \in \Sigma_i^{\text{undetectable}}$ gives the same distribution of m for each r, $\sigma_i|_{\bar{r}_i}$ can detect player j's deviation as well as the faithful strategy for $\bar{r}_i$:

$$\Pr(\bar{m}|\sigma_j, \bar{r}_{-i}, \sigma_i|_{\bar{r}_i}) = \Pr(\bar{m}|\sigma_j, \bar{r}) \neq \Pr(\bar{m}|\bar{r}) = \Pr(\bar{m}|\bar{r}, \sigma_i|_{\bar{r}_i}). \tag{65}$$

Moreover, instead of $\sigma_i \in \Sigma_i^{\text{undetectable}}$, consider σ_i^{truth} such that player i takes the same mixture of actions as σ_i but tells the truth about y_i: $\sigma_i^{\text{truth}}(r_i)(a_i) = \sigma_i(r_i)(a_i)$ for each $r_i, a_i \in A_i$ and $\sigma_i^{\text{truth}}(r_i, a_i, y_i)(m_i) = 1$ if and only if $m_i = y_i$. This truthful strategy σ_i^{truth}

can detect player j's deviation: There exists $\hat{m}_i \in Y_i$ such that

$$\Pr(\hat{m}_i, \bar{m}_{-i}|\sigma_j, \bar{r}_{-i}, \sigma_i^{\text{truth}}|_{\bar{r}_i}) \neq \Pr(\hat{m}_i, \bar{m}_{-i}|\bar{r}, \sigma_i^{\text{truth}}|_{\bar{r}_i}). \tag{66}$$

To see why, suppose that $\Pr(\hat{m}_i, \bar{m}_{-i}|\sigma_j, \bar{r}_{-i}, \sigma_i^{\text{truth}}|_{\bar{r}_i}) = \Pr(\hat{m}_i, \bar{m}_{-i}|\bar{r}, \sigma_i^{\text{truth}}|_{\bar{r}_i})$ for all $\hat{m}_i \in Y_i$. Note that the distribution of $\bar{m}_i$ given σ_i can be written as

$$\sum_{\hat{m}_i} \sigma_i(r_i, a_i, \hat{m}_i)(m_i) \Pr(\hat{m}_i, \bar{m}_{-i}|\sigma_j, \bar{r}_{-i}, \sigma_i^{\text{truth}}|_{\bar{r}_i})$$

since $\sigma_i(r_i, a_i, \hat{m}_i)(m_i)$ is the probability that σ_i sends m_i after receiving r_i, taking a_i, and observing $\hat{m}_i$. Hence, we would have

$$\Pr(\bar{m}|\sigma_j, \bar{r}_{-i}, \sigma_i|_{\bar{r}_i}) = \Pr(\bar{m}|\bar{r}, \sigma_i|_{\bar{r}_i}),$$

which contradicts to (65).

We find $\bar{\sigma}_i \in \Sigma_i^{\text{undetectable}}$ which is a static best response to μ^{full}:

$$\bar{\sigma}_i \in \arg \max_{\sigma_i \in \Sigma_i^{\text{undetectable}}} u_i(\sigma_i, \mu^{\text{full}}).$$

For each $r_i \in A_i$, let $\bar{\sigma}_i|_{r_i}$ denote player i's continuation strategy $\sigma_i|_{r_i}$ after receiving r_i. By (64), this implies that

$$\bar{\sigma}_i \in \arg \max_{\sigma_i \in \Sigma_i^{\text{undetectable}}} u_i(\sigma_i, (\bar{\sigma}_n)_{n \in -i}) \tag{67}$$

for each $i \in I$.

Without loss, we assume that each pure strategy $s_i|_{r_i}$ in the support of $\bar{\sigma}_i|_{r_i}$ generates an extreme distribution of the messages: For each $s_i|r_i \in \text{supp}\,(\bar{\sigma}_i|_{r_i})$, there does not exist a set of pure continuation strategies after receiving r_i, denoted by $\tilde{S}_i|_{r_i}$, and a distribution over $\tilde{S}_i|_{r_i}$, denoted by $\phi \in \Delta(\tilde{S}_i|_{r_i})$ with $\phi(\tilde{s}_i|_{r_i}) > 0$ for all $\tilde{s}_i|_{r_i}$, such that

$$\begin{cases} \Pr(\bar{m}|s_i|_{r_i}, \bar{r}_{-i}) \neq \Pr(\bar{m}|\tilde{s}_i|_{r_i}, \bar{r}_{-i}) \text{ for some } \tilde{s}_i|_{r_i} \in \tilde{S}_i|_{r_i},\ \bar{r}_{-i} \in A_{-i}, \text{ and } \bar{m} \in Y, \\ \Pr(m|s_i|_{r_i}, r_{-i}) = \sum_{\tilde{s}_i|_{r_i} \in \tilde{S}_i|_{r_i}} \phi(\tilde{s}_i|_{r_i}) \Pr(m|\tilde{s}_i|_{r_i}, r_{-i}) \text{ for each } r_{-i} \in A_{-i} \text{ and } m \in Y. \end{cases} \tag{68}$$

Otherwise, we can assume that player i mixes $\tilde{s}_i|_{r_i} \in \tilde{S}_i|_{r_i}$ with probability $\phi(\tilde{s}_i|_{r_i})$ instead of taking $s_i|_{r_i}$.

Note that (68) implies that $s_i|_{r_i}$ is truthful. To see this, if player i tells a lie, then the distribution of the messages can be written as

$$\Pr(m|s_i|_{r_i}, r_{-i}) = \sum_{\hat{m}_i} s_i(r_i, s_i(r_i), \hat{m}_i)(m_i) \Pr(\hat{m}_i, m_{-i}|s_i(r_i), r_{-i})$$

for each $r_{-i} \in A_{-i}$ and $m \in Y$. Recall that $s_i(r_i) \in A_i$ is the action that player i with $s_i|_{r_i}$ takes after receiving r_i, and $s_i(r_i, s_i(r_i), \hat{m}_i)(m_i)$ is the probability that s_i sends m_i after receiving r_i, taking $s_i(r_i)$, and observing $\hat{m}_i$. Hence, this contradicts to (68).

Therefore, we can see $\bar{\sigma}_i$ as the following strategy: After receiving $r_i \in A_i = \text{supp}(\mu_i^{\text{full}})$, player i takes a pure and truthful strategy $s_i|_{r_i}$ with probability $\bar{\sigma}_i(s_i|_{r_i})$.

Suppose now that the mediator recommends r as follows: First, she draws r according to μ^{full}. For each i, given r_i, she draws the recommendation to player i, denoted by $\hat{r}_i$, according to $\sum_{s_i|_{r_i} \text{ takes } \hat{r}_i} \sigma_i(s_i|_{r_i})$. Let $\hat{\mu}^{\text{full}}$ denote such a strategy.

We show that $\underline{k}(\hat{\mu}^{\text{full}}, \lambda) \geq k(\lambda) - \varepsilon$. To this end, we check the following four conditions:

1. [effective full support] Recall that

$$\text{aff}\left(\{q_i(Y_i|r_i, r_{-i})\}_{r_{-i} \in A_{-i}}\right) = \text{aff}\left(\{q_i(Y_i|r_i, r_{-i})\}_{r_{-i} \in \text{supp}(\mu_{-i}^{\text{full}}|_{r_i})}\right).$$

   Since $\bar{\sigma}_n \in \Sigma_n^{\text{undetectable}}$ for each n, we have

$$q_i(Y_i|r_i, r_{-i}) = q_i(Y_i|r_i, (\bar{\sigma}_n(r_n))_{n \in -i})$$

for each r. Therefore, we have

$$\text{aff}\left(\{q_i(Y_i|r_i,r_{-i})\}_{r_{-i}\in A_{-i}}\right) = \text{aff}\left(\{q_i(Y_i|r_i,r_{-i})\}_{r_{-i}\in \text{supp}(\hat{\mu}^{\text{full}}_{-i}|_{r_i})}\right).$$

2. [incentive compatibility] By (66) and (67), with $\hat{\mu}^{\text{full}}$, for each $i \in I$, each supp($\hat{\mu}^{\text{full}}$)-undetectable deviation is not profitable. Hence, by the same proof as Lemma 6, there exists $x(r,m)$ such that, for all $s_i \in S_i$, we have

$$\mathbb{E}\left[u_i(r) + x_i(r,y)|\hat{\mu}^{\text{full}}\right] \geq \mathbb{E}\left[u_i(a_i,r_{-i}) + x_i(r,m_i,y_{-i})|s_i,\hat{\mu}^{\text{full}}\right].$$

3. [strict incentive compatibility] Moreover, by (68), $s_i|_{r_i}$ with $\bar{\sigma}_i(s_i|_{r_i}) > 0$ induces the extreme point in the distribution of the messages. Hence, we can make sure that, unless a deviation induces exactly the same distribution, player i's incentive is strict: For each $r_i \in \text{supp}(\hat{\mu}^{\text{full}}_i)$ and $\tilde{s}_i|_{r_i}$ with

$$\Pr(m|r_i,r_{-i}) \neq \Pr(m|\tilde{s}_i|_{r_i},r_{-i}) \text{ for some } r_{-i} \in A_{-i} \text{ and } m \in Y,$$

we have

$$\mathbb{E}\left[u_i(r) + x_i(r,y)|\mu,r_i\right] > \mathbb{E}\left[u_i(a_i,r_{-i}) + x_i(r,m_i,y_{-i})|\mu,r_i,\tilde{s}_i|_{r_i}\right].$$

4. [promise keeping] By adding/subtracting a constant, we can make sure that $\mathbb{E}\left[x(r,y)|\hat{\mu}^{\text{full}}\right] = 0$ and so $v = u(\hat{\mu}^{\text{full}})$. Hence, we have

$$\begin{aligned} \lambda \cdot v &\geq \lambda \cdot u(\hat{\mu}^{\text{full}}) \\ &\geq \lambda \cdot u(\mu^{\text{full}}) - \frac{\varepsilon}{2} \text{ by (63)} \\ &\geq k(\lambda) - \varepsilon \text{ by (62).} \end{aligned}$$

5. [self generation] We have $\mathbb{E}\left[x(r,y)|\hat{\mu}^{\text{full}}\right] = 0$.

Therefore, we have established that $\underline{k}(\lambda) \geq \underline{k}(\hat{\mu}^{\text{full}}, \lambda) \geq k(\lambda) - \varepsilon$, as desired.

We second consider the case with $\text{supp}(\lambda) = 1$, that is, $\lambda = \pm e_i$ for some i. Since [effective full support] and [strict incentive compatibility] are not issues, we are left to create $\hat{\mu}$ and $\hat{x}(r, m)$ that satisfy [incentive compatibility], [promise keeping], [self generation], and [ex post self generation] such that for each $\mu \in \Delta(A)$,

$$\underline{k}(\hat{\mu}, \lambda) \geq \min \{k(\lambda), l(\lambda)\} - \varepsilon. \tag{69}$$

Since we assume that the realized payoffs are observable, each player n's realized utility can be written as a function $\tilde{u}_n(a_n, y_n)$ of her own actions and signals.

Specifically, with a sufficiently large $K > 0$, we define $\hat{x}(r, m)$ as

$$\begin{aligned} \hat{x}_i(r, y) &= \sum_{j \neq i} \frac{\lambda_i}{K} \left(\tilde{u}_j(r_j, y_j) - \max_{\tilde{r}_j \in A_j, \tilde{y}_j \in Y_j} \tilde{u}_j(\tilde{r}_j, \tilde{y}_j) \right), \\ \hat{x}_j(r, y) &= \sum_{n \neq i, j} \tilde{u}_n(r_n, y_n) + \lambda_i K \tilde{u}_i(r_i, y_i) \text{ for each } j \in -i. \end{aligned}$$

By definition, [ex post self generation] is satisfied. Note that [ex post self generation] implies [self generation].

If $\lambda = e_i$, then we define $\hat{\mu}$ so that $\hat{\mu}$ almost maximizes player i's payoffs:

$$\hat{\mu} \in \arg\max_{\mu \in \Delta(A)} \mathbb{E} \left[u_i(a) + \frac{1}{K} \sum_{n \neq i} u_n(a) \right] \tag{70}$$

The mediator recommends r according to $\hat{\mu}$.

On the other hand, if $\lambda = -e_i$, then we consider the following minimax problem:

$$\min_{\mu_{-i} \in \Delta(A_{-i})} \max_{a_i \in A_i} \mathbb{E} \left[u_i(a) - \frac{1}{K} \sum_{n \neq i} u_n(a) \right]. \tag{71}$$

Let $(\hat{\mu}_{-i}, \hat{a}_i)$ be a solution for this problem. The mediator recommends r_{-i} according to $\hat{\mu}_{-i}$ to players $-i$ and recommends $r_i = \hat{a}_i$ with probability one to player i.

Let us verify [incentive compatibility] and [promise keeping]:

2. [Lower: incentive compatibility] We first consider the case with $\lambda_i = e_i$. For each $n \in I$, since $\hat{x}_n(r, m)$ does not depend on m_n, player n has the incentive to tell the truth. Note that, for each $n \in I$, we have

$$\begin{aligned} & \mathbb{E}\left[u_n(a) + \hat{x}_n(r, y) \mid a\right] \\ = \ & \text{positive constant} \times \left[u_i(a) + \frac{1}{K}\sum_{n' \neq i} u_{n'}(a)\right] + \text{constant.} \end{aligned}$$

Hence, each player has the incentive to follow $\hat{\mu}$ which maximizes (70).

We second consider the case with $\lambda_i = -e_i$. For each $n \in I$, since $\hat{x}_n(r, m)$ does not depend on m_n, player n has the incentive to tell the truth. Note that, for player i, we have

$$\begin{aligned} & \mathbb{E}\left[u_i(a) + \hat{x}_i(r, y) \mid a\right] \\ = \ & \text{positive constant} \times \left[u_i(a) - \frac{1}{K}\sum_{n \neq i} u_n(a)\right] + \text{constant.} \end{aligned}$$

On the other hand, for each $j \in -i$, we have

$$\begin{aligned} & \mathbb{E}\left[u_j(a) + \hat{x}_j(r, y)\right] \\ = \ & \text{negative constant} \times \left[u_i(a) - \frac{1}{K}\sum_{n \neq i} u_n(a)\right] + \text{constant.} \end{aligned}$$

Hence, player i wants to maximize $u_i(a) - \frac{1}{K}\sum_{n \neq i} u_j(a)$ while players $-i$ wants to minimize it. Hence, (71) ensures that $(\hat{\mu}_{-i}, \hat{a}_i)$ is incentive compatible.

3. [promise keeping] For $\lambda = e_i$, for sufficiently large K, we have

$$\begin{aligned} & \mathbb{E}\left[u_i(\mu) + \hat{x}_i(r, y) \mid \hat{\mu}\right] \\ \geq \ & \arg\max_a u_i(a) - \frac{1}{K}\max_{a \in A}\left|u_i(a)\right|. \end{aligned}$$

Hence, for each $\varepsilon > 0$, for sufficiently large K, we have

$$\lambda_i \mathbb{E}\left[u_i(\hat{\mu}) + \hat{x}_i(r, y) \mid \hat{\mu}\right] \geq k(\mu, \lambda) - \varepsilon.$$

For $\lambda = -e_i$, for sufficiently large K, we have

$$\begin{aligned} & \mathbb{E}\left[u_i(\hat{\mu}) + \hat{x}_i(r, y) \mid \hat{\mu}_{-i}, \hat{a}_i\right] \\ \leq & \min_{\mu_{-i} \in \Delta(A_{-i})} \max_{a_i \in A_i} u_i(a) + \frac{1}{K} \max_{a \in A} \left| u_i(a) \right|. \end{aligned}$$

Recall that, by $l(-e_i)$, we should have

$$k(\mu, \lambda) \geq \min_{\mu_{-i} \in \Delta(A_{-i})} \max_{a_i \in A_i} u_i(a).$$

Hence, for each $\varepsilon > 0$, for sufficiently large K, we have

$$\lambda_i \mathbb{E}\left[u_i(r) + \hat{x}_i(r, y) \mid \hat{\mu}_{-i}, \hat{a}_i\right] \geq l(\lambda) - \varepsilon.$$

Therefore, we have (69).

7.3.7 Proof of Lemma 2

Since $\underline{l}(\lambda) = \infty$ for each λ with $\operatorname{supp}(\lambda) = 1$, we focus on the case with $\operatorname{supp}(\lambda) = 2$. Moreover, by the same proof as Proposition 2, if there exists player i with $\lambda_i > 0$, then we have $\underline{l}(\lambda) \geq \underline{k}(\lambda)$. By Lemma 1, this implies $\underline{l}(\lambda) \geq \min\{k(\lambda), l(\lambda)\} - \varepsilon$.

Therefore, we are left to show that, for λ with $\operatorname{supp}(\lambda) \geq 2$ and $\lambda_i \leq 0$ for all i, for each $\varepsilon > 0$, we have $\underline{l}(\lambda) \geq l(\lambda) - \varepsilon$. Notice that the requirement that τ is strictly incentive compatible is parallel to [strict incentive compatibility] in the definition of $\underline{k}(\mu, \lambda)$. Hence, we follow the same steps in the proof of Lemma 1 except that we need to make sure that [effective punishment] and [self generation after unilateral deviation] are satisfied.

First, by Proposition 2, there exist $\tau \in \Delta(A)$ and v such that $l(\lambda) = \lambda \cdot v$ and

$$|\lambda| \cdot v \geq \max_{\sigma \in SD(\text{supp}(\lambda))} \sum_{r \in A} \tau(r) \sum_{i \in I} |\lambda_i| \, u_i(r, \sigma_i).$$

As we created μ^{full} from μ in Section 7.3.7, we perturb τ so that τ^{full} satisfies the full support: $\tau^{\text{full}} = (1-\eta)\tau + \eta \sum_{a \in A} \frac{\eta}{|A|}$. For each $\varepsilon > 0$, for sufficiently small $\eta > 0$, we have

$$\max_{\sigma \in SD(\text{supp}(\lambda))} \sum_{r \in A} \tau^{\text{full}}(r) \sum_{i \in I} |\lambda_i| \, u_i(r, \sigma_i) \leq \max_{\sigma \in SD(\text{supp}(\lambda))} \sum_{a \in A} \tau(r) \sum_{i \in I} |\lambda_i| \, u_i(r, \sigma_i) + \varepsilon.$$

Hence, if τ^{full} is strictly incentive compatible, then we are done.

Suppose that τ^{full} is not strictly incentive compatible. As we created $\hat{\mu}^{\text{full}}$ from μ^{full}, we create $\hat{\tau}^{\text{full}}$ from τ^{full} so that $\hat{\tau}^{\text{full}}$ is strictly incentive compatible. We are left to show that $\hat{\tau}^{\text{full}}$ is sufficiently severe punishment:

$$\max_{\sigma \in SD(\text{supp}(\lambda), \tau)} \sum_{r \in A} \hat{\tau}^{\text{full}}(r) \sum_{i \in I} |\lambda_i| \, u_i(r, \sigma_i) \leq \max_{\sigma \in SD(\text{supp}(\lambda))} \sum_{r \in A} \tau^{\text{full}}(r) \sum_{i \in I} |\lambda_i| \, u_i(r, \sigma_i) + \frac{\varepsilon}{2}.$$

Here, $SD(\text{supp}(\lambda), \tau)$ is the similar deviation with respect to action profiles in $\text{supp}(\tau)$:

$$SD(\text{supp}(\lambda), \tau) \equiv \left\{ \sigma^J = (\sigma_i)_{i \in J} \in \prod\nolimits_{j \in J} \Sigma_j : \begin{array}{c} \forall i, \forall j \in \text{supp}(\lambda), \\ \Pr(m|\sigma_i, r) = \Pr(m|\sigma_j, r) \\ \text{for all } r \in \text{supp}(\tau) \text{ and } m \in Y \end{array} \right\}.$$

By (66), the restriction of $r \in \text{supp}\left(\hat{\tau}^{\text{full}}\right)$ does not matter:

$$SD(\text{supp}(\lambda), \hat{\tau}^{\text{full}}) = SD(\text{supp}(\lambda), \tau^{\text{full}}) = SD(\text{supp}(\lambda)).$$

Hence, we are left to show that

$$\max_{\sigma \in SD(\text{supp}(\lambda))} \sum_{r \in A} \hat{\tau}^{\text{full}}(r) \sum_{i \in I} |\lambda_i| \, u_i(r, \sigma_i) \leq \max_{\sigma \in SD(\text{supp}(\lambda))} \sum_{r \in A} \tau^{\text{full}}(r) \sum_{i \in I} |\lambda_i| \, u_i(r, \sigma_i).$$

Recall that, by (64), for each player i and $a_i \in A_i$, we have

$$\sum_{r \in A} \hat{\tau}^{\text{full}}(r) u_i(a_i, r_{-i}) = \sum_{r \in A} \tau^{\text{full}}(r) u_i(a_i, r_{-i}).$$

Therefore, we have

$$\max_{\sigma \in SD(\text{supp}(\lambda))} \sum_{a \in A} \hat{\tau}^{\text{full}}(a) u_n(a, \sigma_n) = \max_{\sigma \in SD(\text{supp}(\lambda))} \sum_{a \in A} \tau^{\text{full}} u_n(a, \sigma_n),$$

as desired.

7.4 Proof of Lemma 6

We construct an equilibrium to support $v_1 + v_2$ arbitrarily close to 3. To this end, we see the repeated game as repetition of T-period blocks.

In each block, the mediator has two possible states $s \in \{G, B\}$. Intuitively, G is the good state while B is the punishment state. We first define the strategy given s, and second offer the state transition between blocks. Since the equilibrium is recursive between blocks, we focus on the initial block.

The Mediator's Strategy in $s = G$ If $s = G$, then the mediator recommends actions as stated in Section 5.2: In period 1, the mediator recommends r according to μ_0 with

$$\begin{aligned}
\Pr((D(1), D(2), U(3), U(4)) \,|\mu_0) &= \frac{1}{2}, \\
\Pr((M(1), M(2), U(3), U(4)) \,|\mu_0) &= \frac{1}{4}, \\
\Pr((U(1), U(2), U(3), U(4)) \,|\mu_0) &= \frac{1}{4}.
\end{aligned}$$

In each period $t \geq 2$, if player 5 reported the signal of g last period, that is, if $m_{5,t-1} = g$, then $r = (M(1), M(2), U(3), U(4))$ with probability one. If player 5 reported the signal of b last period, that is, if $m_{5,t-1} = b$, then $r = (D(1), D(2), U(3), U(4))$ with probability $\frac{2}{3}$ and

$r = (U(1), U(2), U(3), U(4)) = \frac{1}{3}$.

The Mediator's Strategy in $s = B$ If $s = B$, then the mediator recommends a static best response, $P(1), P(2), U_m(3), D(4)$ repeatedly.

State Transition Once $s = B$ happens, then the state in the next block will be b. Intuitively, once the punishment state B is triggered, then it will last forever. It will be useful to calculate player i's payoff in state B for $i \in \{1, 2\}$. Since

$$u_i(P(1), P(2), U_m(3), D(4)) = 0,$$

player i's value for $s = B$ is $v(B) = 0$.

We are left to define the probability of triggering $s = B$ when the current state is G.

To this end, it will be useful to define the value

$$v(G, \delta, \varepsilon) = \frac{3}{2} + \frac{1-\delta}{1-\delta^T} \mathbb{E}\left[\left\{\begin{array}{c} \sum_{t=2}^{T} \left\{1_{\{m_{5,t-1}=g\}} (1-\delta) \delta^{t-1} 2 + 1_{\{m_{5,t-1}=b\}} (1-\delta) \delta^{t-1} \frac{13}{6}\right\} \\ + \left\{1_{\{m_{5,T}=g\}} \delta^{T-1} \frac{2}{3}\right\} \\ - \sum_{t=1}^{T} \delta^{t-1} \varepsilon \end{array}\right\} \mid \sigma_m^T\right],$$

where σ_m^T is the mediator's strategy defined above, assuming that all the players are faithful. Intuitively, we define the transition probability so that such v is player i's value for $i \in \{1, 2\}$ in state G. Note that

$$\lim_{T \to 1} \lim_{\delta \to 1} v(G, \delta, \varepsilon) = \frac{3}{2} - \varepsilon.$$

In order to incentivize player $i \in \{1, 2\}$ to be faithful, it will be important to make sure that player i's continuation payoff after $\left(r^T, m^T\right)$ is equal to

$$w(r^T, m^T) = v(G, \delta, \varepsilon) + \frac{1-\delta}{\delta^T} \left\{\begin{array}{c} \sum_{t=2}^{T} \left\{1_{\{m_{5,t-1}=g\}} (1-\delta) \delta^{t-1} 2 + 1_{\{m_{5,t-1}=b\}} (1-\delta) \delta^{t-1} \frac{13}{6}\right\} \\ + \left\{1_{\{m_{5,T}=g\}} \delta^{T-1} \frac{2}{3}\right\} \\ - \sum_{t=1}^{T} \delta^{t-1} \varepsilon \end{array}\right\}. \quad (72)$$

Since $v(G,\delta,\varepsilon)$ is the value at $s=g$ and $v(B)=0$ is the value at $s=b$, we should define the transition probability such that

$$\Pr\left(G|r^T,m^T\right)v(G,\delta,\varepsilon)=w(r^T,m^T),$$

that is

$$\Pr\left(G|r^T,m^T\right)=\frac{w(r^T,m^T)}{v(G,\delta,\varepsilon)}.$$

Here, $\Pr\left(G|r^T,m^T\right)$ is the probability that the next state is G when the current state is G and the mediator's history is (r^T,m^T).

For sufficiently large δ, we have $\Pr\left(G|r^T,m^T\right)\in[0,1]$ because of the following two reasons: (i) For sufficiently large δ, for each (r^T,m^T), we have $w(r^T,m^T)\leq v(G,\delta,\varepsilon)$, and (ii) we have $\lim_{\delta\to 1}\max_{r^T,m^T}\left|w(r^T,m^T)-v(G,\delta,\varepsilon)\right|=0$.

Incentive Compatibility Let us now verify that each player has the incentive to be faithful. If $s=B$, since the recommendation is a static Nash equilibrium, the recommendation is stationary, and $s=B$ in the next block, it is optimal for each player to be faithful.

If $s=G$, then player 5 has the incentive to tell the truth about $y_{5,t}$ since her payoff is independent of the mediator's strategy. In addition, as seen in Section 5.2, players 3 and 4 believe that

$$\begin{aligned}
r_t &= (D(1),D(2),U(3),U(4)) \text{ with probability } \Pr(m_{5,t-1}=b)\times\frac{2}{3}=\frac{1}{2},\\
r_t &= (M(1),M(2),U(3),U(4)) \text{ with probability } \Pr(m_{5,t-1}=g)=\frac{1}{4},\\
r_t &= (U(1),U(2),U(3),U(4)) \text{ with probability } \Pr(m_{5,t-1}=b)\times\frac{2}{3}=\frac{1}{3}.
\end{aligned}$$

Moreover, their actions do not affect the distribution of signals. Hence, it is optimal for players 3 and 4 to be faithful.

Hence, we are left to show the incentive of players 1 and 2. Let us focus on player 1's incentive. (The proof for player 2's incentive is the same.) Guess that $v(G,\delta,\varepsilon)$ is player

1's value in state G. Then, by the value function and definition of $\Pr\left(G|r^T,m^T\right)$, player 1 wants to maximize

$$\mathbb{E}\left[(1-\delta)\sum_{t=1}^{T}\delta^{t-1}u_1(a_i,r_{-i})+\delta^T w(r^T,m^T)\mid\sigma_i,\sigma_m^T\right].$$

By (72), after affine transformation, player 1 wants to maximize

$$\mathbb{E}\left[\sum_{t=1}^{T}\delta^{t-1}u_1(a_{i,t},r_{-i,t})+\sum_{t=2}^{T}\left\{1_{\{m_{5,t-1}=g\}}(1-\delta)\,\delta^{t-1}2+1_{\{m_{5,t-1}=b\}}(1-\delta)\,\delta^{t-1}\frac{13}{6}\right\}+\left\{1_{\left\{m_{5,T}=g\right\}}\delta^{T-1}\frac{2}{3}\right\}\right.$$

Let us verify player 1's incentive by backward induction: In period T, player 1 wants to maximize

$$\delta^{T-1}\mathbb{E}\left[u_1(a_{i,T},r_{-i,T})+1_{\left\{m_{5,T}=g\right\}}\frac{2}{3}|h_i^T,\sigma_i,\sigma_m^T\right].$$

if $M(1)$ or $D(1)$ is recommended, then $M(2)$ or $D(2)$ is recommended to player 2, respectively. Since this is a static best response and the distribution of y_5 is independent of player 1's action, player 1 has the incentive to follow the recommendation. If $U(1)$ is recommended, then $U(2)$ is recommended to player 2. Player 1 does not take $a_1\notin\{U(1),U_p(1),U_m(1)\}$ since $U(1)$ is weakly better in terms of payoffs and the distribution of y_5 does not change. Moreover, the term $1_{\left\{m_{5,T}=g\right\}}\frac{2}{3}$ makes player 1 indifferent among $\{U(1),U_p(1),U_m(1)\}$ by (18). Hence, player 1 has the incentive to follow the recommendation.

Suppose that player 1 will follow the recommendation from $t\geq\tau+1$. Then, in period $t=\tau$, player 1 wants to maximize

$$\delta^{\tau-1}\mathbb{E}\left[u_1(a_{i,\tau},r_{-i,\tau})+\delta u_1(r_{\tau+1})+1_{\{m_{5,\tau}=g\}}(1-\delta)\,2+1_{\{m_{5,\tau}=b\}}(1-\delta)\,\frac{13}{6}|h_i^\tau,\sigma_i,\sigma_m^T\right]$$

since (i) from period $\tau+1$, the distribution of $m_{5,t}$ is independent of the action in period τ.

There are following four possibilities:

1. If player 1 follows the recommendation, then $u_1(a_{1,\tau},r_{-1,\tau})=2$. Since $\Pr(m_{5,\tau}=g)=$

$\frac{1}{4}$,

$$\mathbb{E}\left[\delta u_1(r_{\tau+1}) + 1_{\{m_{5,\tau}=g\}}(1-\delta)\,2 + 1_{\{m_{5,\tau}=b\}}(1-\delta)\,\frac{13}{6}\right] = \frac{3}{2}.$$

2. If player 1 takes $U_p(1)$, then $u_1(a_{1,\tau}, r_{-1,\tau}) = \frac{3}{2}$. Since since $\Pr(m_{5,\tau} = g) = 1$,

$$\mathbb{E}\left[\delta u_1(r_{\tau+1}) + 1_{\{m_{5,\tau}=g\}}(1-\delta)\,2 + 1_{\{m_{5,\tau}=b\}}(1-\delta)\,\frac{13}{6}\right] = 2.$$

3. If player 1 takes $U_m(1)$, then $u_1(a_{1,\tau}, r_{-1,\tau}) = \frac{13}{6}$. Since since $\Pr(m_{5,\tau} = g) = 0$,

$$\mathbb{E}\left[\delta u_1(r_{\tau+1}) + 1_{\{m_{5,\tau}=g\}}(1-\delta)\,2 + 1_{\{m_{5,\tau}=b\}}(1-\delta)\,\frac{13}{6}\right] = \frac{13}{6}.$$

4. Player 1 does not take $a_1 \notin \{U(1), U_p(1), U_m(1)\}$ since $U(1)$ is weakly better in terms of payoffs and the distribution of y_5 does not change.

Hence, player 1 is indifferent between $U(1)$, $U_p(1)$, and $U_m(1)$.

Therefore, each player has the incentive to follow the recommendation.

Value Functions Since each player has the incentive to follow the recommendation, player 1's value in state G is actually equal to

$$\begin{aligned}
v(G,\delta,\varepsilon) &= \mathbb{E}\left[(1-\delta)\sum_{t=1}^{T}\delta^{t-1}u_1(a_i, r_{-i}) + \delta^T w(r^T, m^T) \mid \sigma_i, \sigma_m^T\right] \\
&= \mathbb{E}\left[\begin{array}{c}
(1-\delta)\sum_{t=1}^{T}\delta^{t-1}u_1(a_i, r_{-i}) \\
+\delta^T v(G,\delta,\varepsilon) \\
+(1-\delta)\left\{\begin{array}{c}
\sum_{t=2}^{T}\left\{1_{\{m_{5,t-1}=g\}}(1-\delta)\,\delta^{t-1}2 + 1_{\{m_{5,t-1}=b\}}(1-\delta)\,\delta^{t-1}\frac{13}{6}\right\} \\
+\left\{1_{\{m_{5,T}=g\}}\delta^{T-1}\frac{2}{3}\right\} \\
-\sum_{t=1}^{T}\delta^{t-1}\varepsilon
\end{array}\right\}
\end{array} \mid \sigma_m^T\right],
\end{aligned}$$

that is,

$$v(G,\delta,\varepsilon) = \frac{3}{2}+\frac{1-\delta}{1-\delta^T}\mathbb{E}\left[\left\{\begin{array}{c}\sum_{t=2}^{T}\left\{1_{\{m_{5,t-1}=g\}}(1-\delta)\,\delta^{t-1}2+1_{\{m_{5,t-1}=b\}}(1-\delta)\,\delta^{t-1}\frac{13}{6}\right\}\\ +\left\{1_{\{m_{5,T}=g\}}\delta^{T-1}\frac{2}{3}\right\}\\ -\sum_{t=1}^{T}\delta^{t-1}\varepsilon\end{array}\right\} \mid \sigma_m^T\right]$$

$$\to_{\delta\to 1}\frac{3}{2}-\frac{1}{T}\frac{2}{3}-\varepsilon.$$

Hence, for sufficiently large T, for sufficiently large δ, we have $v(G,\delta,\varepsilon)$ sufficiently close to $\frac{3}{2}$, as desired.

References

[1] Ben-Porath, E. (1998): "Correlation without Mediation: Expanding the Set of Equilibrium Outcomes by 'Cheap' Pre-play Procedures," *Journal of Economic Theory*, 80(1), 108-122.

[2] Gerardi, D. (2004): "Unmediated Communication in Games with Complete and Incomplete Information," *Journal of Economic Theory*, 114 (1), 104-131.

[3] Fudenberg, D. and D. K. Levine (1994): "Efficiency and Observability with Long-Run and Short-Run Players," *Journal of Economic Theory*, 62(1), 103-135.

[4] Fudenberg, D., D. Levine, and E. Maskin (1994): "The Folk Theorem with Imperfect Public Information," *Econometrica*, 62(5), 997-1039.

[5] Fudenberg, D., D. K. Levine, and S. Takahashi (2007): "Perfect Public Equilibrium when Players are Patient," *Games and Economic Behavior*, 61(1), 27-49.

[6] Fudenberg, D., and E. Maskin (1986): "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information," *Econometrica*, 53(3), 533-554.

[7] Hörner, J., and W. Olszewski (2006): "The folk theorem for games with private almost-perfect monitoring," *Econometrica*, 74(6), 1499■544.

[8] Kandori, M. and H. Matsushima (1998): "Private Observation, Communication, and Collusion," *Econometrica*, 66(3), 627-652.

[9] Rahman, D. (2012): "But Who will Monitor the Monitor?," *American Economic Review*, 102(6), 2767-2797.

[10] Renault, J. and T. Tomala (2004): "Communication equilibrium payoffs in repeated games with imperfect monitoring," *Games and Economic Behavior*, 49(2), 313-344.

[11] Rockafellar, R. T. (1970): *Convex Analysis*. Princeton University Press.

[12] Sugaya, T. (2014a): "Folk Theorem in Repeated Games with Private Monitoring," *manuscript.*

[13] Sugaya, T. (2014b): "Folk Theorem in Repeated Games with Private Monitoring: A General Two Players," *manuscript.*

[14] Sugaya, T. and A. Wolitzky (2014): "Perfect Versus Imperfect Monitoring in Repeated Games," *manuscript.*