On the Equilibrium Payoff Set in Repeated Games with Imperfect Private Monitoring*

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

April 19, 2015

Abstract

We provide simple sufficient conditions for the existence of a tight, recursive upper bound on the sequential equilibrium payoff set at a fixed discount factor in two-player repeated games with imperfect private monitoring. The bounding set is the sequential equilibrium payoff set with perfect monitoring and a mediator. We show that this bounding set admits a simple recursive characterization, which nonetheless necessarily involves the use of private strategies. Under our conditions, this set describes precisely those payoff vectors that arise in equilibrium for some private monitoring structure.

*This replaces our earlier paper, "Perfect Versus Imperfect Monitoring in Repeated Games." For helpful comments, we thank Dilip Abreu, Gabriel Carroll, Henrique de Oliveira, Glenn Ellison, Drew Fudenberg, Michihiro Kandori, Hitoshi Matsushima, Stephen Morris, Ilya Segal, Tadashi Sekiguchi, Andrzej Skrzypacz, Satoru Takahashi, Juuso Toikka, and Thomas Wiseman.

1 Introduction

Like many dynamic economic models, repeated games are typically studied using recursive methods. In an incisive paper, Abreu, Pearce, and Stacchetti (1990; henceforth APS) recursively characterized the perfect public equilibrium payoff set at a fixed discount factor in repeated games with imperfect public monitoring. Their results (along with related contributions by Fudenberg, Levine, and Maskin (1994) and others) led to fresh perspectives on problems like collusion (Green and Porter, 1984; Athey and Bagwell, 2001), relational contracting (Levin, 2003), and government credibility (Phelan and Stacchetti, 2001). However, other important environments—like collusion with secret price cuts (Stigler, 1964) or relational contracting with subjective performance evaluations (Levin, 2003; MacLeod, 2003; Fuchs, 2007)—involve imperfect *private* monitoring, and it is well-known that the methods of APS do not easily extend to such settings (Kandori, 2002). Whether the equilibrium payoff set in repeated games with private monitoring exhibits any tractable recursive structure at all is thus a major question.

In this paper, we do *not* make any progress toward giving a recursive characterization of the sequential equilibrium payoff set in a repeated game with a *given* private monitoring structure. Instead, working in the context of two-player games, we provide simple conditions for the existence of a tight, recursive upper bound on this set.[1] Equivalently, under these conditions we give a recursive characterization of the set of payoffs that can be attained in equilibrium for *some* private monitoring structure. Thus, from the perspective of an observer who knows the monitoring structure, our results give an upper bound on how well players can do in a repeated game; while from the perspective of an observer who does not know the monitoring structure, our results exactly characterize how well the players can do.

Throughout the paper, the set we use to upper-bound the equilibrium payoff set with private monitoring is the equilibrium payoff set with perfect monitoring and a mediator (or, more precisely, the closure of the strict equilibrium payoff set with this information

[1] The ordering on sets of payoff vectors here is *not* the usual one of set inclusion, but rather dominance in all non-negative Pareto directions. Thus, we say that A upper-bounds B if every payoff vector $b \in B$ is Pareto dominated by some payoff vector $a \in A$.

structure). We do not take a position on the realism of allowing for a mediator, and instead view the model with a mediator as a purely technical device that—as we show—is useful for bounding the equilibrium payoff set with private monitoring. We thus show that the equilibrium payoff set with private monitoring has a simple, recursive upper bound by establishing two main results:

1. Under some conditions, the equilibrium payoff set with perfect monitoring and a mediator is indeed an upper bound on the equilibrium payoff set with any private monitoring structure.

2. The equilibrium payoff set with perfect monitoring and a mediator has a simple recursive structure.

At first glance, it might seem surprising that any conditions at all are needed for the first of these results, as one might think that improving the precision of the monitoring structure and adding a mediator can only expand the equilibrium set. But this is not the case: giving a player more information about her opponents' past actions splits her information sets and thus gives her new ways to cheat, and indeed we show by example that (unmediated) imperfect private monitoring can sometimes outperform (mediated) perfect monitoring. Our first result provides sufficient conditions for this not to happen. Thus, another contribution of our paper is pointing out that perfect monitoring is not necessarily the optimal monitoring structure in a repeated game (even if it is advantaged by giving players access to a mediator), while also giving sufficient conditions under which perfect monitoring is indeed optimal.

Our sufficient condition for mediated perfect monitoring to outperform any private monitoring structure is that there is a feasible continuation payoff vector v such that no player i is tempted to deviate if she gets continuation payoff v_i when she conforms and is minmaxed when she deviates. This is a joint restriction on the stage game and the discount factor, and it is essentially always satisfied when players are at least moderately patient. (They need not be extremely patient: none of our main results concern the limit $\delta \to 1$.) The reason why this condition is sufficient for mediated perfect monitoring to outperform private monitoring in two-player games is fairly subtle, so we postpone a detailed discussion.

Our second main result also involves some subtleties. In repeated games with perfect monitoring *without* a mediator, all strategies are public, so the sequential (equivalently, subgame perfect) equilibrium set coincides with the perfect public equilibrium set, which was recursively characterized by APS. On the other hand, *with* a mediator—who makes private action recommendations to the players—private strategies play a crucial role, and APS's characterization does not apply. We nonetheless show that the sequential equilibrium payoff set with perfect monitoring and a mediator does have a simple recursive structure. Under the sufficient conditions for our first result, a recursive characterization is obtained by replacing APS's generating operator B with what we call a *minmax-threat generating operator* $\tilde{B}$: for any set of continuation payoffs W, the set $\tilde{B}(W)$ is the set of payoffs that can be attained when on-path continuation payoffs are drawn from W and deviators are minmaxed.[2] To see intuitively why deviators can always be minmaxed in the presence of a mediator—and also why private strategies cannot be ignored—suppose that the mediator recommends a target action profile $a \in A$ with probability $1 - \varepsilon$, while recommending every other action profile with probability $\varepsilon / (|A| - 1)$; and suppose further that if some player i deviates from her recommendation, the mediator then recommends that her opponents minmax her in every future period. In such a construction, player i's opponents never learn that a deviation has occurred, and they are therefore always willing to follow the recommendation of minmaxing player i.[3] (This construction clearly relies on private strategies: if the mediator's recommendations were public, players would always see when a deviation occurs, and they then might not be willing to minmax the deviator.) Our recursive characterization of the equilibrium payoff set with perfect monitoring and a mediator takes into account the

[2] The equilibrium payoff set with perfect monitoring and a mediator has a recursive structure whether or not the sufficient conditions for our first result are satisfied, but the characterization is somewhat more complicated in the general case. See Section 9.1.

[3] In this construction, the mediator *virutally implements* the target action profile. For other applications of virtual implementation in games with a mediator, see Lehrer (1992), Mertens, Sorin, and Zamir (1994), Renaul and Tomala (2004), Rahman and Obara (2010), and Rahman (2012).

possibility of minmaxing deviators in this way.

We also consider several extensions of our results. Perhaps most importantly, we establish two senses in which the equilibrium payoff set with perfect monitoring and a mediator is a *tight* upper bound on the equilibrium payoff set with any private monitoring structure. First, mediated perfect monitoring is itself an example of a *nonstationary* monitoring structure, meaning that the distribution of signals can depend on everything that has happened in the past, rather than only on current actions. Thus, our upper bound is trivially tight in the space of nonstationary monitoring structures. Second, with a standard, *stationary* monitoring structure, where the signal distribution depends only on the current actions, we show that the mediator can be replaced by ex ante correlation and cheap talk. Hence, our upper bound is also tight in the space of stationary monitoring structures when an ex ante correlating device and cheap talk are available.

This paper is certainly not the first to develop recursive methods for private monitoring repeated games. In an early and influential paper, Kandori and Matsushima (1998) augment private monitoring repeated games with opportunities for public communication among players, and provide a recursive characterization of the equilibrium payoff set for a subclass of equilibria that is large enough to yield a folk theorem. Tomala (2009) gives related results when the repeated game is augmented with a mediator rather than only public communication. However, neither paper provides a recursive upper bound on the entire sequential equilibrium payoff set at a fixed discount factor in the repeated game.[4] Amarante (2003) does give a recursive characterization of the equilibrium payoff set in private monitoring repeated games, but the state space in his characterization is the set of repeated game histories, which grows over time. In contrast, our upper bound is recursive in payoff space, just like in APS.

[4] Ben-Porath and Kahneman (1996) and Compte (1998) also prove folk theorems for private monitoring repeated games with communication, but they do not emphasize recursive methods away from the $\delta \to 1$ limit. Lehrer (1992), Mertens, Sorin, and Zamir (1994), and Renault and Tomala (2004) characterize the communication equilibrium payoff set in *undiscounted* repeated games. These papers study how imperfect monitoring can limit the equilibrium payoff set without discounting, whereas our focus is on how discounting can limit the equilibrium payoff set independently of the monitoring structure.

A different approach is taken by Phelan and Skrzypacz (2012) and Kandori and Obara (2010), who develop recursive methods for *checking* whether a given finite-state strategy profile is an equilibrium in a private monitoring repeated game. Their results do not give a recursive characterization or upper bound on the equilibrium payoff set. The type of recursive methods used in their papers is also different: their methods for checking whether a given strategy profile is an equilibrium involve a recursion on the sets of beliefs that players can have about each other's states, rather than a recursion on payoffs.

Recently, Awaya and Krishna (2014) and Pai, Roth, and Ullman (2014) derive bounds on payoffs in private monitoring repeated games as a function of the monitoring structure. The bounds in these papers come from the observation that, if an individual's actions can have only a small impact on the distribution of signals, then the shadow of the future can have only a small effect on her incentives. In contrast, our payoff bounds apply for all monitoring structures, including those in which individuals' actions have a large impact on the signal distribution.[5]

Finally, we have emphasized that our results can be interpreted either as giving an upper bound on the equilibrium payoff set in a repeated game for a *particular* private monitoring structure, or as characterizing the set of payoffs that can arise in equilibrium in a repeated game for *some* private monitoring structure. With the latter interpretation, our paper shares a motivation with Bergemann and Morris (2013), who characterize the set of payoffs that can arise in equilibrium in a static incomplete information game for some information structure. Yet another interpretation of our results is that they establish that information is valuable in mediated repeated games, in that—under our sufficient conditions—players cannot benefit from imperfections in the monitoring technology. This interpretation connects our paper to the literature on the value of information in static incomplete information games (e.g., Gossner, 2000; Lehrer, Rosenberg, and Shmaya, 2010; Bergemann and Morris, 2013).

The rest of the paper is organized as follows. Section 2 describes our models of repeated

[5] A working paper by Cherry and Smith (2011) is the first to consider the issue of upper-bounding the equilibrium payoff set with imperfect private monitoring with the equilibrium set with perfect monitoring and correlating devices. However, the result of theirs which is most relevant for us–their Theorem 3–appears to be contradicted by our example in Section 3 below.

games with imperfect private monitoring and repeated games with perfect monitoring and a mediator, which are standard. Section 3 gives an example showing that private monitoring can sometimes outperform perfect monitoring with a mediator. Section 4 develops some preliminary results about repeated games with perfect monitoring and a mediator. Section 5 presents our first main result, which gives sufficient conditions for such examples not to exist: the proof of this result is lengthy and is deferred to Section 8. Section 6 presents our second main result: a simple recursive characterization of the equilibrium payoff set with perfect monitoring and a mediator. Combining the results of Sections 5 and 6 gives the desired recursive upper bound on the equilibrium payoff set with private monitoring. Section 7 illustrates the calculation of the upper bound with an example. Section 9 discusses partial versions of our results that apply when our sufficient conditions do not hold, as in the case of more than two players, as well as the tightness of our upper bound. Section 10 concludes.

2 Model

A stage game $G = \left(I, (A_i)_{i \in I}, (u_i)_{i \in I}\right)$ is repeated in periods $t = 1, 2, \ldots$, where $I = \{1, \ldots, |I|\}$ is the set of players, A_i is the finite set of player i's actions, and $u_i : A \to \mathbb{R}$ is player i's payoff function. Players maximize expected discounted payoffs with common discount factor δ. We compare the equilibrium payoff sets in this repeated game under *private monitoring* and *mediated perfect monitoring*.

2.1 Private Monitoring

In each period t, the game proceeds as follows: Each player i takes an action $a_{i,t} \in A_i$. A signal $z_t = (z_{i,t})_{i \in I} \in (Z_i)_{i \in I} = Z$ is drawn from distribution $p(z_t | a_t)$, where Z_i is the finite set of player i's signals and $p(\cdot | a)$ is the monitoring structure. Player i observes $z_{i,t}$.

A period t history of player i's is an element of $H_i^t = (A_i \times Z_i)^{t-1}$, with typical element $h_i^t = (a_{i,\tau}, z_{i,\tau})_{\tau=1}^{t-1}$, where H_i^1 consists of the null history $\emptyset$. A (behavior) strategy of player i's is a map $\sigma_i : \bigcup_{t=1}^{\infty} H_i^t \to \Delta(A_i)$. Let $H^t = (H_i^t)_{i \in I}$.

We do not impose the common assumption that a player's payoff is measurable with respect to her own action and signal (i.e., that players observe their own payoffs), because none of our results need this assumption. All examples considered in the paper do however satisfy this measurability condition.

The solution concept is sequential equilibrium. More precisely, a belief system of player i's is a map $\beta_i : \bigcup_{t=1}^{\infty} H_i^t \to \bigcup_{t=1}^{\infty} \Delta(H^t)$ satisfying $\operatorname{supp} \beta_i(h_i^t) \subseteq (h_i^t, H_{-i}^t)$ for all t; we also write $\beta_i(h^t|h_i^t)$ for the probability of h^t under $\beta_i(h_i^t)$. We say that an assessment (σ, β) constitutes a sequential equilibrium if the following two conditions are satisfied:

1. [Sequential rationality] For each player i and history h_i^t, σ_i maximizes player i's expected continuation payoff at history h_i^t under belief $\beta_i(h_i^t)$.

2. [Consistency] There exists a sequence of completely mixed strategy profiles (σ^n) such that the following two conditions hold:

   (a) σ^n converges to σ (pointwise in t): For all $\varepsilon > 0$ and t, there exists N such that, for all $n > N$,
   $$\left|\sigma_i^n(h_i^t) - \sigma_i(h_i^t)\right| < \varepsilon \text{ for all } i \in I, h_i^t \in H_i^t.$$

   (b) Conditional probabilities converge to β (pointwise in t): For all $\varepsilon > 0$ and t, there exists N such that, for all $n > N$,
   $$\left|\frac{\Pr^{\sigma^n}(h_i^t, h_{-i}^t)}{\sum_{\tilde{h}_{-i}^t} \Pr^{\sigma^n}(h_i^t, \tilde{h}_{-i}^t)} - \beta_i(h_i^t, h_{-i}^t \mid h_i^t)\right| < \varepsilon \text{ for all } i \in I, h_i^t \in H_i^t, h_{-i}^t \in H_{-i}^t.$$

We choose this relatively permissive definition of consistency (requiring that strategies and beliefs converge only pointwise in t) to make our results upper-bounding the sequential equilibrium payoff set stronger. The results with a more restrictive definition (requiring uniform convergence) would be essentially the same.

2.2 Mediated Perfect Monitoring

In each period t, the game proceeds as follows: A mediator sends a private message $m_{i,t} \in M_i$ to each player i, where M_i is a finite message set for player i. Each player i takes an action $a_{i,t} \in A_i$. All players and the mediator observe the action profile $a_t \in A$.

A period t history of the mediator's is an element of $H_m^t = (M \times A)^{t-1}$, with typical element $h_m^t = (m_\tau, a_\tau)_{\tau=1}^{t-1}$, where H_m^1 consists of the null history. A strategy of the mediator's is a map $\mu : \bigcup_{t=1}^{\infty} H_m^t \to \Delta(M)$. A period t history of player i's is an element of $H_i^t = (M_i \times A)^{t-1} \times M_i$, with typical element $h_i^t = \left((m_{i,\tau}, a_\tau)_{\tau=1}^{t-1}, m_{i,t}\right)$, where $H_i^1 = M_i$.[6] A strategy of player i's is a map $\sigma_i : \bigcup_{t=1}^{\infty} H_i^t \to \Delta(A_i)$.

The definition of sequential equilibrium is the same as with private monitoring, except that sequential rationality is imposed (and beliefs are defined) only at histories consistent with the mediator's strategy. The interpretation is that the mediator is not a player in the game, but rather a "machine" that cannot tremble. Note that with this definition, an assessment (including the mediator's strategy) (μ, σ, β) is a sequential equilibrium with mediated perfect monitoring if and only if (σ, β) is a sequential equilibrium with the "non-stationary" private monitoring structure where $Z_i = M_i \times A$ and $p(\cdot|h_m^{t+1})$ coincides with perfect monitoring of actions with messages given by $\mu(h_m^{t+1})$ (see Section 9.2).

As in Forges (1986) and Myerson (1986), any equilibrium distribution over (infinite paths of) actions arises in an equilibrium of the following form:

1. [Messages are action recommendations] $M = A$.

2. [Obedience/incentive compatibility] At history $h_i^t = ((r_{i,\tau}, a_\tau)_{\tau=1}^{t-1}, r_{i,t})$, player i plays $a_{i,t} = r_{i,t}$.

Without loss of generality, we restrict attention to such *obedient* equilibria throughout.[7]

[6] We also occasionally write h_i^t for $(m_{i,\tau}, a_\tau)_{\tau=1}^{t-1}$, omitting the period t message $m_{i,t}$.

[7] Dhillon and Mertens (1996) show that the revelation principle fails for trembling-hand perfect equilibria. Nonetheless, with our "machine" interpretation of the mediator, the revelation principle applies for sequential equilibrium by precisely the usual argument of Forges (1986).

Finally, we say that a sequential equilibrium with mediated perfect monitoring is *on-path strict* if following the mediator's recommendation is strictly optimal for each player i at every on-path history h_i^t. Let $E_{\text{med}}(\delta)$ denote the set of on-path strict sequential equilibrium payoffs. For the rest of the paper, we slightly abuse terminology by omitting the qualifier "on-path" when discussing such equilibria.

3 An Illustrative (Counter)Example

The goal of this paper is to provide sufficient conditions for the equilibrium payoff set with mediated perfect monitoring to be a (recursive) upper bound on the equilibrium payoff set with private monitoring. Before giving these main results, we provide an illustrative example showing why, in the absence of our sufficient conditions, private monitoring (without a mediator) can outperform mediated perfect monitoring. Readers eager to get to the results can skip this section without loss of continuity.

Consider the repetition of the following stage game, with $\delta = \frac{1}{6}$:

	L	M	R
U	$2,2$	$-1,0$	$-1,0$
D	$3,0$	$0,0$	$0,0$
T	$0,3$	$6,-3$	$-6,-3$
B	$0,-3$	$0,3$	$0,3$

We show the following:

Proposition 1 *In this game, there is no sequential equilibrium where the players' per-period payoffs sum to more than* 3 *with perfect monitoring and a mediator, while there is such a sequential equilibrium with some private monitoring structure.*

Proof. See appendix. ■

In the constructed private monitoring structure in the proof of Proposition 1, players' payoffs are measurable with respect to their own actions and signals. In addition, a similar argument shows that imperfect public monitoring (with private strategies) can also outperform mediated perfect monitoring. This shows that some conditions are also required to guarantee that the sequential equilibrium payoff set with mediated perfect monitoring is an upper bound on the sequential equilibrium payoff set with imperfect public monitoring. Since imperfect public monitoring is a special case of imperfect private monitoring, our sufficient conditions for the private monitoring case are enough.

Here is a sketch of the proof of Proposition 1. Note that (U, L) is the only action profile where payoffs sum to more than 3. Because δ is so low, player 1 (row player, "she") can be induced to play U in response to L only if action profile (U, L) is *immediately* followed by (T, M) with high enough probability: specifically, this probability must exceed $\frac{3}{5}$. With perfect monitoring, player 2 (column player, "he") must then "see (T, M) coming" with probability at least $\frac{3}{5}$ following (U, L), and this probability is so high that player 2 will deviate from M to L (regardless of the specification of continuation play). This shows that payoffs cannot sum to more than 3 with perfect monitoring.

On the other hand, with private monitoring, player 2 may not know whether (U, L) has just occurred, and therefore may be unsure of whether the next action profile will be (T, M) or (B, M), which can give him the necessary incentive to play M rather than L. In particular, suppose that player 1 mixes $\frac{1}{3}U + \frac{2}{3}D$ in period 1, and the monitoring structure is such that player 2 gets signal m ("play M") with probability 1 following (U, L), and gets signals m and r ("play R") with probability $\frac{1}{2}$ each following (D, L). Suppose further that player 1 plays T in period 2 if she played U in period 1, and plays B in period 2 if she played D in period 1. Then, when player 2 sees signal m in period 1, his posterior belief that player 1 played U in period 1 is

$$\frac{\frac{1}{3}(1)}{\frac{1}{3}(1) + \frac{2}{3}\left(\frac{1}{2}\right)} = \frac{1}{2}.$$

Player 2 therefore expects to face T and B in period 2 with probability $\frac{1}{2}$ each, so he is willing to play M rather than L. Meanwhile, player 1 is always rewarded with (T, M) in period 2 when she plays U in period 1, so she is willing to play U (as well as D) in period 1.

To summarize, the advantage of private monitoring is that pooling players' information sets (in this case, player 2's information sets after (U, L) and (D, L)) can make providing incentives easier.[8]

To preview our results, we will show that private monitoring cannot outperform mediated perfect monitoring when there exists a feasible payoff vector that is appealing enough to both players that neither is tempted to deviate when it is promised to them in continuation if they conform. This condition is violated in the current example because, for example, no feasible continuation payoff for player 2 is high enough to induce him to respond to T with M rather than L. Specifically, our condition is satisfied in the current example if and only if δ is greater than $\frac{19}{25}$. Thus, in this example, private monitoring can outperform mediated perfect monitoring if $\delta = \frac{1}{6}$, but not if $\delta > \frac{19}{25}$.[9]

4 Preliminary Results about $E_{\text{med}}(\delta)$

We begin with two preliminary results about the equilibrium payoff set with mediated perfect monitoring. These results are important for both our result on when private monitoring cannot outperform mediated perfect monitoring (Theorem 1) and our characterization of the equilibrium payoff set with mediated perfect monitoring (Theorem 2).

Let $\underline{u}_i$ be player i's correlated minmax payoff, given by

$$\underline{u}_i = \min_{\alpha_{-i} \in \Delta(A_{-i})} \max_{a_i \in A_i} u_i(a_i, \alpha_{-i}).$$

Let $\alpha^*_{-i} \in \Delta(A_{-i})$ be a solution to this minmax problem. Let d_i be player i's greatest

[8] As far as we know, the observation that players can benefit from imperfections in monitoring even in the presence of a mediator is original. Examples by Kandori (1991), Sekiguchi (2002), Mailath, Matthews, and Sekiguchi (2002), and Miyahara and Sekiguchi (2013) show that players can benefit from imperfect monitoring in finitely repeated games. However, in their examples this conclusion relies on the absence of a mediator, and is thus due to the possibilities for correlation opened up by private monitoring. The broader point that giving players more information can be bad for incentives is of course an old one.

[9] We do not find the lowest possible discount factor $\bar{\delta}$ such that mediated perfect monitoring outperforms private monitoring for all $\delta > \bar{\delta}$.

possible gain from a deviation at any recommendation profile, given by

$$d_i = \max_{r \in A, a_i \in A_i} u_i(a_i, r_{-i}) - u_i(r).$$

Let $\underline{w}_i$ be the lowest continuation payoff such that player i does not want to deviate at any recommendation profile when she is minmaxed forever if she deviates, given by

$$\underline{w}_i = \underline{u}_i + \frac{1-\delta}{\delta} d_i.$$

Let

$$W_i = \left\{ w \in \mathbb{R}^{|I|} : w_i \geq \underline{w}_i \right\}.$$

Finally, let $u\left(A\right)$ be the convex hull of the set of feasible payoffs, and denote the interior of $\bigcap_{i \in I} W_i \cap u(A)$ as a subspace of $u\left(A\right)$ by

$$\mathring{W}^* \equiv \text{int} \left(\bigcap_{i \in I} W_i \cap u(A) \right)$$

with

$$W^* \equiv \bigcap_{i \in I} W_i \cap u(A)$$

Our first preliminary result is that all payoffs in $\mathring{W}^*$ are attainable in equilibrium with mediated perfect monitoring. See Figure 1. The intuition is that—as discussed in the introduction—the mediator can virtually implement any payoff vector in W^* by minmaxing deviators.

We will actually prove the slightly stronger result that all payoffs in $\mathring{W}^*$ are attainable in a strict "full-support" equilibrium with mediated perfect monitoring. Formally, we say that an equilibrium has *full support* if for each player i and history $h_i^t = (r_{i,\tau}, a_\tau)_{\tau=1}^{t-1}$ such that there exists $(r_{-i,\tau})_{\tau=1}^{t-1}$ with $\Pr^\mu(r_\tau | (r_{\tau'}, a_{\tau'})_{\tau'=1}^{\tau-1}) > 0$ for each $\tau = 1, ..., t-1$, there exists

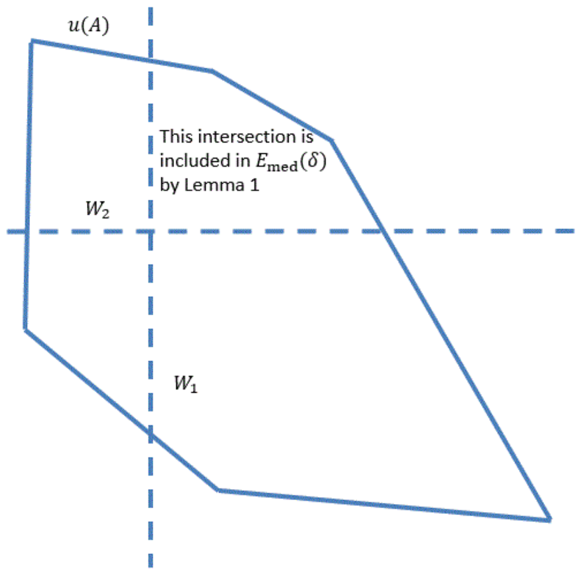

Figure 1: $W^* \subseteq E_{\text{med}}(\delta)$

$(\bar{r}_{-i,\tau})_{\tau=1}^{t-1}$ such that for each $\tau = 1, ..., t-1$ we have

$$\Pr^{\mu}(r_{i,\tau}, \bar{r}_{-i,\tau} | (r_{i,\tau'}, \bar{r}_{-i,\tau'}, a_{\tau'})_{\tau'=1}^{\tau-1}) > 0 \text{ and } \bar{r}_{-i,\tau} = a_{-i,\tau}.$$

That is, any history h_i^t consistent with the mediator's strategy is also consistent with i's opponents' equilibrium strategies (even if player i herself has deviated, noting that we allow $r_{i,\tau} \neq a_{i,\tau}$ in h_i^t). This is weaker than requiring that the mediator's recommendation has full support at all histories (on- and off-path), but stronger than requiring that the recommendation has full support at all on-path histories only. Note that, if the equilibrium has full support, player i never believes that any of the other players has deviated.

Lemma 1 *For all $v \in \mathring{W}^*$, there exists a strict full-support equilibrium with mediated perfect monitoring with payoff v. In particular, $\mathring{W}^* \subseteq E_{\text{med}}(\delta)$.*

Proof. For each $v \in \mathring{W}^*$, there exists $\mu \in \Delta(A)$ such that $u(\mu) = v$ and $\mu(r) > 0$ for all $r \in A$. On the other hand, for each $i \in I$ and $\varepsilon \in (0,1)$, approximate the minmax strategy α_{-i}^* by the full-support strategy $\alpha_{-i}^{\varepsilon} \equiv (1-\varepsilon)\,\alpha_{-i}^* + \varepsilon \sum_{a_{-i} \in A_{-i}} \frac{a_{-i}}{|A_{-i}|}$. Since $v \in \text{int}\left(\bigcap_{i \in I} W_i\right)$, there exists $\varepsilon \in (0,1)$ such that, for each $i \in I$, we have

$$v_i > \max_{a_i \in A_i} u_i(a_i, \alpha^{\varepsilon}_{-i}) + \frac{1-\delta}{\delta} d_i. \tag{1}$$

Consider the following recommendation schedule: The mediator follows an automaton strategy whose state is identical to a subset of players $J \subseteq I$. Hence, the mediator has $2^{|I|}$ states. In the following construction of the mediator's strategy, J will represent the set of players who have ever deviated from the mediator's recommendation.

If the state J is equal to $\emptyset$ (no player has deviated), then the mediator recommends μ. If there exists i with $J = \{i\}$ (only player i has deviated), then the mediator recommends r_{-i} to players $-i$ according to $\alpha^{\varepsilon}_{-i}$, and recommends some best response to $\alpha^{\varepsilon}_{-i}$ to player i. Finally, if $|J| \geq 2$ (several players have deviated), then for each $i \in J$, the mediator recommends the best response to $\alpha^{\varepsilon}_{-i}$, while she recommends each profile $a_{-J} \in A_{-J}$ to the other players $-J$ with probability $\frac{1}{|A_{-J}|}$. The state transitions as follows: if the current state is J and players J' deviate, then the state transitions to $J \cup J'$.[10]

Player i's strategy is to follow her recommendation $r_{i,t}$ in period t. She believes that the mediator's state is $\emptyset$ if she herself has never deviated, and believes that the state is $\{i\}$ if she has deviated.

Since the mediator's recommendation has full support, player i's belief is consistent. (In particular, no matter how many times player i has been instructed to minmax some player j, it is always infinitely more likely that these instructions resulted from randomization by the mediator rather than a deviation by player j.) If player i has deviated, then (given her belief) it is optimal for her to always play a static best response to $\alpha^{\varepsilon}_{-i}$, since the mediator always recommends $\alpha^{\varepsilon}_{-i}$ in state $\{i\}$. Given that a unilateral deviation by player i is punished in this way, (1) implies that on path player i has a strict incentive to follow her recommendation $r_{i,t}$ at any recommendation profile $r_t \in A$. Hence, she has a strict incentive to follow her recommendation when she believes that $r_{-i,t}$ is distributed according to $\Pr^{\mu}(r_{-i,t}|h_i^t)$. ■

The condition that $\mathring{W}^* \neq \emptyset$ can be more transparently stated as a lower bound on the

[10] We thank Gabriel Carroll for suggestions which helped simplify this construction.

discount factor. In particular, $\mathring{W}^* \neq \emptyset$ if and only if there exists $v \in u(A)$ such that

$$v_i > \underline{u}_i + \frac{1-\delta}{\delta} d_i \text{ for all } i \in I,$$

or equivalently

$$\delta > \delta^* \equiv \min_{v \in u(A)} \max_{i \in I} \frac{d_i}{d_i + v_i - \underline{u}_i}.$$

For instance, it can be checked that $\delta^* = \frac{19}{25}$ in the example of Section 3. Note that δ^* is strictly less than 1 if and only if the stage game admits a feasible and strictly individually rational payoff vector (relative to correlated minmax payoffs).[11] For most games of interest, δ^* will be some "intermediate" discount factor that is not especially close to either 0 or 1.

Our second preliminary result is that, if a strict full-support equilibrium exists, then any payoff vector that can be attained by a mediator's strategy that is incentive compatibility on path is (virtually) attainable in strict equilibrium. The intuition is that mixing such a mediator's strategy with an ε probability of playing any strict full-support equilibrium yields a strict equilibrium with nearby payoffs.

Lemma 2 *With mediated perfect monitoring, fix a payoff vector v, and suppose there exists a mediator's strategy μ that (1) attains v when players obey the mediator, and (2) has the property that obeying the mediator is optimal for each player at each on-path history, when she is minmaxed forever if she deviates: that is, for each player i and on-path history h_m^{t+1},*

$$\begin{aligned} & (1-\delta)\mathbb{E}\left[u_i(r_t) \mid h_m^t, r_{i,t}\right] + \delta \mathbb{E}\left[(1-\delta)\sum_{\tau=t+1}^{\infty} \delta^{\tau-t-1} u_i(\mu(h_m^\tau)) \mid h_m^t, r_{i,t}\right] \\ \geq \; & \max_{a_i \in A_i} (1-\delta)\mathbb{E}\left[u_i(a_i, r_{-i,t}) \mid h_m^t, r_{i,t}\right] + \delta \underline{u}_i. \end{aligned} \tag{2}$$

Suppose also that there exists a strict full-support equilibrium. (For example, such an equilibrium exists if $\mathring{W}^* \neq \emptyset$, by Lemma 1.) Then $v \in \overline{E_{\text{med}}(\delta)}$.

Proof. See appendix. ■

[11] Recall that a payoff vector v is *strictly individually rational* if $v_i > \underline{u}_i$ for all $i \in I$.

5 A Sufficient Condition for $\overline{E_{\text{med}}(\delta)}$ to Give an Upper Bound

Our sufficient condition for mediated perfect monitoring to outperform private monitoring in two-player games is that $\delta > \delta^*$. In Section 9, we discuss what happens when there are more than two players or the condition that $\delta > \delta^*$ is relaxed.

Let $E(\delta, p)$ be the set of (possibly weak) sequential equilibrium payoffs with private monitoring structure p. The following is our first main result. Note that $E(\delta, p)$ is closed, as we use the product topology on assessments (Fudenberg and Levine, 1983), so the maxima in the theorem exist.

Theorem 1 *If $|I| = 2$ and $\delta > \delta^*$, then for every private monitoring structure p and every non-negative Pareto weight $\lambda \in \Lambda_+ \equiv \{\lambda \in \mathbb{R}^2_+ : \|\lambda\| = 1\}$, we have*

$$\max_{v \in E(\delta,p)} \lambda \cdot v \leq \max_{v \in \overline{E_{\text{med}}(\delta)}} \lambda \cdot v.$$

Theorem 1 says that, in games involving two players of at least moderate patience, the Pareto frontier of the (closure of the strict) equilibrium payoff set with mediated perfect monitoring extends farther in any non-negative direction than does the Pareto frontier of the equilibrium payoff set with any private monitoring structure.[12] We emphasize that—like all our main results—Theorem 1 concerns equilibrium payoff sets at fixed discount factors, not in the $\delta \to 1$ limit.

We describe the idea of the proof of Theorem 1, deferring the proof itself to Section 8.

Let $\mathcal{E}(\delta)$ be the equilibrium payoff set in the mediated repeated game with the following *universal information structure*: the mediator directly observes the recommendation profile r_t and the action profile a_t in each period t, while each player i observes nothing beyond her own recommendation $r_{i,t}$ and her own action $a_{i,t}$.[13] With this monitoring structure, the

[12] We do not know if the same result holds for negative Pareto weights.

[13] This information structure may not result from mediated communication among the players, as actions are not publicly observed. Again, we simply view $\mathcal{E}(\delta)$ as a technical device.

mediator can clearly replicate any private monitoring structure p by setting $\mu(h_m^t)$ equal to $p(\cdot|a_{t-1})$ for every history $h_m^t = (r_\tau, a_\tau)_{\tau=1}^{t-1}$. It particular, we have $E(\delta, p) \subseteq \mathcal{E}(\delta)$ for every p,[14] so to prove Theorem 1 it suffices to show that

$$\max_{v \in \mathcal{E}(\delta)} \lambda \cdot v \leq \max_{v \in \overline{E_{\text{med}}(\delta)}} \lambda \cdot v. \tag{3}$$

(The maximum over $v \in \mathcal{E}(\delta)$ is well defined since $\mathcal{E}(\delta)$ is closed, by the same reasoning as for $E(\delta, p)$.)

To show this, the idea is to start with an equilibrium in $\mathcal{E}(\delta)$—where players only observe their own recommendations—and then show that the players' recommendations can be "publicized" without violating anyone's obedience constraints.[15] To see why this is possible (when $|I| = 2$ and $\delta > \delta^*$, or equivalently $\mathring{W}^* \neq \emptyset$), first note that we can restrict attention to equilibria with Pareto-efficient on-path continuation payoffs, as improving both players' on-path continuation payoffs improves their incentives (assuming that deviators are minmaxed, which is possible when $\mathring{W}^* \neq \emptyset$, by Lemma 2). Next, if $|I| = 2$ and $\mathring{W}^* \neq \emptyset$, then if a Pareto-efficient payoff vector v lies outside of W_i for one player (say player 2), it must then lie inside of W_j for the other player (player 1). Hence, at each history h^t, there can be only one player—here player 2—whose obedience constraint could be violated if we publicized both players' past recommendations.

Now, suppose that at history h^t we do publicize the entire vector of players' past recommendations $r^t = (r_\tau)_{\tau=1}^{t-1}$, but the mediator then issues period t recommendations according to the original equilibrium distribution of recommendations conditional on *player 2's past recommendations* $r_2^t = (r_{2,\tau})_{\tau=1}^{t-1}$ *only.* We claim that doing this violates neither player's obedience constraint: Player 1's obedience constraint is easy to satisfy, as we can always ensure that continuation payoffs lie in W_1. And, since player 2 already knew r_2^t in the

[14] In light of this fact, the reader may wonder why we do not simply take $\mathcal{E}(\delta)$ rather than $\overline{E_{\text{med}}(\delta)}$ to be our bound on equilibrium payoffs with private monitoring. The answer is that, as far as we know, $\mathcal{E}(\delta)$ does not admit a recursive characterization, while we recursively characterize $\overline{E_{\text{med}}(\delta)}$ in Section 6.

[15] More precisely, the construction in the proof both publicizes the players' recommendations and modifies the equilibrium in ways that only improve the players' λ-weighted payoffs.

original equilibrium, publicizing h^t while issuing recommendations based only on r_2^t does not affect his incentives.

An important missing step in this proof sketch is that, in the original equilibrium in $\mathcal{E}(\delta)$, at some histories it may be player *1* who is tempted to deviate when we publicize past recommendations (while it is player 2 who is tempted at other histories). For instance, it is not clear how we can publicize past recommendations when ex ante equilibrium payoffs are very good for player 1 and so $v \in W_1$ (so player 2 is tempted to deviate in period 1), but continuation payoffs at some later history are very good for player 2 (so then player 1 is tempted to deviate). The proof of Theorem 1 shows that we can ignore this possibility, because—somewhat unexpectedly—equilibrium paths like this one are never needed to sustain Pareto-efficient payoffs. In particular, to sustain an ex ante payoff that is very good for player 1 (i.e., outside W_2), we never need to promise continuation payoffs that are very good for player 2 (i.e., outside W_1). The intuition is that, rather than promising player 2 a very good continuation payoff outside W_1, we can instead promise him a fairly good continuation inside W_1, while compensating him for this change by also occasionally transitioning to this fairly good continuation payoff at histories where the original promised continuation payoff is less good for him. Finally, since the feasible payoff set is convex, the resulting "compromise" continuation payoff vector is also acceptable to player 1.

6 Recursively Characterizing $\overline{E_{\text{med}}(\delta)}$

We have seen that $\overline{E_{\text{med}}(\delta)}$ is an upper bound on $E(\delta, p)$ for two-player games satisfying $\delta > \delta^*$. As our goal is to give a recursive upper bound on $E(\delta, p)$, it remains to recursively characterize $\overline{E_{\text{med}}(\delta)}$. Our characterization assumes that $\delta > \delta^*$, but it applies for any number of players.

Recall that APS characterize the perfect public equilibrium set with imperfect public monitoring as the iterative limit of a generating operator B, where $B(W)$ is defined as the set of payoffs that can be sustained when on- and off-path continuation payoffs are drawn from W. What we show is that the sequential equilibrium payoff set with mediated perfect monitoring is the iterative limit of a generating operator $\tilde{B}$, where $\tilde{B}(W)$ is the set of payoffs that can be sustained when on-path continuation payoffs are drawn from W, and deviators are minmaxed off path. Intuitively, there are two things to show: (1) we can indeed minmax deviators off path, and (2) on-path continuation payoffs must themselves be sequential equilibrium payoffs. The first of these facts is Lemma 2. For the second, note that, in an obedient equilibrium with perfect monitoring, players can perfectly infer each other's private history on path. Continuation play at on-path histories (but not off-path histories) is therefore "common knowledge," which gives the desired recursive structure.

In the following formal development, we assume familiarity with APS (see also Section 7.3 of Mailath and Samuelson, 2006), and focus on the new features that emerge when mediation is available.

Definition 1 *For any set $V \subseteq \mathbb{R}^{|I|}$, a correlated action profile $\alpha \in \Delta(A)$ is* minmax-threat enforceable *on V by a mapping $\gamma : A \to V$ if, for each player i and action $a_i \in \operatorname{supp} \alpha_i$,*

$$\begin{aligned} & \mathbb{E}^{\alpha}\left[(1-\delta) u_i(a_i, a_{-i}) + \delta\gamma(a_i, a_{-i}) \mid a_i\right] \\ \geq\; & \max_{a_i' \in A_i} \mathbb{E}^{\alpha}\left[(1-\delta) u_i(a_i', a_{-i}) \mid a_i\right] + \delta \min_{\alpha_{-i} \in \Delta(A_{-i})} \max_{a_i \in A_i} u_i(a_i, \alpha_{-i}). \end{aligned}$$

Definition 2 *A payoff vector $v \in \mathbb{R}^{|I|}$ is* minmax-threat decomposable *on V if there exists a correlated action profile $\alpha \in \Delta(A)$ which is minmax-threat enforced on V by a mapping γ such that*

$$v = \mathbb{E}^{\alpha}\left[(1-\delta) u(a) + \delta\gamma(a)\right].$$

Let $\tilde{B}(V) = \left\{v \in \mathbb{R}^{|I|} : v \text{ is minmax-threat decomposable on } V\right\}$.

We show that the following algorithm recursively computes $\overline{E_{\text{med}}(\delta)}$: let $W^1 = u(A)$, $W^n = \tilde{B}(W^{n-1})$ for $n > 1$, and $W^\infty = \lim_{n\to\infty} W^n$.

Theorem 2 *If $\delta > \delta^*$, then $\overline{E_{\text{med}}(\delta)} = W^\infty$.*

With the exception of the following two lemmas, the proof of Theorem 2 is entirely standard, and omitted. The lemmas correspond to facts (1) and (2) above. In particular, Lemma 3 follows directly from APS and Lemma 2, while Lemma 4 establishes on-path recursivity. For both lemmas, assume $\delta > \delta^*$.

Lemma 3 *If a set $V \subseteq \mathbb{R}^{|I|}$ is bounded and satisfies $V \subseteq \tilde{B}(V)$, then $\tilde{B}(V) \subseteq \overline{E_{\text{med}}(\delta)}$.*

Proof. See appendix. ■

Lemma 4 $\overline{E_{\text{med}}(\delta)} = \tilde{B}\left(\overline{E_{\text{med}}(\delta)}\right)$.

Proof. See appendix. ■

Combining Theorems 1 and 2 yields our main conclusion: *in two-player games with $\delta > \delta^*$, the equilibrium payoff set with mediated perfect monitoring is a recursive upper bound on the equilibrium payoff set with any imperfect private monitoring structure.*

7 The Upper Bound in an Example

We illustrate our results with an application to a simple repeated Bertrand game. Specifically, in such a game we find greatest equilibrium payoff that each firm can attain for any private monitoring structure.

Consider the following Bertrand game: there are two firms $i \in \{1, 2\}$, and each firm i's possible price level is $p_i \in \{W, L, M, H\}$ (price war, low price, medium price, high price). Given p_1 and p_2, firm i's profit is determined by the following payoff matrix

	W	L	M	H
W	$15, 15$	$30, 25$	$50, 15$	$80, 0$
L	$25, 30$	$40, 40$	$60, 35$	$90, 15$
M	$15, 50$	$35, 60$	$55, 55$	$85, 35$
H	$0, 80$	$15, 90$	$35, 85$	$65, 65$

Note that L (low price) is a dominant strategy in the stage game, W (price war) is a costly action that hurts the other firm, and (H, H) maximizes the sum of the firms' profits. The feasible payoff set is given by

$$u(A) = \text{co}\{(0, 80), (15, 15), (15, 90), (35, 85), (65, 65), (80, 0), (85, 35), (90, 15)\}.$$

In addition, each firm's minmax payoff $\underline{u}_i$ is 25, so the feasible and individually rational payoff set is given by

$$\text{co}\{(25, 25), (25, 87.5), (35, 85), (65, 65), (85, 35), (87.5, 25)\}.$$

In particular, the greatest feasible and individually rational payoff for each firm is 87.5.

In this game, each firm's maximum deviation gain d_i is 25. Since the game is symmetric, the critical discount factor δ^* above which we can apply Theorems 1 and 2 is given by plugging the best symmetric payoff of 65 into the formula for δ^*, which gives

$$\delta^* = \frac{25}{25 + 65 - 25} = \frac{5}{13}.$$

To illustrate our results, we find the greatest equilibrium payoff that each firm can attain for any private monitoring structure when $\delta = \frac{1}{2} > \frac{5}{13}$.

When $\delta = \frac{1}{2}$, we have

$$\underline{w}_i = 25 + \frac{1-\delta}{\delta} 25 = 50.$$

Hence, Lemma 1 implies that

$$\begin{aligned} & \{v \in \text{int } u(A) : v_i > 50 \text{ for each } i\} \\ = \; & \text{int co}\{(50, 50), (75, 50), (50, 75), (65, 65)\} \subseteq E_{\text{med}}(\delta). \end{aligned}$$

We now compute the best payoff vector for firm 1 in $\tilde{B}(u(A))$. By Theorems 1 and 2, any Pareto-efficient payoff profile not included $\tilde{B}(u(A))$ is not included in $E(\delta, p)$ for any p.

In computing the best payoff vector for firm 1, it is natural to conjecture that firm 1's incentive compatibility constraint is not binding. We thus consider a relaxed problem with only firm 2's incentive constraint, and then verify that firm 1's incentive constraint is satisfied. Note that playing L is always the best deviation for firm 2. Furthermore, the corresponding deviation gain decreases as firm 1 increases its price from W to L, and (weakly) increases as it increases its price from L to M or H. On the other hand, firm 1's payoff increases as firm 1 increases its price from W to L and decreases as it increases its price from L to M or H. Hence, in order to maximize firm 1's payoff, firm 1 should play L.

Suppose that firm 2 plays H. Then, firm 2's incentive compatibility constraint is

$$(1-\delta)\underbrace{25}_{\text{maximum deviation gain}} \leq \delta(w_2 - \underbrace{25}_{\text{minmax payoff}}),$$

where w_2 is firm 2's continuation payoff. That is, $w_2 \geq 50$.

By feasibility, $w_2 \geq 50$ implies that $w_1 \leq 75$. Hence, if $r_2 = H$, the best minmax-threat decomposable payoff for firm 1 is

$$(1-\delta)\begin{pmatrix} 90 \\ 15 \end{pmatrix} + \delta \begin{pmatrix} 75 \\ 50 \end{pmatrix} = \begin{pmatrix} 82.5 \\ 32.5 \end{pmatrix}.$$

Since 82.5 is larger than any payoff that firm 1 can get when firm 2 plays W, M, or L, firm 2 should indeed play H to maximize firm 1's payoff. Moreover, since $75 \geq \underline{w}_1$, firm 1's incentive constraint is not binding. Thus, we have shown that 82.5 is the best payoff for firm 1 in $\tilde{B}(u(A))$.

On the other hand, with mediated perfect monitoring it is indeed possible to (virtually) implement an action path in which firm 1's payoff is 82.5: play (L, H) in period 1 (with payoffs $(90, 15)$), and then play $\frac{1}{2}(M, H) + \frac{1}{2}(H, H)$ forever (with payoffs $\frac{1}{2}(85, 35) + \frac{1}{2}(65, 65) = (75, 50)$), while minmaxing deviators.

Thus, when $\delta = \frac{1}{2}$, each firm's greatest feasible and individually rational payoff is 87.5, but the greatest payoff it can attain with any imperfect private monitoring structure is only

82.5. In this simple game, we can therefore say exactly how much of a constraint is imposed on each firm's greatest equilibrium payoff by the firms' impatience alone, independently of the monitoring structure.

8 Proof of Theorem 1

8.1 Preliminaries and Plan of Proof

We wish to establish (3) for every Pareto weight $\lambda \in \Lambda_+$. Note that Lemma 1 implies that

$$W^* \subseteq \overline{E_{\text{med}}(\delta)}.$$

Therefore, for every Pareto weight $\lambda \in \Lambda_+$, if there exists $v \in \arg\max_{v' \in \mathcal{E}(\delta)} \lambda \cdot v'$ such that $v \in W^*$, then there exists $v^* \in \overline{E_{\text{med}}(\delta)}$ such that $\lambda \cdot v \leq \lambda \cdot v^*$, as desired.

Hence, we are left to consider $\lambda \in \Lambda_+$ with

$$\arg\max_{v' \in \mathcal{E}(\delta)} \lambda \cdot v' \cap W^* = \emptyset. \tag{4}$$

Since we consider two-player games, we can order $\lambda \in \Lambda_+$ as follows: $\lambda \leq \lambda'$ if and only if $\frac{\lambda_1}{\lambda_2} \leq \frac{\lambda_1'}{\lambda_2'}$, that is, the vector λ is steeper than λ'. For each player i, let $\bar{w}^i$ be the Pareto-efficient point in W_i satisfying

$$\bar{w}^i \in \arg\max_{v \in W_i \cap u(A)} v_{-i}.$$

Note that the assumption that $\mathring{W}^* \neq \emptyset$ implies that $\bar{w}^i \in W^*$. Let $\alpha^i \in \Delta(A)$ be a recommendation that attains $\bar{w}^i$: $u(\alpha^i) = \bar{w}^i$. Let Λ^i be the (non-empty) set of Pareto weight λ^i such that $\bar{w}^i \in \arg\max_{v \in u(A)} \lambda^i \cdot v$:

$$\Lambda^i = \left\{ \lambda^i \in \mathbb{R}^2_+ : \left\| \lambda^i \right\| = 1, \bar{w}^i \in \arg\max_{v \in u(A)} \lambda^i \cdot v \right\}.$$

As $u\left(A\right)$ is convex, if λ satisfies (4) then either $\lambda < \lambda^1$ for each $\lambda^1 \in \Lambda^1$ or $\lambda > \lambda^2$ for each

$\lambda^2 \in \Lambda^2$. See Figure 2. We focus on the case where $\lambda > \lambda^2$. (The proof for the $\lambda < \lambda^1$ case is symmetric and thus omitted.)

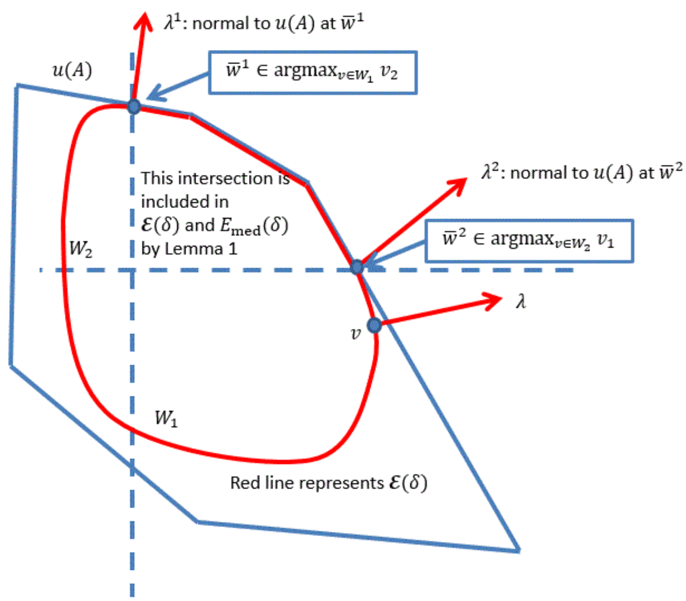

Figure 2: Setup for the construction.

Fix $v \in \arg\max_{v' \in \mathcal{E}(\delta)} \lambda \cdot v'$. Let $(\mu, (\sigma_i)_{i \in I})$ be an equilibrium that attains v with the universal information structure (where players do not observe each other's actions). By Lemma 2, it suffices to construct a mediator's strategy μ^* yielding payoffs v^* such that (2) ("perfect monitoring incentive compatibility") holds and $\lambda \cdot v \leq \lambda \cdot v^*$. The rest of the proof constructs such a strategy.

The plan for constructing the strategy μ^* is as follows: First, from μ, we construct a mediator's strategy $\bar{\mu}$ that yields payoffs v and satisfies perfect monitoring incentive compatibility for player 2, but possibly not for player 1. The idea is to set the distribution of recommendations under $\bar{\mu}$ equal to the distribution of recommendations under μ conditional on player 2's information only. Second, from $\bar{\mu}$, we construct a mediator's strategy μ^* that yields payoffs v^* with $\lambda \cdot v \leq \lambda \cdot v^*$ and satisfies perfect monitoring incentive compatibility for both players.

8.2 Construction and Properties of $\bar{\mu}$

For each on-path history of player 2's recommendations, denoted by $r_2^t = (r_{2,\tau})_{\tau=1}^{t-1}$, let $\Pr^{\mu}(\cdot|r_2^t)$ be the conditional distribution of recommendations in period t; and let $w^{\mu}(r_2^t)$ be the continuation payoff vector from period t onward conditional on r_2^t:

$$w^{\mu}(r_2^t) = \mathbb{E}^{\mu}\left[\sum_{\tau=0}^{\infty} \delta^{\tau} u(r_{t+\tau}) \mid r_2^t\right].$$

Define $\bar{\mu}$ so that, for every on-path history $r^t = (r_\tau)_{\tau=1}^{t-1}$, the mediator draws r_t according to $\Pr^{\mu}(r_t|r_2^t)$:

$$\Pr^{\bar{\mu}}(r_t|r^t) \equiv \Pr^{\mu}(r_t|r_2^t). \tag{5}$$

We claim that $\bar{\mu}$ yields payoffs v and satisfies (2) for player 2. To see this, let $w^{\bar{\mu}}(r^t)$ be the continuation payoff vector from period t onward conditional on r^t under $\bar{\mu}$, and note that $w^{\bar{\mu}}(r^t) = w^{\mu}(r_2^t)$. In particular, $w^{\bar{\mu}}(r^1) = w^{\mu}(r_2^1) = v$. In addition, the fact that μ is an equilibrium with the universal information structure implies that, for every on-path history r^{t+1},

$$(1-\delta)\,\mathbb{E}^{\mu}\left[u_2(r_t)\,|r_2^{t+1}\right] + \delta w^{\mu}\left(r_2^{t+1}\right) \geq \max_{a_2 \in A_2} (1-\delta)\,\mathbb{E}^{\mu}\left[u_2(r_{1,t}, a_2)\,|r_2^{t+1}\right] + \delta \underline{u}_2.$$

As $w_2^{\mu}\left(r_2^{t+1}\right) = w_2^{\bar{\mu}}(r^{t+1})$ and $\Pr^{\mu}\left(r_t|r_2^{t+1}\right) = \Pr^{\bar{\mu}}(r_t|r^t, r_{2,t})$, this implies that (2) holds for player 2.

8.3 Construction of μ^*

The mediator's strategy μ^* will involve mixing over continuation payoffs at certain histories r^{t+1}, and we will denote the mixing probability at history r^{t+1} by $p(r^{t+1})$. Our approach is to first construct the mediator's strategy μ^* for an arbitrary function $p : \bigcup_{t=1}^{\infty} A^{t-1} \to [0,1]$ specifying these mixing probabilities, and to then specify the function p.

Given a function $p : \bigcup_{t=1}^{\infty} A^{t-1} \to [0,1]$, the mediator's strategy μ^* is defined as follows:

In each period $t = 0, 1, 2, \ldots$, the mediator is in one of two states, $\omega_t \in \{S_1, S_2\}$ (where "period 0" is a purely notational, and as usual the game begins in period 1). Given the state, recommendations in period $t \geq 1$ are as follows:

1. In state S_1, at history $r^t = (r_\tau)_{\tau=1}^{t-1}$, the mediator recommends r_t according to $\Pr^{\bar{\mu}}(r_t|r^t)$.

2. In state S_2, the mediator recommends r_t according to some $\alpha^1 \in \Delta(A)$ such that $u(\alpha^1) = \bar{w}^1$.

The initial state is $\omega_0 = S_1$. State S_2 is absorbing: if $\omega_t = S_2$ then $\omega_{t+1} = S_2$. Finally, the transition rule in state S_1 is as follows:

1. If $w^{\bar{\mu}}(r^{t+1}) \notin W_1$, then $\omega_{t+1} = S_2$ with probability one.

2. If $w^{\bar{\mu}}(r^{t+1}) \in W_1$, then $\omega_{t+1} = S_2$ with probability $1 - p(r^{t+1})$.

Thus, strategy μ^* agrees with $\bar{\mu}$, with the exception that μ^* occasionally transitions to an absorbing state where actions yielding payoffs $\bar{w}^1$ are recommended forever. In particular, such a transition always occurs when continuation payoffs under $\bar{\mu}$ lie outside W_1, and otherwise this transition occurs with probability $1 - p\left(r^{t+1}\right)$.

To complete the construction of μ^*, it remains only to specify the function p. To this end, it is useful to define an operator F, which maps functions $w : \bigcup_{t=1}^{\infty} A^{t-1} \to \mathbb{R}^2$ to functions $F(w_2) : \bigcup_{t=1}^{\infty} A^{t-1} \to \mathbb{R}^2$. The operator F will be defined so that its unique fixed point is precisely the continuation value function in state S_1 under μ^* for a particular function p, and this function will be the one we use to complete the construction of μ^*.

Given $w : \bigcup_{t=1}^{\infty} A^{t-1} \to \mathbb{R}^2$, define $w^*(w) : \bigcup_{t=1}^{\infty} A^{t-1} \to \mathbb{R}$ so that, for every $r^t \in A^{t-1}$, we have

$$w^*(w)(r^t) = (1 - \delta)\, u\left(\bar{\mu}(r^t)\right) + \delta \mathbb{E}\left[w(r^{t+1})|r^t\right]. \tag{6}$$

On the other hand, given $w^*(w) : \bigcup_{t=1}^{\infty} A^{t-1} \to \mathbb{R}$, define $F(w) : \bigcup_{t=1}^{\infty} A^{t-1} \to \mathbb{R}$ so that, for every $r^t \in A^{t-1}$, we have

$$F(w)\left(r^t\right) = 1_{\{w^{\bar{\mu}}(r^t)\in W_1\}} \left\{ \begin{array}{c} p(w)(r^t) \times w^*(w)(r^t) \\ +(1-p(w)(r^t)) \times \bar{w}^1 \end{array} \right\} + 1_{\{w^{\bar{\mu}}(r^t)\notin W_1\}}\bar{w}^1, \qquad (7)$$

where, when $w^{\bar{\mu}}(r^t) \in W_1$, $p(w)(r^t)$ is the largest number in $[0,1]$ such that

$$p(w)(r^t) \times w_2^*(w)(r^t) + \left(1 - p(w)(r^t)\right) \times \bar{w}_2^1 \geq w_2^{\bar{\mu}}(r^t). \qquad (8)$$

That is, if $w_2^*(w)\left(r^t\right) \geq w_2^{\bar{\mu}}(r^t)$, then $p(w)(r^t) = 1$; and otherwise, since $w^{\bar{\mu}}(r^t) \in W_1$ implies that $w_2^{\bar{\mu}}(r^t) \leq \bar{w}_2^1$, $p(w)(r^t) \in [0,1]$ solves

$$p(w)(r^t) \times w_2^*(w)(r^t) + \left(1 - p(w)(r^t)\right) \times \bar{w}_2^1 = w_2^{\bar{\mu}}(r^t).$$

(Intuitively, the term $1_{\{w^{\bar{\mu}}(r^t)\notin W_1\}}\bar{w}^1$ in (7) reflects the fact that we have replaced continuation payoffs outside of W_1 with player 2's most favorable continuation payoff within W_1, namely $\bar{w}^1$. This replacement may reduce player 2's value below his original value of $w_2^{\bar{\mu}}(r^t)$. However, (8) ensures that, by also replacing continuation payoffs *within* W_1 with $\bar{w}^1$ with high enough probability, player 2's value does not fall below $w_2^{\bar{\mu}}(r^t)$.)

To show that F has a unique fixed point, it suffices to show that F is a contraction.

Lemma 5 *For all w and $\tilde{w}$, we have $\|F(w) - F(\tilde{w})\| \leq \delta \|w - \tilde{w}\|$, where $\|w - \tilde{w}\| \equiv \sup_{r^t} \|w(r^t) - \tilde{w}(r^t)\|$.*

Proof. See appendix. ■

Let w be the unique fixed point of F. Given this function w, let $w^* = w^*(w)$ (given by (6)) and let $p = p(w)$ (given by (8)). This completes the construction of the mediator's strategy μ^*.

8.4 Properties of μ^*

Observe that

$$w^*(r^t) = (1-\delta)\, u\left(\bar{\mu}(r^t)\right) + \delta \mathbb{E}\left[w(r^{t+1})|r^t\right] \tag{9}$$

and

$$w\left(r^t\right) = 1_{\{w^{\bar{\mu}}(r^t)\in W_1\}} \left\{p(r^t)w^*(r^t) + \left(1-p(r^t)\right)\bar{w}^1\right\} + 1_{\{w^{\bar{\mu}}(r^t)\notin W_1\}}\bar{w}^1. \tag{10}$$

Thus, for $i = 1, 2$, $w_i^*(r^t)$ is player i's expected continuation payoff from period t given r^t and $\omega_t = S_1$ (before she observes $r_{i,t}$), and $w_i(r^t)$ is player i's expected continuation payoff from period t given r^t and $\omega_{t-1} = S_1$ (before she observes $r_{i,t}$). In particular, recalling that $\omega_1 = S_1$ and $v = w^{\mu}(\emptyset) \in W_1$, (8) implies that the ex ante payoff vector v^* is given by

$$v^* = w(\emptyset) = p(\emptyset)w^*(\emptyset) + (1-p(\emptyset))\,\bar{w}^1.$$

We prove the following key lemma in the appendix.

Lemma 6 *For all $t \geq 1$, if $w^{\bar{\mu}}(r^t) \in W_1$, then $p(r^t)w^*(r^t) + (1-p(r_2^t))\,\bar{w}^1$ Pareto dominates $w^{\bar{\mu}}(r^t)$.*

Here is a graphical explanation of Lemma 6: By (9), $w^*(r^t) - w^{\bar{\mu}}(r^t)$ is parallel to $w(r^{t+1}) - w^{\bar{\mu}}(r^{t+1})$. To evaluate this difference, consider (10) for period $t+1$. The term $1_{\{w^{\bar{\mu}}(r^{t+1})\notin W_1\}}\bar{w}^1$ indicates that we construct $w(r^{t+1})$ by replacing some continuation payoff not included in W_1 with $\bar{w}^1$. Hence, $w(r^{t+1}) - w^{\bar{\mu}}(r^{t+1})$ (and thus $w^*(r^t) - w^{\bar{\mu}}(r^t)$) is parallel to $\bar{w}^1 - \hat{w}(r^{t+1})$ for some $\hat{w}(r^{t+1}) \in u(A) \setminus W_1$. See Figure 3 for an illustration.

Recall that $p(r^t)$ is determined by (8). Since the vector $w^*(r^t) - w^{\bar{\mu}}(r^t)$ is parallel to $\bar{w}^1 - \hat{w}(r^{t+1})$ for some $\hat{w}(r^{t+1}) \in u(A) \setminus W_1$ and $u(A)$ is convex, we have $w_1^*(r^t) \geq w_1^{\bar{\mu}}(r_2^t)$. Hence, if we take $p(r^t)$ so that the convex combination of $w_2^*(r^t)$ and $\bar{w}_2^1$ is equal to $w_2^{\bar{\mu}}(r^t)$, then player 1 is better off compared to $w_1^{\bar{\mu}}(r^t)$. See Figure 4.

Given Lemma 6, we show that μ^* satisfies perfect monitoring incentive compatibility ((2)) for both players, and $\lambda \cdot v \leq \lambda \cdot v^*$.

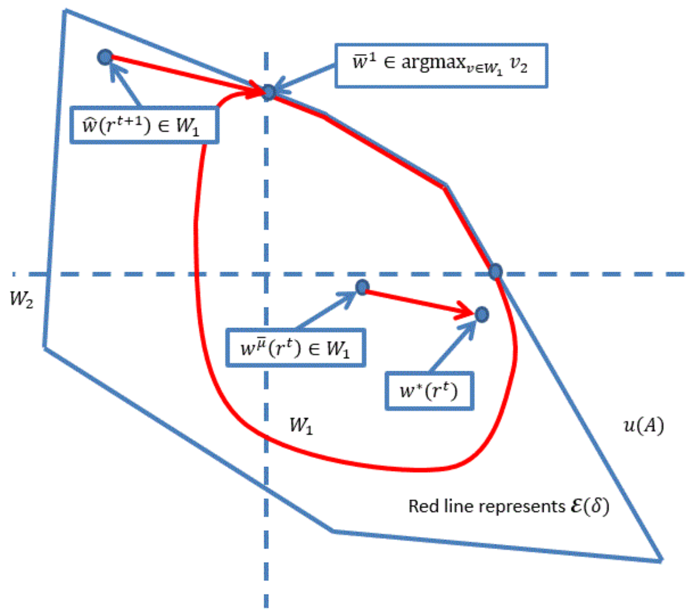

Figure 3: The vector from $w^{\bar{\mu}}(r^t)$ to $w^*(r^t)$ is parallel to the one from $\hat{w}(r^{t+1})$ to $\bar{w}^1$.

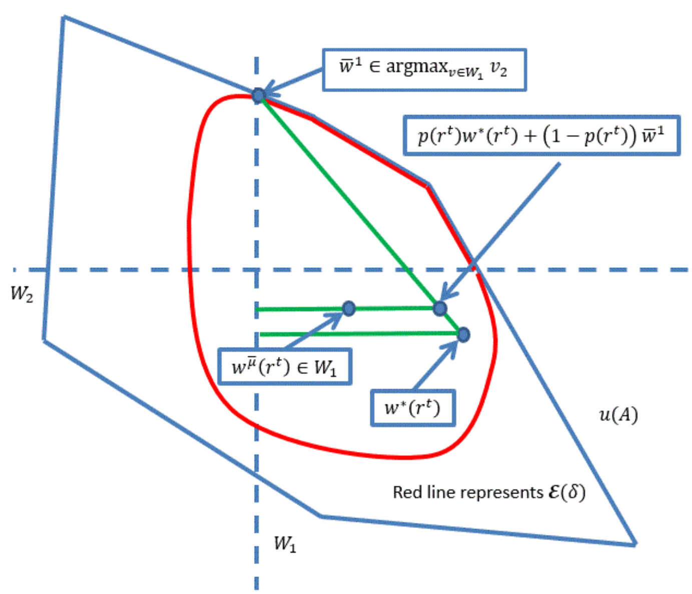

Figure 4: $p(r^t)w^*(r^t) + (1 - p(r^t))\,\bar{w}^1$ and $w^{\bar{\mu}}(r^t)$ have the same value for player 2.

1. Incentive compatibility for player 1: It suffices to show that, conditional on any on-path history r^t and period t recommendation $r_{1,t}$, the expected continuation payoff from period $t+1$ onward lies in W_1. If $\omega_t = S_2$, then this continuation payoff is $\bar{w}^1 \in W_1$. If $\omega_t = S_1$, then it suffices to show that $w\left(r^{t+1}\right) \in W_1$ for all r^{t+1}. If $w^{\bar{\mu}}(r^{t+1}) \in W_1$, then, by Lemma 6, $w(r^{t+1}) = p(r^{t+1})w^*(r^{t+1}) + (1 - p(r^{t+1}))\,\bar{w}^1$ Pareto dominates $w^{\bar{\mu}}(r^{t+1}) \in W_1$, so $w(r^{t+1}) \in W_1$. If $w^{\bar{\mu}}(r^{t+1}) \notin W_1$, then $w(r^{t+1}) = \bar{w}^1 \in W_1$. Hence, $w\left(r^{t+1}\right) \in W_1$ for all r^{t+1}.

2. Incentive compatibility for player 2: Fix an on-path history r^t and a period t recommendation $r_{2,t}$. If $\omega_t = S_2$, or if both $\omega_t = S_1$ and $w^{\bar{\mu}}(r^{t+1}) \notin W_1$, then the expected continuation payoff from period $t+1$ onward conditional on $(r^t, r_{2,t})$ is $\bar{w}^1 \in W_1$, so (2) holds. If instead $\omega_t = S_1$ and $w^{\bar{\mu}}(r^{t+1}) \in W_1$, then $w(r^{t+1}) = p(r^{t+1})w^*(r^{t+1}) + (1 - p(r^{t+1}))\,\bar{w}^1 = w^{\bar{\mu}}\left(r^{t+1}\right)$ by (8). As μ is an equilibrium with the universal information structure and $\Pr^{\mu^*}\left(r_t | r_2^{t+1}\right) = \Pr^{\bar{\mu}}\left(r_t | r^t, r_{2,t}\right)$, this implies that (2) holds for player 2, by the same argument as in Section 8.2.

3. $\lambda \cdot v \leq \lambda \cdot v^*$: Immediate from Lemma 6 with $t = 1$.

9 Extensions

This section discusses what happens when the conditions for Theorem 1 are violated, as well as the extent to which the payoff bound is tight.

9.1 What if $\mathring{W}^* = \emptyset$?

The assumption that $\mathring{W}^* \neq \emptyset$ guarantees that all action profiles are supportable in equilibrium, which in turn implies that deviators can be held to their correlated minmax payoffs. This fact plays a key role both in showing that $\overline{E_{\text{med}}(\delta)}$ is an upper bound on $E\left(\delta, p\right)$ for any private monitoring structure p and in recursively characterizing $\overline{E_{\text{med}}(\delta)}$. However, the assumption that $\mathring{W}^* \neq \emptyset$ is restrictive, in that it is violated when players are too impatient.

Furthermore, it implies that the Pareto frontier of $\overline{E_{\text{med}}(\delta)}$ coincides with the Pareto frontier of the feasible payoff set for some Pareto weights λ (but of course not for others), so this assumption must also be relaxed for our approach to be able to give non-trivial payoff bounds for all Pareto weights.

To address these concerns, this subsection shows that even if $\mathring{W}^* = \emptyset$, $\overline{E_{\text{med}}(\delta)}$ may still be an upper bound on $E(\delta, p)$ for any private monitoring structure p, and $\overline{E_{\text{med}}(\delta)}$ can still be characterized recursively. The idea is that, even if not all action profiles are supportable, our approach still applies if a condition analogous to $\mathring{W}^* \neq \emptyset$ holds with respect to the subset of action profiles that are supportable.

Recall that $\mathcal{E}(\delta)$ is the equilibrium payoff set with the universal monitoring structure where a mediator observes the recommendation profile r_t and action profile a_t in each period t, while each player i only observes her own recommendation $r_{i,t}$ and her own action $a_{i,t}$. Denote this monitoring structure by p^*. We provide a sufficient condition—which is more permissive than $\mathring{W}^* \neq \emptyset$—under which we show that the Pareto frontier of $\overline{E_{\text{med}}(\delta)}$ dominates the Pareto frontier of $\mathcal{E}(\delta)$ and characterize $\overline{E_{\text{med}}(\delta)}$.

Let $\text{supp}(\delta)$ be the set of supportable actions with monitoring structure p^*:

$$\text{supp}(\delta) = \left\{ a \in A : \begin{array}{c} \text{with monitoring structure } p^*, \\ \text{there exist an equilibrium strategy } \mu \\ \text{and history } h_m^t \text{ with } a \in \text{supp}(\mu(h_m^t)) \end{array} \right\}.$$

Note that in this definition h_m^t can be an off-path history.

On the other hand, given a product set of action profiles $\bar{A} = \prod_{i \in I} \bar{A}_i \subseteq A$, let $S_i(\bar{A})$ be the set of actions $a_i \in \bar{A}_i$ such that there exists a correlated action $\alpha_{-i} \in \Delta(\bar{A}_{-i})$ such that

$$(1-\delta)\, u_i(a_i, \alpha_{-i}) + \delta \max_{\bar{a} \in \bar{A}} u_i(\bar{a}) \geq (1-\delta) \max_{\hat{a}_i \in A_i} u_i(\hat{a}_i, \alpha_{-i}) + \delta \min_{\hat{\alpha}_{-i} \in \Delta(\bar{A}_{-i})} \max_{a_i \in A_i} u_i(a_i, \hat{\alpha}_{-i}). \tag{11}$$

(That is, $a_i \in S_i(\bar{A})$ if there exists $\alpha_{-i} \in \Delta(\bar{A}_{-i})$ such that, if her opponents play α_{-i}, player i's reward for playing a_i is the best payoff possible among those with support in $\bar{A}$, and player

i's punishment for deviating from a_i is the worst possible among those with support in $\bar{A}_{-i}$, then player i plays a_i.) Let $S\left(\bar{A}\right) = \prod_{i \in I} S_i\left(A_i\right) \subseteq \bar{A}$. Let $\mathcal{A}^1 = A$, let $\mathcal{A}^n = S\left(\mathcal{A}^{n-1}\right)$ for $n > 1$, and let $\mathcal{A}^\infty = \lim_{n \to \infty} \mathcal{A}^n$. Note that the problem of computing $\mathcal{A}^\infty$ is tractable, as the set $S\left(\bar{A}\right)$ is defined by a finite number of linear inequalities.

Finally, in analogy with the definition of w_i from Section 5, let

$$\min_{\alpha_{-i} \in \Delta\left(\mathcal{A}^\infty_{-i}\right)} \max_{a_i \in A_i} u_i(a_i, \alpha_{-i}) + \frac{1-\delta}{\delta} \max_{r \in \mathcal{A}^\infty, a_i \in A_i} \left\{u_i(a_i, r_{-i}) - u_i(r)\right\}.$$

be the lowest continuation payoff such that player i does not want to deviate to any $a_i \in A_i$ at any recommendation profile $r \in \mathcal{A}^\infty$, when she is minmaxed forever if she deviates, *subject to the constraint that punishments are drawn from* $\mathcal{A}^\infty$. In analogy with the definition of W_i from Section 5, let

$$\bar{W}_i = \left\{ w \in \mathbb{R}^{|I|} : w_i \geq \min_{\alpha_{-i} \in \Delta\left(\mathcal{A}^\infty_{-i}\right)} \max_{a_i \in A_i} u_i(a_i, \alpha_{-i}) + \frac{1-\delta}{\delta} \max_{r \in \mathcal{A}^\infty, a_i \in A_i} \left\{u_i(a_i, r_{-i}) - u_i(r)\right\} \right\}.$$

We show the following.

Proposition 2 *Assume that* $|I| = 2$. *If*

$$\operatorname{int}\left(\bigcap_{i \in I} \bar{W}_i \cap u(\mathcal{A}^\infty)\right) \neq \emptyset \tag{12}$$

in the topology induced from $u(\mathcal{A}^\infty)$*, then for every private monitoring structure* p *and every non-negative Pareto weight* $\lambda \in \Lambda_+$*, we have*

$$\max_{v \in E(\delta, p)} \lambda \cdot v \leq \max_{v \in \overline{E_{\text{med}}(\delta)}} \lambda \cdot v.$$

In addition, $\operatorname{supp}(\delta) = \mathcal{A}^\infty$.

Proof. Given that $\operatorname{supp}(\delta) = \mathcal{A}^\infty$, the proof is analogous to the proof of Theorem 1, everywhere replacing $u(A)$ with $u(\mathcal{A}^\infty)$ and replacing "full support" with "full support within

$\mathcal{A}^\infty$." The proof that $\text{supp}(\delta) = \mathcal{A}^\infty$ whenever $\text{int}\left(\bigcap_{i\in I} \bar{W}_i \cap u(\mathcal{A}^\infty)\right) \neq \emptyset$ is deferred to the appendix. ■

Note that Proposition 2 only improves on Theorem 1 at low discount factors: for a high enough discount factor, $A = S(A) = \mathcal{A}^\infty$, so (12) reduces to $\mathring{W}^* \neq \emptyset$. However, as $\mathring{W}^*$ can be empty only for low discount factors, this is precisely the case where an improvement is needed.[16]

In order to be able to use Proposition 2 to give a recursive upper bound on $E(\delta, p)$ when $\mathring{W}^* \neq \emptyset$, we must characterize $\overline{E_{\text{med}}(\delta)}$ under (12). Our earlier characterization generalizes easily. In particular, the following definitions are analogous to Definitions 1 and 2.

Definition 3 *For any set $V \subseteq \mathbb{R}^{|I|}$, a correlated action profile $\alpha \in \Delta(\text{supp}(\delta))$ is* supp(δ) enforceable *on V by a mapping $\gamma : \text{supp}(\delta) \to V$ such that, for each player i, and action $a_i \in \text{supp}\,\alpha_i$,*

$$\mathbb{E}^\alpha\left[(1-\delta)\, u_i(a_i, a_{-i}) + \delta\gamma(a_i, a_{-i})\right] \geq \max_{a_i' \in A_i} \mathbb{E}^\alpha\left[(1-\delta)\, u_i(a_i', a_{-i})\right] + \delta \min_{\alpha_{-i}\in\Delta(\text{supp}(\delta))} \max_{\hat{a}_i \in A_i} u_i(\hat{a}_i, \alpha_{-i}).$$

Definition 4 *A payoff vector $v \in \mathbb{R}^{|I|}$ is* supp(δ) decomposable *on V if there exists a correlated action profile $\alpha \in \Delta(\text{supp}(\delta))$ which is* supp(δ) *enforced on V by some mapping γ such that*

$$v = \mathbb{E}^\alpha\left[(1-\delta)\, u(a) + \delta\gamma(a)\right].$$

Let $\tilde{B}^{\text{supp}(\delta)}(V) = \left\{v \in \mathbb{R}^{|I|} : v \text{ is supp}(\delta) \text{ decomposable on } V\right\}$.

Let $W^{\text{supp}(\delta),1} = u(\text{supp}(\delta))$, let $W^{\text{supp}(\delta),n} = \tilde{B}^{\text{supp}(\delta)}\left(W^{\text{supp}(\delta),n-1}\right)$ for $n > 1$, and let $W^{\text{supp}(\delta),\infty} = \lim_{n\to\infty} W^{\text{supp}(\delta),n}$. We have the following.

Proposition 3 *If* $\text{int}\left(\bigcap_{i\in I} \bar{W}_i \cap u(\mathcal{A}^\infty)\right) \neq \emptyset$, *then* $\overline{E_{\text{med}}(\delta)} = W^{\mathcal{A}^\infty,\infty}$.

Proof. Given that $\text{supp}(\delta) = \mathcal{A}^\infty$ by Proposition 2, the proof is analogous to the proof of Theorem 2. ■

[16] To be clear, it is possible for W^* to be empty while $\mathcal{A}^\infty = A$. Theorem 1 and Proposition 2 only give sufficient conditions: we are not claiming that they cover every possible case.

As an example of how Propositions 2 and 3 can be applied, one can check that applying the operator S in the Bertrand example in Section 7 for any $\delta \in \left(\frac{1}{4}, \frac{5}{18}\right)$ yields $\mathcal{A}^\infty = \{W, L, M\} \times \{W, L, M\}$—ruling out the efficient action profile (H, H)—and $\operatorname{int}\left(\bigcap_{i \in I} \bar{W}_i \cap u(\mathcal{A}^\infty)\right) \neq \emptyset$. We can then compute $\overline{E_{\text{med}}(\delta)}$ by applying the operator $\tilde{B}^{\mathcal{A}^\infty}$.

Finally, we mention that $\overline{E_{\text{med}}(\delta)}$ can be characterized recursively even if $\operatorname{int}\left(\bigcap_{i \in I} \bar{W}_i \cap u(\mathcal{A}^\infty)\right) = \emptyset$. This fact is not directly relevant for the current paper, so we omit the proof. It is available from the authors upon request.

9.2 Tightness of the Bound

There are at least two senses in which $\overline{E_{\text{med}}(\delta)}$ is a *tight* bound on the equilibrium payoff set with private monitoring.

First, thus far our model of repeated games with private monitoring has maintained the standard assumption that the distribution of period t signals depends only on period t actions: that is, that this distribution can be written as $p(\cdot|a_t)$. In many settings, it would be desirable to relax this assumption and let the distribution of period t signals depend on the entire history of actions and signals up to period t, leading to a conditional distribution of the form $p_t(\cdot|a^t, z^t)$, as well as letting players receive signals before the first round of play. (Recall that $a^t = (a_\tau)_{\tau=1}^{t-1}$ and $z^t = (z_\tau)_{\tau=1}^{t-1}$.) For example, colluding firms do not only observe their sales in every period, but also occasionally get more information about their competitors' past behavior from trade associations, auditors, tax data, and the like.[17] In the space of such *nonstationary* private monitoring structures, the bound $\overline{E_{\text{med}}(\delta)}$ is clearly tight: $\overline{E_{\text{med}}(\delta)}$ is an upper bound on $E(\delta, p)$ for any non-stationary private monitoring structure p, because the equilibrium payoff set with the universal monitoring structure p^*, $\mathcal{E}(\delta)$, remains an upper bound on $E(\delta, p)$; and the bound is tight because mediated perfect monitoring is itself a particular nonstationary private monitoring structure.

Second, maintaining the assumption that the monitoring structure is stationary, the

[17] Rahman (2014, p. 1) quotes from the European Commission decision on the amino acid cartel: a typical cartel member "reported its citric acid sales every month to a trade association, and every year, Swiss accountants audited those figures."

bound $\overline{E_{\text{med}}(\delta)}$ is tight if the players can communicate through cheap talk, as they can then "replicate" the mediator among themselves. For this result, we also need to slightly generalize our definition of a private monitoring structure by letting the players receive signals before the first round of play. This seems innocuous, especially if we take the perspective of an an outside observer who does not know the game's start date. Equivalently, the players have access to a mediator at the beginning of the game only. The monitoring structure is required to be stationary thereafter. We call such a monitoring structure a *private monitoring structure with ex ante correlation.*

Proposition 4 *Let $E_{\text{talk}}(\delta, p)$ be the sequential equilibrium payoff set in the repeated game with private monitoring structure with ex ante correlation p and with finitely many rounds of public cheap talk before each round of play. If $|I| = 2$ and $\delta > \delta^*$, then there exists a private monitoring structure with ex ante correlation p such that $\overline{E_{\text{talk}}(\delta, p)} = \overline{E_{\text{med}}(\delta)}$.*

The proof is long and is deferred to the online appendix. The main idea is as in the literature on implementing correlated equilibria without a mediator (see Forges (2009) for a survey). More specifically, Proposition 4 is similar to Theorem 9 of Heller, Solan, and Tomala (2012), which shows that communication equilibria in repeated games with perfect monitoring can always be implemented by ex ante correlation and cheap talk. As the private monitoring structure used in the proof of Proposition 4 is in fact perfect monitoring, the main difference between the results is that theirs is for Nash rather than sequential equilibrium, so they are concerned only with detecting deviations rather than providing incentives to punish deviations once detected. In our model, when $\delta > \delta^*$, incentives to minmax the deviator can be provided (as in Lemma 1) *if* her opponent does not realize that the punishment phase has begun. The additional challenge in the proof of Proposition 4 is thus that we sometimes need a player to switch to the punishment phase for her opponent without realizing that this switch has occurred.

If one insists on stationary monitoring and does not allow communication, then we believe that there are some games in which our bound is not tight, in that there are points in $\overline{E_{\text{med}}(\delta)}$

which are not attainable in equilibrium for any stationary private monitoring structure. We leave this as a conjecture.[18]

9.3 What if There are More Than Two Players?

The condition that $\mathring{W}^* \neq \emptyset$ no longer guarantees that mediated perfect monitoring outperforms private monitoring when there are more than two players. We record this as a proposition.

Proposition 5 *There are games with $|I| > 2$ where $\mathring{W}^* \neq \emptyset$ but $\max_{v \in E(\delta,p)} \lambda \cdot v > \max_{v \in \overline{E_{\text{med}}(\delta)}} \lambda \cdot v$ for some private monitoring structure p and some non-negative Pareto weight $\lambda \in \Lambda_+$.*

Proof. We give an example in the appendix. ■

To see where the proof of Theorem 1 breaks down when $|I| > 2$, recall that the proof is based on fact that, for any Pareto-efficient payoff v, if $v \notin W_i$ for one player i, then it must be the case that $v \in W_j$ for the other player j. This implies that incentive compatibility is a problem only for one player at a time, which lets us construct an equilibrium with perfect monitoring by basing continuation play only on that player's past recommendations (which she necessarily knows in any private monitoring structure). On the other hand, if there are more than two players, several players' incentive compatibility constraints might bind at once when we publicize past recommendations. The proof of Theorem 1 then cannot get off the ground.

We can however say some things about what happens with more than two players.

First, the argument in the proof of Proposition 2 that $\text{supp}(\delta) = \mathcal{A}^\infty$ whenever $\text{int}\left(\bigcap_{i \in I} \bar{W}_i \cap u(\mathcal{A}^\infty)\right) \neq \emptyset$ does not rely on $|I| = 2$. Thus, when $\text{int}\left(\bigcap_{i \in I} \bar{W}_i \cap u(\mathcal{A}^\infty)\right) \neq \emptyset$, we can characterize the

[18] Strictly speaking, since our maintained definition of a private monitoring structure does not allow ex ante correlation, if $\delta = 0$ then there are points in $\overline{E_{\text{med}}(\delta)}$ which are not attainable with any private monitoring structure whenever the stage game's correlated equilibrium payoff set strictly contains its Nash equilibrium payoff set. The non-trivial conjecture is that the bound is still not tight when ex ante correlation is allowed, but communication is not, or when discount factors are moderately high.

set of supportable actions for any number of players. This is sometimes already enough to imply a non-trivial upper bound on payoffs.

Second, Proposition 6 below shows that if a payoff vector $v \in \text{int}\,(u\,(A))$ satisfies $v_i > \underline{u}_i + \frac{1-\delta}{\delta} d_i$ for all $i \in I$, then $v \in \overline{E_{\text{med}}(\delta)}$. This shows that private monitoring cannot do "much" better than mediated perfect monitoring when the players are at least moderately patient (e.g., it cannot do more than order $1-\delta$ better).

Finally, suppose there is a player i whose opponents $-i$ all have identical payoff functions: $\exists i : \forall j, j' \in I \setminus \{i\}$, $u_j(a) = u_{j'}(a)$ for all $a \in A$. Then the proof of Theorem 1 can be adapted to show that private monitoring cannot outperform mediated perfect monitoring in a direction where the extremal payoff vector v lies in $\bigcap_{j \in -i} W_j$ (but not necessarily in W_i). For example, if the game involves one firm and many identical consumers, then the consumers' best equilibrium payoff under mediated perfect monitoring is at least as good as under private monitoring. We can also show that the same result holds if the preferences of players $-i$ are sufficiently close to each other.

9.4 The Folk Theorem with Mediated Perfect Monitoring

The point of this paper is bounding the equilibrium payoff set with private monitoring at a fixed discount factor. Nonetheless, it is worth pointing out that—as a corollary of our results—a strong version of the folk theorem holds with mediated perfect monitoring, without any conditions on the feasible payoff set (e.g., full-dimensionality, non-equivalent utilities), and with a rate of convergence of $1-\delta$. One reason why this result may be of interest is that it shows that, if players are fairly patient, they cannot do "much" better with private monitoring than with mediated perfect monitoring, even if the sufficient conditions for Theorem 1 fail. Another reason is that the mediator can usually be replaced by unmediated communication among the players (we spell out exactly when below), and it seems worth knowing that the folk theorem holds with unmediated communication without conditions on the feasible payoff set.

Recall that $\underline{u}_i$ is player i's correlated minmax payoff and d_i is player i's greatest possible deviation gain. Let $\underline{u} = (u_i)_{i \in I}$ and $d = (d_i)_{i \in I}$.

Proposition 6 *If a payoff vector $v \in \text{int}\,(u\,(A))$ satisfies*

$$v > \underline{u} + \frac{1-\delta}{\delta} d, \tag{13}$$

then $v \in \overline{E_{\text{med}}(\delta)}$.

Proof. Fix $\alpha \in \Delta\,(A)$ such that $v = u\,(\alpha)$. If v satisfies (13), then $v \in \mathring{W}^*$ (so in particular $\mathring{W}^* \neq \emptyset$), and the infinite repetition of α satisfies (2). Lemma 1 then gives $v \in \overline{E_{\text{med}}(\delta)}$. ■

The folk theorem is a corollary of Proposition 6.[19]

Corollary 1 (Folk Theorem) *For every strictly individually rational payoff vector $v \in u\,(A)$, there exists $\bar{\delta} < 1$ such that if $\delta > \bar{\delta}$ then $v \in E_{\text{med}}(\delta)$.*

Proof. It is immediate that $v \in \overline{E_{\text{med}}(\delta)}$ for high enough δ: let $\bar{\delta} = \max_{i \in I} \frac{d_i}{d_i + v_i - \underline{u}_i}$ and apply Proposition 6. The proof that taking the closure is unnecessary is in the appendix. ■

In earlier work (Sugaya and Wolitzky, 2014), we have shown that the mediator can usually be dispensed with in Proposition 6 and Corollary 1 if players can communicate through cheap talk. In particular, this is possible unless there are exactly three players, of whom exactly two have equivalent utilities in the sense of Abreu, Dutta, and Smith (1994).

10 Conclusion

This paper gives a simple sufficient condition ($\delta > \delta^*$) under which the equilibrium payoff set in a two-player repeated game with perfect monitoring and a mediator is a tight, recursive upper bound on the equilibrium payoff set in the same game with any imperfect private monitoring structure. There are at least three perspectives from which this result may be

[19] The restriction in Corollary 1 to *strictly* individually rational payoff vectors cannot be dropped. In particular, the counterexample to the folk theorem due to Forges, Mertens, and Neyman (1986)—in which no payoff vector is strictly individually rational—remains valid in the presence of a mediator.

of interest. First, it shows that simple, recursive methods can be used to upper-bound the equilibrium payoff set in a repeated game with imperfect private monitoring at a fixed discount factor, even though the problem of recursively characterizing this set seems intractable. Second, it characterizes the set of payoffs that can arise in a repeated game for some monitoring structure. Third, it shows that information is valuable in mediated repeated games, in that players cannot benefit from imperfections in the monitoring technology when $\delta > \delta^*$.

These different perspectives on our results suggest different questions for future research. Do moderately patient players always benefit from *any* improvement in the monitoring technology, or only from going all the way to perfect monitoring? Is it possible to characterize the set of payoffs that can arise for some monitoring structure even if $\delta < \delta^*$? If we do know the monitoring structure under which the game is being played, is there a general way of using this information to tighten our upper bound? Answering these questions may improve our understanding of repeated games with private monitoring at fixed discount factors, even if a full characterization of the equilibrium payoff set in such games remains out of reach.

11 Appendix

11.1 Proof of Proposition 1: Mediated Perfect Monitoring

As the players' stage game payoffs from any profile other than (U, L) sum to at most 3, it follows that the players' per-period payoffs may sum to more than 3 only if (U, L) is played in some period t with positive probability. For this to occur in equilibrium, player 1's expected continuation payoff from playing U must exceed her expected continuation payoff from playing D by more than $(1 - \delta) 1 = \frac{5}{6}$, her instantaneous gain from playing D rather than U. In addition, player 1 can guarantee herself a continuation payoff of 0 by always playing D, so her expected continuation payoff from playing U must exceed $\frac{1}{\delta}\left(\frac{5}{6}\right) = 5$. This is possible only if the probability that (T, M) is played in period $t+1$ when U is played in period t exceeds the number p such that

$$\left(1-\frac{1}{6}\right)[p(6)+(1-p)(3)]+\frac{1}{6}(6)=5,$$

or $p=\frac{3}{5}$. In particular, there must exist a period $t+1$ history h_2^{t+1} of player 2's such that (T, M) is played with probability at least $\frac{3}{5}$ in period $t+1$ conditional on reaching h_2^{t+1}. At such a history, player 2's payoff from playing M is at most

$$\left(1-\frac{1}{6}\right)\left[\frac{3}{5}(-3)+\frac{2}{5}(3)\right]+\frac{1}{6}(3)=0.$$

On the other hand, noting that player 2 can guarantee himself a continuation payoff of 0 by playing $\frac{1}{2}L+\frac{1}{2}M$, player 2's payoff from playing L at this history is at least

$$\left(1-\frac{1}{6}\right)\left[\frac{3}{5}(3)+\frac{2}{5}(-3)\right]+\frac{1}{6}(0)=\frac{1}{2}.$$

Therefore, player 2 has a profitable deviation, so no such equilibrium can exist.

11.2 Proof of Proposition 1: Private Monitoring

Consider the following imperfect private monitoring structure. Player 2's action is perfectly observed. Player 1's action is perfectly observed when it equals T or B. When player 1 plays U or D, player 2 observes one of two possible private signals, m and r. Whenever player 1 plays U, player 2 observes signal m with probability 1; whenever player 1 plays D, player 2 observes signals m and r with probability $\frac{1}{2}$ each.

We now describe a strategy profile under which the players' payoffs sum to $\frac{23}{7}\approx 3.29$.

Player 1's strategy: In each odd period $t=2n+1$ with $n=0,1,...$, player 1 plays $\frac{1}{3}U+\frac{2}{3}D$. Let $a_1(n)$ denote the realization of this mixture. In the even period $t=2n+2$, if the previous action $a_1(n)$ equals U, then player 1 plays T; if the previous action $a_1(n)$ equals D, then player 1 plays B.

Player 2's strategy: In each odd period $t=2n+1$ with $n=0,1,...$, player 2 plays L.

Let $y_2(n)$ denote the realization of player 2's private signal. In the even period $t = 2n + 2$, if the previous private signal $y_2(n)$ equals m, then player 2 plays M; if the previous signal $y_2(n)$ equals r, then player 2 plays R.

We check that this strategy profile, together with any consistent belief system, is a sequential equilibrium.

In an odd period, player 1's payoff from U is the solution to $v = \frac{5}{6}(2) + \frac{1}{6}\frac{5}{6}(6) + \frac{1}{6^2}v$. On the other hand, her payoff from D is $\frac{5}{6}(3) + \frac{1}{6}\frac{5}{6}(0) + \frac{1}{6^2}v$. Hence, player 1 is indifferent between U and D (and clearly prefers either of these to T or B).

In addition, playing L is a myopic best response for player 2, player 1's continuation play is independent of player 2's action, and the distribution of player 2's signal is also independent of player 2's action. Hence, playing L is optimal for player 2.

Next, in an even period, it suffices to check that both players always play myopic best responses, as in even periods continuation play is independent of realized actions and signals. If player 1's last action was $a_1(n) = U$, then she believes that player 2's signal is $y_2(n) = m$ with probability 1 and thus that he will play M. Hence, playing T is optimal. If instead player 1's last action was $a_1(n) = D$, then she believes that player 2's signal is equal to m and r with probability $\frac{1}{2}$ each, and thus that he will play $\frac{1}{2}M + \frac{1}{2}R$. Hence, both T and B are optimal.

On the other hand, if player 2 observes signal $y_2(n) = m$, then his posterior belief that player 1's last action was $a_1(n) = U$ is

$$\frac{\frac{1}{3}(1)}{\frac{1}{3}(1) + \frac{2}{3}\left(\frac{1}{2}\right)} = \frac{1}{2}.$$

Hence, player 2 is indifferent among all of his actions. If player 2 observes $y_2(n) = r$, then his posterior is that $a_1(n) = D$ with probability 1, so that M and R are optimal.

Finally, expected payoffs under this strategy profile in odd periods sum to $\frac{1}{3}(4) + \frac{2}{3}(3) =$

$\frac{10}{3}$, and in even periods sum to 3. Therefore, per-period expected payoffs sum to

$$\left(1-\frac{1}{6}\right)\left(\frac{10}{3}+\frac{1}{6}(3)\right)\left(1+\frac{1}{6^2}+\frac{1}{6^4}+\ldots\right)=\frac{23}{7}.$$

Three remarks on the proof: First, the various indifferences in the above argument result only because we have chosen payoffs to make the example as simple as possible. One can modify the example to make all incentives strict.[20] Second, players' payoffs are measurable with respect to their own actions and signals. In particular, the required realized payoffs for player 2 are as follows:

(Action,Signal) Pair:	(L,m)	(L,r)	(M,m)	(M,r)	(R,m)	(R,r)
Realized Payoff:	2	-2	0	0	0	0

Third, a similar argument shows that imperfect public monitoring with private strategies can also outperform mediated perfect monitoring.[21]

11.3 Proof of Lemma 2

Fix such a strategy μ and any strict full-support equilibrium μ^{strict}. We construct a strict equilibrium that attains a payoff close to v.

In period 1, the mediator draws one of two states, R_v and R_{perturb}, with probabilities $1-\varepsilon$ and ε, respectively. In state R_v, the mediator's recommendation is determined as follows: If no player has deviated up to period t, the mediator recommends r_t according to $\mu(h_m^t)$. If only player i has deviated, the mediator recommends $r_{-i,t}$ to players $-i$ according to α_{-i}^*, and recommends some best response to α_{-i}^* to player i. Multiple deviations are treated as in the proof of Lemma 1. On the other hand, in state R_{perturb}, the mediator follows

[20] The only non-trivial step in doing so is giving player 1 a strict incentive to mix in odd periods. This can be achieved by introducing correlation between the players' actions in odd periods.

[21] Here is a sketch: Modify the current example by adding a strategy L' for player 2, which is an exact duplicate of L as far as payoffs are concerned, but which switches the interpretation of signals m and r. Assume that player 1 cannot distinguish between L and L', and modify the equilibrium by having player 2 play $\frac{1}{2}L+\frac{1}{2}L'$ in odd periods. Then, even if the signals m and r are publicly observed, their interpretations will be private to player 2, and essentially the same argument as with private monitoring applies.

the equilibrium μ^{strict}. Player i follows the recommendation $r_{i,t}$ in period t. Since the constructed recommendation schedule has full support, player i never believes that another player has deviated. Moreover, since μ^{strict} has full support, player i believes that the mediator's state is R_{perturb} with positive probability after any history. Therefore, by (2) and the fact that μ^{strict} is a strict equilibrium, it is always strictly optimal for each player i to follow her recommendation on path. Taking $\varepsilon \to 0$ yields a sequence of strict equilibria with payoffs converging to v.

11.4 Proof of Lemma 3

Standard arguments imply that, for every $v \in \tilde{B}(V)$, there exists an on-path recommendation strategy μ that yields payoff v and satisfies (2). Under the assumption that $\delta > \delta^*$, the conclusion of the lemma follows from Lemma 2.

11.5 Proof of Lemma 4

By Lemma 3 and boundedness, we need only show that $\overline{E_{\text{med}}(\delta)} \subseteq \tilde{B}\left(\overline{E_{\text{med}}(\delta)}\right)$.

Let $E_{\text{med}}^{\text{weak}}(\delta)$ be the set of (possibly weak) sequential equilibrium payoffs with mediated perfect monitoring. Note that, in any sequential equilibrium, player i's continuation payoff at any history h_i^t must be at least $\underline{u}_i$. Therefore, if μ is an on-path recommendation strategy in a (possibly weak) sequential equilibrium, then it must satisfy (2). Hence, under the assumption that $\mathring{W}^* \neq \emptyset$, we have $E_{\text{med}}^{\text{weak}}(\delta) \subseteq \overline{E_{\text{med}}(\delta)}$.

Now, for any $v \in E_{\text{med}}(\delta)$, let μ be a corresponding equilibrium mediator's strategy. In the corresponding equilibrium, if some player i deviates in period 1 while her opponents are obedient, player i's continuation payoff must be at least $\underline{u}_i$. Hence, we have

$$\mathbb{E}^{\mu}\left[(1-\delta)\, u_i(a_i, a_{-i}) + \delta w_i(a_i, a_{-i}) \mid a_i\right] \geq \max_{a_i' \in A_i} E^{\mu}\left[(1-\delta)\, u_i(a_i', a_{-i}) \mid a_i\right] + \delta \underline{u}_i,$$

where $w_i(a_i, a_{-i})$ is player i's equilibrium continuation payoff when action profile (a_i, a_{-i}) is recommended and obeyed in period 1. Finally, since action profile (a_i, a_{-i}) is in the support of the mediator's recommendation in period 1, each player assigns probability 1 to the true mediator's history when (a_i, a_{-i}) is recommended and played in period 1. Therefore, continuation play from this history is itself at least a weak sequential equilibrium. In particular, we have $w_i(a_i, a_{-i}) \in E^{\text{weak}}_{\text{med}}(\delta) \subseteq \overline{E_{\text{med}}(\delta)}$ for all $(a_i, a_{-i}) \in \operatorname{supp} \mu(h^t)$. Hence, v is minmax-threat decomposable on $\overline{E_{\text{med}}(\delta)}$ by action profile $\mu(\emptyset)$ and continuation payoff function w, so in particular $v \in \tilde{B}\left(\overline{E_{\text{med}}(\delta)}\right)$.

We have shown that $E_{\text{med}}(\delta) \subseteq \tilde{B}\left(\overline{E_{\text{med}}(\delta)}\right)$. As $\overline{E_{\text{med}}(\delta)}$ is compact and $\tilde{B}$ preserves compactness, taking closures yields $\overline{E_{\text{med}}(\delta)} \subseteq \overline{\tilde{B}\left(\overline{E_{\text{med}}(\delta)}\right)} = \tilde{B}\left(\overline{E_{\text{med}}(\delta)}\right)$.

11.6 Proof of Lemma 5

By (6), $\|w^*(w) - w^*(\tilde{w})\| \leq \delta \|w - \tilde{w}\|$. By (7),

$$\begin{aligned} \left|F(w)\left(r^t\right) - F(\tilde{w})\left(r^t\right)\right| &= 1_{\{w^{\bar{\mu}}(r^t) \in W_1\}} \left| \begin{array}{c} \{p(w)(r^t)w^*(w)(r^t) + (1-p(w)(r^t))\,\bar{w}^1\} \\ -\{p(\tilde{w})(r^t)w^*(\tilde{w})(r^t) + (1-p(\tilde{w})(r^t))\,\bar{w}^1\} \end{array} \right| \\ &\leq \|w^*(w) - w^*(\tilde{w})\|. \end{aligned}$$

Combining these inequalities yields $\|F(w) - F(\tilde{w})\| \leq \delta \|w - \tilde{w}\|$.

11.7 Proof of Lemma 6

It is useful to introduce a family of auxiliary value functions $\left(w^T\right)_{T=1}^{\infty}$ and $\left(w^{*,T}\right)_{T=1}^{\infty}$, which will converge to w and w^* pointwise in r^t as $T \to \infty$. For periods $t \geq T$, define

$$w^T(r^t) = w^{\bar{\mu}}(r^t) \text{ and } w^{*,T}(r^{t-1}) = w^{\bar{\mu}}(r^{t-1}). \tag{14}$$

On the other hand, for periods $t \leq T-1$, define $w^{*,T}(r^t)$, $p^T(r^t)$, and $w^T(r^t)$ given $w^T(r^{t+1})$ recursively, as follows. First, define

$$w^{*,T}(r^t) = (1-\delta)\, u\left(\bar{\mu}(r^t)\right) + \delta \mathbb{E}\left[w^T(r^{t+1})|r^t\right]. \tag{15}$$

Note that, for $t = T - 1$, this definition is compatible with (14). Second, given $w^{*,T}(r^t)$, define

$$w^T(r^t) = 1_{\{w^{\bar{\mu}}(r^t) \in W_1\}} \left\{ p^T(r^t) w^{*,T}(r^t) + \left(1 - p^T(r^t)\right) \bar{w}^1 \right\} + 1_{\{w^{\bar{\mu}}(r^t) \notin W_1\}} \bar{w}^1, \tag{16}$$

where, when $w^{\bar{\mu}}(r^t) \in W_1$, $p^T(r^t)$ is the largest number in $[0,1]$ such that

$$p^T(r^t) w_2^{*,T}(r^t) + \left(1 - p^T(r^t)\right) \bar{w}_2^1 \geq w_2^{\bar{\mu}}(r^t). \tag{17}$$

We show that $w^{*,T}$ converges to w^*.

Lemma 7 $\lim_{T \to \infty} w^{*,T}(r^t) = w^*(r^t)$ *for all* $r^t \in A^t$.

Proof. By Lemma 5, it suffices to show that $F(w^T) = w^{T+1}$. For $t \geq T+1$, (14) implies that $w^{*,T+1}(r^{t-1}) = w^{\bar{\mu}}(r^{t-1})$. On the other hand, given w^T, $w^*(w_2^T)$ is the value calculated according to (6). Since $w^T(r^t) = w^{\bar{\mu}}(r^t)$ by (14), we have $w^*(w_2^T)(r^{t-1}) = w^{\bar{\mu}}(r^{t-1})$ by (6). Hence,

$$w^*(w_2^T)(r^{t-1}) = w^{*,T+1}(r^{t-1}). \tag{18}$$

For $t \leq T$, by (15), we have

$$w^{*,T+1}(r^t) = (1-\delta)\, u\left(\bar{\mu}(r^t)\right) + \delta \mathbb{E}\left[w^{T+1}(r^{t+1}) | r^t\right].$$

By (16),

$$\begin{aligned} w^{T+1}(r^t) &= 1_{\{w^{\bar{\mu}}(r^t) \in W_1\}} \left\{ p^{T+1}(r^t) w^{*,T+1}(r^t) + \left(1 - p^{T+1}(r^t)\right) \bar{w}^1 \right\} + 1_{\{w^{\bar{\mu}}(r^t) \notin W_1\}} \bar{w}^1 \\ &= 1_{\{w^{\bar{\mu}}(r^t) \in W_1\}} \left\{ p^{T+1}(r^t) w^*(w^T)(r^t) + \left(1 - p^{T+1}(r^t)\right) \bar{w}^1 \right\} + 1_{\{w^{\bar{\mu}}(r^t) \notin W_1\}} \bar{w}^1, \end{aligned}$$

where the second equality follows from (18) for $t = T$, and follows by induction for $t < T$.

Recall that p^{T+1} is defined by (17). On the other hand, $p(w^T)(r^t)$ is defined in (8) using

$w^* = w^*(w^T)$. Since $w^{*,T+1}(r^t) = w^*(w_2^T)(r^t)$, we have $p^{T+1}(r^t) = p(w^T)(r^t)$. Hence,

$$\begin{aligned} w^{T+1}(r^t) &= 1_{\{w^{\bar{\mu}}(r^t)\in W_1\}}\left\{p(w^T)(r^t)w^*(w^T)(r^t) + \left(1-p(w^T)(r^t)\right)\bar{w}^1\right\} + 1_{\{w^{\bar{\mu}}(r^t)\notin W_1\}}\bar{w}^1. \\ &= F(w^T)(r^t), \end{aligned}$$

as desired. ■

As $w^{*,T}(r^t)$, $w^T(r^t)$, and $p^T(r^t)$ converge to $w^*(r^t)$, $w(r^t)$, and $p(r^t)$ by Lemma 7, the following lemma implies Lemma 6:

Lemma 8 *For all $t = 1, ..., T-1$, if $w^{\bar{\mu}}(r^t) \in W_1$, then $p^T(r^t)w^{*,T}(r^t) + \left(1-p^T(r^t)\right)\bar{w}^1$ Pareto dominates $w^{\bar{\mu}}(r^t)$.*

Proof. For $t = T-1$, the claim is immediate since $w^{*,T}(r^t) = w^{\bar{\mu}}(r^t)$ and so $p^T(r^t) = 1$.

Suppose that the claim holds for each period $\tau \geq t+1$. We show that it also hold for period t. By construction, $p^T(r^t)w_2^{*,T}(r^t) + \left(1-p^T(r^t)\right)\bar{w}_2^1 \geq w_2^{\bar{\mu}}(r^t)$. Thus, it suffices to show that $p^T(r^t)w_1^{*,T}(r^t) + \left(1-p^T(r^t)\right)\bar{w}_1^1 \geq w_1^{\bar{\mu}}(r^t)$.

Note that

$$\begin{aligned} & w^{*,T}(r^t) \\ = \; & (1-\delta)\, u\left(\bar{\mu}(r^t)\right) + \delta\mathbb{E}\left[w^T(r^{t+1})|r^t\right] \\ = \; & (1-\delta)\, u\left(\bar{\mu}(r^t)\right) + \delta\left\{\begin{array}{c} \sum_{r^{t+1}:w^{\bar{\mu}}(r^{t+1})\in W_1} \Pr^{\bar{\mu}}\left(r^{t+1}|r^t\right)\left\{p^T(r^{t+1})w^{*,T}(r^{t+1}) + \left(1-p^T(r^{t+1})\right)\bar{w}^1\right\} \\ + \sum_{r^{t+1}:w^{\bar{\mu}}(r^{t+1})\notin W_1} \Pr^{\bar{\mu}}\left(r^{t+1}|r^t\right)\bar{w}^1 \end{array}\right\}, \end{aligned}$$

while

$$w^{\bar{\mu}}(r^t) = (1-\delta)\, u\left(\bar{\mu}(r^t)\right) + \delta\sum_{r^{t+1}} \Pr^{\bar{\mu}}\left(r^{t+1}|r^t\right) w^{\bar{\mu}}(r^t).$$

Hence,

$$\begin{aligned} & w^{*,T}(r^t) - w^{\bar{\mu}}(r^t) \\ = \; & \delta\left\{\begin{array}{c} \sum_{r^{t+1}:w^{\bar{\mu}}(r^{t+1})\in W_1} \Pr^{\bar{\mu}}\left(r^{t+1}|r^t\right)\left\{p^T(r^{t+1})w^{*,T}(r^{t+1}) + \left(1-p^T(r^{t+1})\right)\bar{w}^1 - w^{\bar{\mu}}(r^{t+1})\right\} \\ + \sum_{r^{t+1}:w^{\bar{\mu}}(r^{t+1})\notin W_1} \Pr^{\bar{\mu}}\left(r^{t+1}|r^t\right)\left\{\bar{w}^1 - w^{\bar{\mu}}(r^{t+1})\right\} \end{array}\right\}. \end{aligned}$$

When $w^{\bar{\mu}}(r^{t+1}) \in W_1$, the inductive hypothesis implies that

$$p^T(r^{t+1})w^{*,T}(r^{t+1}) + \left(1 - p^T(r^{t+1})\right)\bar{w}^1 - w^{\bar{\mu}}(r^{t+1}) \geq 0.$$

On the other hand, note that

$$\sum_{r^{t+1}:w^{\bar{\mu}}(r^{t+1})\notin W_1} \Pr^{\bar{\mu}}\left(r^{t+1}|r^t\right)\left\{\bar{w}^1 - w^{\bar{\mu}}(r^{t+1})\right\} = l(r^t)(\bar{w}^1 - \tilde{w}(r^t))$$

for some number $l(r^t) \geq 0$ and vector $\tilde{w}(r^t) \notin W_1$. In total, we have

$$w^{*,T}(r^t) = w^{\bar{\mu}}(r^t) + l(r^t)(\bar{w}^1 - \hat{w}(r^t)) \tag{19}$$

for some number $l(r^t) \geq 0$ and vector $\hat{w}(r^t) \leq \tilde{w}(r^t) \notin W_1$. Since $\bar{w}_1^1 \geq \hat{w}_1(r^t)$, if $\bar{w}_1^1 \geq w_1^{\bar{\mu}}(r^t)$ then (19) implies that $\min\left\{w_1^{*,T}(r^t), \bar{w}_1^1\right\} \geq w_1^{\bar{\mu}}(r^t)$, and therefore $p^T(r^t)w_1^{*,T}(r^t) + \left(1 - p^T(r^t)\right)\bar{w}_1^1 \geq w_1^{\bar{\mu}}(r^t)$. In addition, if $w^{\bar{\mu}}(r^{t+1}) \in W_1$ with probability one, then the inductive hypothesis implies that $w_2^{*,T}(r^t) \geq w_2^{\bar{\mu}}(r^t)$, and therefore $p^T(r^t) = 1$ and

$$\begin{aligned} p^T(r^t)w_1^{*,T}(r^t) + \left(1 - p^T(r^t)\right)\bar{w}_1^1 &= w_1^{*,T}(r^t) \\ &= w_1^{\bar{\mu}}(r^t) + l(r^t)(\bar{w}_1^1 - \hat{w}_1(r^t)) \\ &\geq w_1^{\bar{\mu}}(r^t). \end{aligned}$$

Hence, it remains only to consider the case where $\bar{w}_1^1 < w_1^{\bar{\mu}}(r^t)$ and $l(r^t) > 0$.

In this case, take a normal vector λ^1 of the supporting hyperplane of $u(A)$ at $\bar{w}^1$. We have $\lambda_1^1 \geq 0$ and $\lambda_2^1 > 0$, and in addition (as $\hat{w}(r^t) \leq \tilde{w}(r^t) \in u(A)$ and $w^{\bar{\mu}}(r^t) \in u(A)$)

$$\begin{aligned} \lambda^1 \cdot \left(\bar{w}^1 - \hat{w}(r^t)\right) &\geq 0, \\ \lambda^1 \cdot \left(\bar{w}^1 - w^{\bar{\mu}}(r^t)\right) &\geq 0. \end{aligned}$$

As $\bar{w}_1^1 - \hat{w}_1(r^t) > 0$ and $\bar{w}_1^1 - w_1^{\bar{\mu}}(r^t) < 0$, we have

$$\frac{\hat{w}_2(r^t) - \bar{w}_2^1}{\bar{w}_1^1 - \hat{w}_1(r^t)} \leq \frac{\lambda_1^1}{\lambda_2^1} \leq \frac{\bar{w}_2^1 - w_2^{\bar{\mu}}(r^t)}{w_1^{\bar{\mu}}(r^t) - \bar{w}_1^1}.$$

Next, by (19), the slope of the line from $w^{\bar{\mu}}(r^t)$ to $w^{*,T}(r^t)$ equals the slope of the line from $\hat{w}(r^t)$ to $\bar{w}^1$. Hence,

$$\frac{w_2^{\bar{\mu}}(r^t) - w_2^{*,T}(r^t)}{w_1^{*,T}(r^t) - w_1^{\bar{\mu}}(r^t)} \leq \frac{\bar{w}_2^1 - w_2^{\bar{\mu}}(r^t)}{w_1^{\bar{\mu}}(r^t) - \bar{w}_1^1}.$$

In this inequality, the denominator of the left-hand side and the numerator of the right-hand side are both positive: $w_1^{*,T}(r^t) > w_1^{\bar{\mu}}(r^t)$ by (19) and $l(r^t) > 0$, while $\bar{w}_2^1 > w_2^{\bar{\mu}}(r^t)$ because $w^{\bar{\mu}}(r^t) \in W_1$ and $w^{\bar{\mu}}(r^t) \neq \bar{w}_2^1$. Therefore, the inequality is equivalent to

$$\frac{w_2^{\bar{\mu}}(r^t) - w_2^{*,T}(r^t)}{\bar{w}_2^1 - w_2^{\bar{\mu}}(r^t)} \leq \frac{w_1^{*,T}(r^t) - w_1^{\bar{\mu}}(r^t)}{w_1^{\bar{\mu}}(r^t) - \bar{w}_1^1}.$$

Now, let $q \in [0,1]$ be the number such that

$$qw_1^{*,T}(r^t) + (1-q)\,\bar{w}_1^1 = w_1^{\bar{\mu}}(r^t).$$

Note that

$$1 - p^T(r^t) = \frac{w_2^{\bar{\mu}}(r^t) - w_2^{*,T}(r^t)}{\bar{w}_2^1 - w_2^{\bar{\mu}}(r^t)},$$

while

$$1 - q = \frac{w_1^{*,T}(r^t) - w_1^{\bar{\mu}}(r^t)}{w_1^{\bar{\mu}}(r^t) - \bar{w}_1^1}.$$

Hence, $p^T(r^t) > q$. Finally, we have seen that $\bar{w}_1^1 \leq w_1^{\bar{\mu}}(r^t) \leq w_1^{*,T}(r^t)$, so we have

$$p^T(r^t)w_1^{*,T}\left(r^t\right) + \left(1 - p^T(r^t)\right)\bar{w}_1^1 \geq qw_1^{*,T}\left(r^t\right) + (1-q)\,\bar{w}_1^1 = w_1^{\bar{\mu}}(r^t).$$

■

11.8 Proof of Proposition 2

We wish to show that $\mathrm{supp}(\delta) = \mathcal{A}^\infty$ whenever $\mathrm{int}\left(\bigcap_{i\in I} \bar{W}_i \cap u\left(\mathcal{A}^\infty\right)\right) \neq \emptyset$. The proof consists of two steps: we show that $\mathrm{supp}(\delta) \subseteq \mathcal{A}^\infty$ in Section 11.8.1, and show that $\mathrm{supp}(\delta) \supseteq \mathcal{A}^\infty$ in Section 11.8.2

11.8.1 Proof of $\mathrm{supp}(\delta) \subseteq \mathcal{A}^\infty$

Recall that $\mathcal{A}^\infty$ is the largest fixed point of the monotone operator S. Hence, to prove that $\mathrm{supp}(\delta) \subseteq \mathcal{A}^\infty$, it suffices to show that there exists a set $\mathrm{supp}_{\varepsilon\downarrow 0}(\delta)$ such that $\mathrm{supp}(\delta) \subseteq \mathrm{supp}_{\varepsilon\downarrow 0}(\delta)$ and $\mathrm{supp}_{\varepsilon\downarrow 0}(\delta) = S(\mathrm{supp}_{\varepsilon\downarrow 0}(\delta))$.

To this end, it will be useful to consider ε-sequential equilibria in the repeated game with monitoring structure p^*. We say that an assessment is an ε-sequential equilibrium if it is consistent, and, for each player i and (on- or off-path) history h_i^t, player i's deviation gain at history h_i^t is no more than ε. Let

$$\mathrm{supp}_\varepsilon^1(\delta) = \left\{ a \in A : \left(\begin{array}{c} \text{there exist an } \varepsilon\text{-sequential equilibrium } \mu \\ \text{such that } a \in \mathrm{supp}\left(\mu(\emptyset)\right) \end{array} \right) \right\}$$

be the on-path support of actions in the initial period in ε-sequential equilibrium with monitoring structure p^*. In addition, let

$$\mathrm{supp}_\varepsilon^t(\delta) = \left\{ a \in A : \left(\begin{array}{c} \text{there exist an } \varepsilon\text{-sequential equilibrium } \mu \text{ and history } h_m^t \\ \text{such that } a \in \mathrm{supp}\left(\mu(h_m^t)\right) \end{array} \right) \right\}$$

be the (possibly off-path) support of actions in period $t \geq 1$.

We will establish the following equality in Lemma 10:

$$\bigcap_{\varepsilon>0} \bigcup_t \mathrm{supp}_\varepsilon^t(\delta) = \bigcap_{\varepsilon>0} \mathrm{supp}_\varepsilon^1(\delta). \tag{20}$$

As a preliminary result toward the proof of Lemma 10, we will show in Lemma 9 that $\bigcap_{\varepsilon>0} \mathrm{supp}_\varepsilon^1(\delta)$ is a product set.

Let $\text{supp}_{\varepsilon\downarrow 0}(\delta) \equiv \bigcap_{\varepsilon>0} \text{supp}_\varepsilon^1(\delta)$. Since a sequential equilibrium is an ε-sequential equilibrium for every ε, we have $\text{supp}(\delta) \subseteq \bigcap_{\varepsilon>0} \bigcup_t \text{supp}_\varepsilon^t(\delta)$. Together with (20), this implies that $\text{supp}(\delta) \subseteq \text{supp}_{\varepsilon\downarrow 0}(\delta)$. Hence, it remains to show that $\text{supp}_{\varepsilon\downarrow 0}(\delta) = S(\text{supp}_{\varepsilon\downarrow 0}(\delta))$.

Finally, as $\text{supp}_{\varepsilon\downarrow 0}(\delta)$ is a product set by Lemma 9 and $S(\bar{A}) \subseteq \bar{A}$ for every product set $\bar{A}$, it suffices to show that $\text{supp}_{\varepsilon\downarrow 0}(\delta) \subseteq S(\text{supp}_{\varepsilon\downarrow 0}(\delta))$. This is established in Lemma 11, completing the proof.

We now prove Lemmas 9, 10, and 11.

Given a mediator's strategy μ, let μ_i denote the marginal distribution of player i's recommendation, and let

$$\text{supp}_{i,\varepsilon}^1(\delta) = \left\{ a_i \in A_i : \begin{pmatrix} \text{there exists an } \varepsilon\text{-sequential equilibrium } \mu \\ \text{such that } a_i \in \text{supp}\left(\mu_i(\emptyset)\right) \end{pmatrix} \right\}.$$

Lemma 9 *$\bigcap_{\varepsilon>0} \text{supp}_\varepsilon^1(\delta)$ is a product set: $\bigcap_{\varepsilon>0} \text{supp}_\varepsilon^1(\delta) = \bigcap_{\varepsilon>0} \prod_{i\in I} \text{supp}_{i,\varepsilon}^1(\delta)$.*

Proof. Since $\bigcap_{\varepsilon>0} \text{supp}_\varepsilon^1(\delta) \subseteq \bigcap_{\varepsilon>0} \prod_{i\in I} \text{supp}_{i,\varepsilon}^1(\delta)$, we are left to show $\bigcap_{\varepsilon>0} \prod_{i\in I} \text{supp}_{i,\varepsilon}^1(\delta) \subseteq \bigcap_{\varepsilon>0} \text{supp}_\varepsilon^1(\delta)$. To this end, fix an arbitrary ε-sequential equilibrium strategy μ. It suffices to show that, for each $\varepsilon > 0$, there exists a 2ε-sequential equilibrium strategy μ_{product} such that $\text{supp}\left(\mu_{\text{product}}(\emptyset)\right) = \prod_{i\in I} \text{supp}(\mu_i(\emptyset))$. (This is sufficient because it implies that $\bigcap_{\varepsilon>0} \prod_{i\in I} \text{supp}_{i,\varepsilon}^1(\delta) \subseteq \bigcap_{\varepsilon>0} \text{supp}_{2\varepsilon}^1(\delta) = \bigcap_{\varepsilon>0} \text{supp}_\varepsilon^1(\delta)$.)

For $\eta > 0$, define the period 1 recommendation distribution under $\mu_{\text{product}}(\emptyset)$ by

$$\mu_{\text{product}}(\emptyset)(a_1) = \begin{cases} (1-\eta)\mu(\emptyset)(a_1) & \text{if } a_1 \in \text{supp}(\mu(\emptyset)), \\ \dfrac{\eta}{\left|\prod_{i\in I} \text{supp}(\mu_i(\emptyset))\right| - |\text{supp}(\mu(\emptyset))|} & \begin{array}{c}\text{if } a_1 \in \prod_{i\in I} \text{supp}(\mu_i(\emptyset)) \\ \text{but } a_1 \notin \text{supp}(\mu(\emptyset)),\end{array} \\ 0 & \text{otherwise.} \end{cases}$$

Note that the mediator recommends action a_1 with positive probability if $a_{i,1}$ is in the support of the marginal distribution of player i's recommendation under μ. In particular, $\text{supp}\left(\mu_{\text{product}}(\emptyset)\right) = \prod_{i\in I} \text{supp}(\mu_i(\emptyset))$.

In subsequent periods $t \geq 2$, if $a_1 \in \text{supp}(\mu(\emptyset))$ was recommended in period 1, then the mediator recommends actions according to $\mu_{\text{product}}(h_m^t) = \mu(h_m^t)$. If $a_1 \in \prod_{i \in I} \text{supp}(\mu_i(\emptyset))$ but $a_1 \notin \text{supp}(\mu(\emptyset))$ was recommended, then the mediator "resets" the history and recommends actions according to μ as if the game started from period 2: formally, $\mu_{\text{product}}(h_m^t) = \mu(\tilde{h}_m^{t-1})$ with $\left((r_1, a_1), \tilde{h}_m^{t-1}\right) = h_m^t$.

As $\eta \to 0$, player i's belief about $a_{-i,1}$ given $a_{i,1} \in \text{supp}(\mu_i(\emptyset))$ under μ_{product} converges to her belief given $a_{i,1}$ in the original equilibrium μ. Since μ_{product} is an ε-sequential equilibrium, this implies that player i's deviation gain in period 1 converges to at most ε as $\eta \to 0$. In addition, if $a_1 \in \text{supp}(\mu(\emptyset))$, then continuation play from period 2 on is the same under μ_{product} and μ. Furthermore, if $a_1 \in \prod_{i \in I} \text{supp}(\mu_i(\emptyset))$ but $a_1 \notin \text{supp}(\mu(\emptyset))$, then continuation play from period 2 on under μ_{product} is the same as play from period 1 on under μ. In sum, for any $\varepsilon > 0$, there exists $\eta > 0$ such that player i's deviation gain at any history under μ_{product} is at most 2ε, so μ_{product} is an 2ε-sequential equilibrium strategy. ■

Lemma 10 $\bigcap_{\varepsilon>0} \bigcup_t \text{supp}_\varepsilon^t(\delta) = \bigcap_{\varepsilon>0} \text{supp}_\varepsilon^1(\delta)$.

Proof. Since $\bigcap_{\varepsilon>0} \bigcup_t \text{supp}_\varepsilon^t(\delta) \supseteq \bigcap_{\varepsilon>0} \text{supp}_\varepsilon^1(\delta)$ by definition, we are left to show $\bigcap_{\varepsilon>0} \bigcup_t \text{supp}_\varepsilon^t(\delta) \subseteq \bigcap_{\varepsilon>0} \text{supp}_\varepsilon^1(\delta)$. Fix an ε-sequential equilibrium strategy μ together with a consistent belief system β. As in Lemma 9, it suffices to show that, for each t and ε, there exists a 2ε-sequential equilibrium strategy $\bar{\mu}$ with consistent belief system $\bar{\beta}$ such that $\text{supp}\left(\bar{\mu}(\emptyset)\right) = \bigcup_{h_m^t} \text{supp}(\mu(h_m^t))$, where the union is taken over all histories h_m^t, including off-path histories.

Fix t and ε arbitrarily. Since (μ, β) is an ε-sequential equilibrium, there exists $\eta > 0$ such that, for each strategy-belief system pair $\left(\bar{\mu}, \bar{\beta}\right)$, if $\left(\bar{\mu}, \bar{\beta}\right)$ satisfies

$$\bar{\mu}(h_m^\tau) = \mu(h_m^\tau) \tag{21}$$

and

$$\bar{\beta}(h_m^\tau | h_i^\tau) \in \left[(1-\eta)^2 \beta(h_m^\tau | h_i^\tau), \frac{1}{(1-\eta)^2} \beta(h_m^\tau | h_i^\tau)\right] \tag{22}$$

for each i, τ, h_i^τ, and h_m^τ—that is, if the recommendation schedule is the same as μ and the posterior is close to β after each history—then $(\bar{\mu}, \bar{\beta})$ is 2ε-sequentially rational. Fix such an $\eta > 0$.

Since μ is an ε-sequential equilibrium, there exists a sequence of completely mixed strategy profiles $(\sigma^k)_{k=1}^{\infty}$ and associated belief systems $(\beta^k)_{k=1}^{\infty}$ such that σ^k converges to the obedient strategy profile at μ and β^k converges to β. Consider the following mediator's strategy $\bar{\mu}^n$. In period 1, the mediator draws a "fictitious" history h_m^t according to $\beta^n(h_m^t)$, the probability of h_m^t given σ^n. Given h_m^t, the mediator draws r according to $\mu(h_m^t)$. In period 1, the mediator sends (h_i^t, r_i) to each player i, where h_i^t is player i's component in h_m^t. Note that, since σ^n is completely mixed, we have $\text{supp}(\bar{\mu}(\emptyset)) = \bigcup_{h_m^t} \text{supp}(\mu(h_m^t))$.

In period $\tau \geq 2$, let $h_m^{2,\tau} = \left(a_1, (r_{\tau'}, a_{\tau'})_{\tau'=2}^{\tau-1}\right)$ with $h_m^{2,2} = a_1$ be the mediator's history at the beginning of period τ except for the recommendation profile r in period 1. The mediator recommends actions as if the history $h_m^{2,\tau}$ happened after (h_m^t, r) in the original equilibrium μ: that is, she recommends action profiles according to $\mu(h_m^t, r, h_m^{2,\tau})$, where $(h_m^t, r, h_m^{2,\tau})$ denotes the concatenation of h_m^t, r, and $h_m^{2,\tau}$. Similarly, let $h_i^t = (r_{i,s}, a_s)_{s=1}^{t-1}$ be player i's fictitious history, and let $h_i^{2,\tau} = \left(a_1, (r_{i,\tau'}, a_{\tau'})_{\tau'=2}^{\tau-1}\right)$ be her history at the beginning of period τ.

When we view $\mathfrak{h}_m^\tau = (h_m^t, r, h_m^{2,\tau})$ and $\mathfrak{h}_i^\tau = \left(h_i^t, r_i, h_i^{2,\tau}\right)$ as the history (including the fictitious one drawn at the beginning of the game), this new strategy $\bar{\mu}^n$ satisfies (21). Hence, it suffices to show that we can construct a consistent belief system $\bar{\beta}$ which satisfies (22). To this end, denote player i's belief about the mediator's history $\mathfrak{h}_m^\tau$ when player i's history in period τ is $\mathfrak{h}_i^\tau = \left(h_i^t, r_i, h_i^{2,\tau}\right)$ by $\Pr(\mathfrak{h}_m^\tau \mid \mathfrak{h}_i^\tau)$.

We construct the following sequence of perturbed full-support strategies: each player i plays $\sigma_i^k(h_i^t, r_i)$ in period 1 and then plays $\sigma_i^k(h_i^t, r_i, h_i^{2,\tau}, r_{i,\tau})$ in period $\tau \geq 2$. Recall that $(\sigma^k)_{k=1}^{\infty}$ is the sequence converging to the original equilibrium σ.

Let $\Pr^{\sigma^n, \sigma^k}(\mathfrak{h}_m^\tau \mid \mathfrak{h}_i^\tau)$ be player i's belief about $\mathfrak{h}_m^\tau$ conditional on her history $\mathfrak{h}_i^\tau$, given that the fictitious history h_m^t is drawn from β^n at the beginning of the game and the players take σ^k in the subsequent periods. We have

$$\begin{aligned}
\Pr^{\sigma^n,\sigma^k}\left(\mathfrak{h}_m^\tau \mid \mathfrak{h}_i^\tau\right) &= \frac{\Pr^{\sigma^k}\left(h_m^{2,\tau} \mid h_m^t, r\right)\Pr^\mu\left(r \mid h_m^t\right)\Pr^{\sigma^n}\left(h_m^t\right)}{\Pr^{\sigma^k}\left(h_i^{2,\tau} \mid h_i^t, r_i\right)\Pr^\mu\left(r_i \mid h_i^t\right)\Pr^{\sigma^n}\left(h_i^t\right)} \\
&= \frac{\Pr^{\sigma^k}\left(h_m^{2,\tau} \mid h_m^t, r\right)\Pr^\mu\left(r \mid h_m^t\right)}{\sum_{\tilde{h}_m^t,\tilde{r}_{-i}}\Pr^{\sigma^k}\left(h_i^{2,\tau} \mid \tilde{h}_m^t, r_i, \tilde{r}_{-i}\right)\Pr^\mu\left(r_i, \tilde{r}_{-i} \mid \tilde{h}_m^t\right)\Pr^{\sigma^n}\left(\tilde{h}_m^t \mid h_i^t\right)}\frac{\Pr^{\sigma^n}\left(h_m^t\right)}{\Pr^{\sigma^n}\left(h_i^t\right)}.
\end{aligned}$$

By consistency of (μ, β), for sufficiently large n, we have

$$(1-\eta)\,\beta\left(h_m^t \mid h_i^t\right) \le \frac{\Pr^{\sigma^n}\left(h_m^t\right)}{\Pr^{\sigma^n}\left(h_i^t\right)} = \beta^n\left(h_m^t \mid h_i^t\right) \le \frac{1}{1-\eta}\beta\left(h_m^t \mid h_i^t\right)$$

for all h_m^t and h_i^t. Moreover,

$$(1-\eta)\,\beta\left(\tilde{h}_m^t \mid h_i^t\right) \le \Pr^{\sigma^n}\left(\tilde{h}_m^t \mid h_i^t\right) = \beta^n\left(\tilde{h}_m^t \mid h_i^t\right) \le \frac{1}{1-\eta}\beta\left(\tilde{h}_m^t \mid h_i^t\right)$$

for each $\tilde{h}_m^t$ and h_i^t.

Fix such a sufficiently large n, and define

$$\bar{\beta}\left(\mathfrak{h}_m^\tau \mid \mathfrak{h}_i^\tau\right) = \lim_{k\to\infty}\Pr^{\sigma^n,\sigma^k}\left(\mathfrak{h}_m^\tau \mid \mathfrak{h}_i^\tau\right).$$

Since the belief system β is consistent, we have

$$\lim_{k\to\infty}\Pr^{\sigma^k}\left(h_m^{2,\tau} \mid h_m^t, r\right) = \beta(h_m^{2,\tau} \mid h_m^t, r)$$

Hence, for each $\mathfrak{h}_m^\tau$ and $\mathfrak{h}_i^\tau$, we have

$$\begin{aligned}
\bar{\beta}\left(\mathfrak{h}_m^\tau \mid \mathfrak{h}_i^\tau\right) &= \frac{\beta\left(h_m^{2,\tau} \mid h_m^t, r\right)\Pr^\mu\left(r \mid h_m^t\right)\Pr^{\sigma^n}\left(h_m^t \mid h_i^t\right)}{\sum_{\tilde{h}_m^t,\tilde{r}_{-i}}\beta\left(h_i^{2,\tau} \mid \tilde{h}_m^t, r_i, \tilde{r}_{-i}\right)\Pr^\mu\left(r_i, \tilde{r}_{-i} \mid \tilde{h}_m^t\right)\Pr^{\sigma^n}\left(\tilde{h}_m^t \mid h_i^t\right)} \\
&= \left[(1-\eta)^2\,\beta\left(\mathfrak{h}_m^\tau \mid \mathfrak{h}_i^\tau\right), \frac{1}{(1-\eta)^2}\beta\left(\mathfrak{h}_m^\tau \mid \mathfrak{h}_i^\tau\right)\right].
\end{aligned}$$

That is, $\bar{\beta}$ satisfies (22), as desired. ■

Lemma 11 $\text{supp}_{\varepsilon\downarrow 0}(\delta) \subseteq S\left(\text{supp}_{\varepsilon\downarrow 0}(\delta)\right)$.

Proof. Fix $a \in \text{supp}_{\varepsilon\downarrow 0}(\delta)$ arbitrarily. Since $\text{supp}_{\varepsilon\downarrow 0}(\delta) = \lim_{\varepsilon\to 0}\text{supp}^1_{\varepsilon}(\delta)$ by definition, there exists a sequence of ε-sequential equilibria $\{\mu^{\varepsilon_n}\}_{n=1}^{\infty}$ with $\lim_{n\to\infty}\varepsilon_n = 0$ such that $a \in \lim_{n\to\infty}\text{supp}(\mu^{\varepsilon_n}(\{\emptyset\}))$. Taking a subsequence if necessary, we can assume that $a \in \text{supp}(\mu^{\varepsilon_n}(\{\emptyset\}))$ for each n.

For each n, incentive compatibility of a player i who is recommended a_i in period 1 implies the following three conditions:

1. Let $\alpha^n_{-i} \in \Delta\left(\left(\text{supp}^1_{\varepsilon_n}(\delta)\right)_{-i}\right)$ be the conditional distribution of player i's opponents' actions when a_i is recommended to player i. In addition, let $w^n_i\left(\hat{a}_i|a_i\right)$ be player i's expected continuation payoff when she is recommended a_i and plays $\hat{a}_i$. Then

$$(1-\delta)\,u_i\left(a_i,\alpha^n_{-i}\right) + \delta w^n_i\left(a_i|a_i\right) + \varepsilon_n \geq \max_{\hat{a}_i\in A_i}\left\{(1-\delta)\,u_i\left(\hat{a}_i,\alpha^n_{-i}\right) + \delta w^n_i\left(\hat{a}_i|a_i\right)\right\}.$$

2. Player i's continuation payoff does not exceed her best feasible payoff with the support of actions restricted to $\bigcup_t \text{supp}^t_{\varepsilon_n}(\delta)$:

$$w^n_i\left(a_i|a_i\right) \leq \max_{\bar{a}\in\bigcup_t \text{supp}^t_{\varepsilon_n}(\delta)} u_i\left(\bar{a}\right).$$

3. For each $\hat{a}_i \in A_i$, by ε-sequential rationality for player i after taking $\hat{a}_i$, player i's continuation payoff after $\hat{a}_i$ is at least her minmax payoff with the support of punishment actions restricted to $\left(\bigcup_t \text{supp}^t_{\varepsilon_n}(\delta)\right)_{-i}$:

$$w^n_i\left(\hat{a}_i|a_i\right) \geq \min_{\hat{\alpha}_{-i}\in\Delta\left(\left(\bigcup_t \text{supp}^t_{\varepsilon_n}(\delta)\right)_{-i}\right)} \max_{\tilde{a}_i\in A_i} u_i\left(\tilde{a}_i,\hat{\alpha}_{-i}\right) - \varepsilon_n.$$

Since $\Delta(A_{-i})$ and $u(A)$ are compact, taking a subsequence if necessary, there exist $\alpha_{-i} = \lim_{n\to\infty}\alpha^n_{-i}$ and $w_i\left(\hat{a}_i|a_i\right) = \lim_{n\to\infty} w^n_i\left(\hat{a}_i|a_i\right)$ for each $\hat{a}_i$. Moreover, since A is finite, there exists N such that for each $n \geq N$, we have

$$\begin{cases} \operatorname{supp}^1_{\varepsilon_n}(\delta) = \lim_{n\to\infty} \operatorname{supp}^1_{\varepsilon_n}(\delta) = \lim_{\varepsilon\to 0} \operatorname{supp}^1_{\varepsilon}(\delta) = \operatorname{supp}_{\varepsilon\downarrow 0}(\delta) = \prod_{i\in I} \operatorname{supp}_{\varepsilon\downarrow 0,i}(\delta), \\ \bigcup_t \operatorname{supp}^t_{\varepsilon_n}(\delta) = \lim_{n\to\infty} \bigcup_t \operatorname{supp}^t_{\varepsilon_n}(\delta) = \lim_{\varepsilon\to 0} \bigcup_t \operatorname{supp}^t_{\varepsilon}(\delta) \underbrace{=}_{\text{by Lemma (10)}} \operatorname{supp}_{\varepsilon\downarrow 0}(\delta) = \prod_{i\in I} \operatorname{supp}_{\varepsilon\downarrow 0,i}(\delta). \end{cases} \tag{23}$$

Hence, for $n \geq N$, we have

$$\begin{aligned} \alpha^n_{-i} &\in \Delta\left(\left(\operatorname{supp}^1_{\varepsilon_n}(\delta)\right)_{-i}\right) = \Delta\left(\left(\operatorname{supp}_{\varepsilon\downarrow 0}(\delta)\right)_{-i}\right) = \Delta\left(\operatorname{supp}_{\varepsilon\downarrow 0,-i}(\delta)\right), \\ w^n_i(a_i|a_i) &\leq \max_{\bar{a}\in\bigcup_t \operatorname{supp}^t_{\varepsilon_n}(\delta)} u_i(\bar{a}) = \max_{\bar{a}\in\operatorname{supp}_{\varepsilon\downarrow 0}(\delta)} u_i(\bar{a}), \\ w^n_i(\hat{a}_i|a_i) &\geq \min_{\hat{\alpha}_{-i}\in\Delta\left(\left(\bigcup_t \operatorname{supp}^t_{\varepsilon_n}(\delta)\right)_{-i}\right)} \max_{\tilde{a}_i\in A_i} u_i(\tilde{a}_i, \hat{\alpha}_{-i}) - \varepsilon_n = \min_{\hat{\alpha}_{-i}\in\Delta\left(\operatorname{supp}_{\varepsilon\downarrow 0,-i}(\delta)\right)} \max_{\tilde{a}_i\in A_i} u_i(\tilde{a}_i, \hat{\alpha}_{-i}) - \varepsilon_n. \end{aligned}$$

Therefore, $\alpha_{-i} = \lim_{n\to\infty} \alpha^n_{-i}$ and $w_i(\hat{a}_i|a_i) = \lim_{n\to\infty} w^n_i(\hat{a}_i|a_i)$ satisfy the following three conditions:

1. If a_i is recommended, player i's opponents play α_{-i}, player i's on-path continuation payoff is $w_i(a_i|a_i)$, and player i's continuation payoff after deviating to $\hat{a}_i$ is $w_i(\hat{a}_i|a_i)$, then player i wants to take a_i:

$$(1-\delta)\, u_i(a_i, \alpha_{-i}) + \delta w_i(a_i|a_i) \geq \max_{\hat{a}_i\in A_i} \left\{(1-\delta)\, u_i(\hat{a}_i, \alpha_{-i}) + \delta w_i(\hat{a}_i|a_i)\right\}, \tag{24}$$

2. Player i's continuation payoff does not exceed her best feasible payoff with the support of actions restricted to $\operatorname{supp}_{\varepsilon\downarrow 0}(\delta)$:

$$w_i(a_i|a_i) \leq \max_{\bar{a}\in\operatorname{supp}_{\varepsilon\downarrow 0}(\delta)} u_i(\bar{a})$$

3. Player i's continuation payoff following a deviation is at least her minmax payoff with

the support of punishment actions restricted to $\mathrm{supp}_{\varepsilon\downarrow 0,-i}(\delta)$:

$$w_i\left(\hat{a}_i|a_i\right) \geq \min_{\hat{\alpha}_{-i}\in\Delta\left(\mathrm{supp}_{\varepsilon\downarrow 0,-i}(\delta)\right)} \max_{\tilde{a}_i\in A_i} u_i\left(\tilde{a}_i,\hat{\alpha}_{-i}\right).$$

Therefore, (24) implies that (11) holds with $\bar{A} = \mathrm{supp}_{\varepsilon\downarrow 0}(\delta)$. Hence, $a_i \in S_i\left(\mathrm{supp}_{\varepsilon\downarrow 0}(\delta)\right)$ for all $i \in I$, so $a \in S\left(\mathrm{supp}_{\varepsilon\downarrow 0}(\delta)\right)$. ■

11.8.2 Proof of $\mathrm{supp}(\delta) \supseteq \mathcal{A}^\infty$

We show that if $\mathrm{int}\left(\bigcap_{i\in I}\bar{W}_i\cap u\left(\mathcal{A}^\infty\right)\right)\neq\emptyset$ then. $\mathcal{A}^\infty\subseteq\mathrm{supp}(\delta)$. The argument is similar to the proof of Lemma 1.

Fix $v \in\mathrm{int}\left(\bigcap_{i\in I}\bar{W}_i\cap u\left(\mathcal{A}^\infty\right)\right)\neq\emptyset$, and let $\mu\in\Delta\left(\mathcal{A}^\infty\right)$ be such that $u\left(\mu\right)=v$ and $\mu\left(r\right)>0$ for all $r\in\mathcal{A}^\infty$. Let α^*_{-i} be a solution to the problem

$$\min_{\hat{\alpha}_{-i}\in\Delta(\mathcal{A}^\infty)} \max_{a_i\in A_i} u_i\left(\hat{a}_i,\alpha_{-i}\right).$$

Let α^ε_{-i} be the following full-support (within $\mathcal{A}^\infty$) approximation of α^*_{-i}: $\alpha^\varepsilon_{-i}=(1-\varepsilon)\,\alpha^*_{-i}+\varepsilon\sum_{a_{-i}\in\mathcal{A}^\infty_{-i}}\frac{a_{-i}}{\left|\mathcal{A}^\infty_{-i}\right|}$. Since $v \in\mathrm{int}\left(\bigcap_{i\in I}\bar{W}_i\right)$, there exists $\varepsilon>0$ such that, for each $i\in I$, we have

$$v_i > \max_{a_i\in A_i} u_i\left(a_i,\alpha^\varepsilon_{-i}\right)+\frac{1-\delta}{\delta}\max_{r\in\mathcal{A}^\infty,a_i\in A_i}\left\{u_i(a_i,r_{-i})-u_i(r)\right\}. \tag{25}$$

In addition, choose $\varepsilon>0$ small enough such that, for each player i, some best response to α^ε_{-i} is included in $\mathcal{A}^\infty_i$: this is always possible, as every best response to α^*_{-i} is included in $\mathcal{A}^\infty_i$.

We can now construct an equilibrium strategy μ^* with $\mathrm{supp}\left(\mu^*\left(\emptyset\right)\right)=\mathcal{A}^\infty$. The construction is identical to that in the proof of Lemma 1, except that μ is recommended on path, and player i's deviations are punished by recommending α^ε_{-i} to her opponents. Incentive compatibility follows from (25), just as it follows from (1) in the proof of Lemma 1.

11.9 Proof of Proposition 5

Consider the following five-player game with common discount factor $\delta = \frac{3}{4}$. Player 1 has two actions $\{1_1, 2_1\}$, players 2 and 4 have three actions $\{1_2, 2_2, 3_2\}$ and $\{1_4, 2_4, 3_4\}$, and player 3 has four actions $\{1_3, 2_3, 3_3, 4_3\}$. Player 5 is a dummy player with one action, and her utility is always equal to $-u_1(a)$.

When player 3 plays 1_3, the payoff matrix for players 1 and 2 is as follows, regardless of player 4's actions:

	1_2	2_2	3_2
1_1	$0, 1$	$20, 7$	$20, 1$
2_1	$-3, 0$	$20, 0$	$20, 6$

In addition, player 3's payoff and player 4's payoff are 0, regardless of (a_1, a_2, a_4).

When player 3 plays 2_3, 3_3, or 4_3, the payoff matrix is as follows, regardless of (a_1, a_2):

	1_4	2_4	3_4
2_3	$0, 1, 0, 0$	$1, 0, 1, 1$	$1, 0, 1, -1$
3_3	$1, 1, 0, 0$	$1, 0, 1, -1$	$1, 0, 1, 1$
4_3	$1, 1, 0, 0$	$1, 0, 1, -1$	$10, 3, 1, 1$

Note that the minmax payoff is zero for every player except player 5 (whose minmax payoff is -20), while the players' maximum deviation gains are $d_1 = 3$, $d_2 = 6$, $d_3 = 1$, $d_4 = 2$, and $d_5 = 0$. Letting $v = (10, 3, 1, 1, -10)$ in the formula for δ^*, we see that a lower bound on δ^* is

$$\max \left\{ \frac{3}{3+10}, \frac{6}{6+3}, \frac{1}{1+1}, \frac{2}{2+1}, \frac{0}{0-10+20} \right\} = \frac{2}{3}.$$

Thus, $\delta > \delta^*$, so $\mathring{W}^* \neq \emptyset$.

Consider the Pareto weight $\lambda = (0, 0, 0, 0, 1)$: that is, we want to maximize player 5's payoff, or equivalently minimize player 1's payoff. We show that player 5's best payoff is strictly less than zero with mediated perfect monitoring, while her best payoff equals zero with some private monitoring structure.

11.9.1 Mediated Perfect Monitoring

We show that $\overline{E_{\text{med}}(\delta)}$ does not include a payoff vector v with $v_1 = 0$. Since $\overline{E_{\text{med}}(\delta)}$ coincides with the set of (possibly weak) sequential equilibrium payoffs when $\mathring{W}^* \neq \emptyset$ (see the proof of Lemma 4), it suffices to show player 1's payoff is strictly positive in every sequential equilibrium with mediated perfect monitoring. To do so, we fix a putative sequential equilibrium with $v_1 = 0$, and establish a series of claims leading to a contradiction.

Claim 1: At any on-path history where player 1 has always played 1_1 in the past (where this includes the null history), player 3 must play 1_3 with probability 1.

Proof: Fix an on-path history h_m^t where player 1 has always played 1_1. Let π denote the ex ante probability of reaching history h_m^t, and let $\Pr(a_i)$ be the probability that action a_i is played in period t conditional on reaching history h_m^t.

Note that player 1's equilibrium payoff is bounded from below by

$$(1-\delta)\,\delta^t \pi \left\{\Pr(2_3)\Pr(2_4 \text{ or } 3_4 \mid 2_3) + \Pr(3_3) + \Pr(4_3)\right\},$$

as if she always plays 1_1 she gets payoff at least 0 in every period and reaches history h_m^t with probability at least π. Hence, $\Pr(3_3) = \Pr(4_3) = 0$.

Given $\Pr(3_3) = \Pr(4_3) = 0$, we must also have $\Pr(2_3) = 0$. To see why, suppose otherwise. Then, we must have $\Pr(2_4 \text{ or } 3_4 \mid 2_3) = 0$, that is, $\Pr(1_4 \mid 2_3) = 1$. Hence, player 4's deviation gain from 1_4 to 2_4 is no less than

$$\Pr(2_3 \mid 1_4) = \frac{\Pr(1_4 \mid 2_3)\Pr(2_3)}{\Pr(1_4)} = \frac{\Pr(2_3)}{\Pr(1_4)}.$$

Therefore, to ensure that player 4 does not deviate, we must guarantee him a payoff $\frac{1-\delta}{\delta}\frac{\Pr(2_3)}{\Pr(1_4)}$ whenever 1_4 is recommended. But then player 4's ex ante payoff at history h_m^t (before his recommendation is drawn) is at least

$$\Pr(1_4)\left\{(1-\delta)\,0 + \frac{1-\delta}{\delta}\frac{\Pr(2_3)}{\Pr(1_4)}\right\} + \Pr(a_4 \neq 1_4)0 > 0,$$

since player 4 can always guarantee a payoff of 0 after any recommendation by playing 1_4 forever. Hence, player 4's equilibrium continuation payoff is strictly positive. But, by feasibility, this implies that player 1's equilibrium continuation payoff is also strictly positive, which is a contradiction. Therefore, we must have $\Pr(2_3) = \Pr(3_3) = \Pr(4_3) = 0$, or equivalently $\Pr(1_3) = 1$.

Claim 2: At any on-path history where player 1 has always played 1_1 in the past, player 2 must play 1_2 with probability 1.

Proof: Given that $\Pr(1_3) = 1$, this follows by the same argument as Claim 1: if $\Pr(2_2)$ or $\Pr(3_2)$ were positive, then player 1 could guarantee a positive payoff by always playing 1_1.

Claim 3: At any on-path history where player 1 has always played 1_1 in the past, player 1 must play each of 1_1 and 2_1 with strictly positive probability.

Proof: Given that $\Pr(1_3) = 1$ and $\Pr(2_1) = 1$, player 2's deviation gain is 6 if player 1 plays a pure action (either 1_1 or 2_1). Therefore, to ensure that player 2 does not deviate, his continuation payoff must be at least $\frac{1-\delta}{\delta}(6) = 2$. But, by feasibility, this implies that player 1's continuation payoff must be at least $\frac{10}{3}$ (noting that the line segment between $(0,1)$ and $(20,7)$ has slope $\frac{10}{3}$). Hence, even if player 1 receives payoff -3 in period t, her payoff from period t on is at least $(1-\delta)(-3) + \delta\left(\frac{10}{3}\right) = \frac{7}{4} > 0$. So player 1 could guarantee a positive payoff by playing 1_1 up to period t and then following her equilibrium strategy.

Claim 4: At any on-path history where player 1 has always played 1_1 in the past, player 1's continuation payoff must equal 0 after 1_1 is recommended and must equal 1 after 2_1 is recommended.

Proof: Player 1's continuation payoff cannot be less than her minmax payoff of 0. If her continuation payoff after 1_1 is recommended were strictly positive, then she could guarantee a positive payoff by always playing 1_1. Hence, her continuation payoff after 1_1 is recommended must equal 0. If her continuation payoff after 2_1 is recommended were greater than 1, then she could guarantee a positive payoff by playing 1_1 up to period t and then following her equilibrium strategy (as $(1-\delta)(-3) + \delta(1) = 0$). If her continuation payoff after 2_1 is recommended were less than 1, she would deviate from 2_1 to 1_1.

Claim 5: Let v^* *be the supremum of player 2's continuation payoff over all on-path histories where player 1's continuation payoff equals* 1. *Then* $v^* \geq \frac{4}{3}$.

Proof: Let v be the supremum of player 2's continuation payoff over all on-path histories where player 1's continuation payoff equals 0. Let P be the set of mixing probabilities p such that, at some on-path history h_m^t where player 1 has always played 1_1 in the past, player 1 plays $p1_1 + (1-p)\,2_1$. As $\Pr(1_3) = 1$ and $\Pr(2_1) = 1$ at such a history, player 2's deviation gain is at least

$$\frac{1-\delta}{\delta}\max\{6p, 6-6p\} = \max\{2p, 2-2p\}.$$

Therefore, to ensure that player 2 does not deviate, his continuation payoff at such a history must be at least $\max\{2p, 2-2p\}$. Hence,

$$v \geq \max\{2p, 2-2p\} \text{ for all } p \in P.$$

On the other hand, by the definition of v and v^* and Claim 4, we have

$$v \leq \sup_{p\in P}(1-\delta)\,p + \delta p v + \delta(1-p)\,v^*.$$

Hence, there exists a number $p \in [0,1]$ such that

$$(1-\delta)\,p + \delta p v + \delta(1-p)\,v^* \geq \max\{2p, 2-2p\},$$

or equivalently

$$v^* \geq \frac{1-\delta p}{\delta(1-p)}\max\{2p, 2-2p\} - \frac{(1-\delta)\,p}{\delta(1-p)}.$$

It is straightforward to show that the right-hand side of this inequality is minimized at $p = \frac{1}{2}$, which yields

$$v^* \geq \frac{1}{\delta} = \frac{4}{3}.$$

Finally, note that the payoff vector $(1, v^*)$ is not in the feasible set for any $v^* \geq \frac{4}{3}$, as $\frac{13}{10}$

is the largest payoff of player 2's that is consistent with player 1 receiving payoff 1. We thus obtain a contradiction.

11.9.2 Private Monitoring

Suppose that players 1, 2, 3, and 5 observe everyone's actions perfectly, while player 4 only observes her own action and payoff and player 2's action. We claim that $(0,1,0,0,0)$ is a sequential equilibrium payoff with this private monitoring structure.

On the equilibrium path, in period 1, players take $(1_1, 1_2, 1_3, 1_4)$ and $(2_1, 1_2, 1_3, 1_4)$ with probabilities $\frac{1}{2}, \frac{1}{2}$. From period 2 on, player 3 takes 2_3 if player 1 took 1_1 in period 1, and she takes 3_3 if player 1 took 2_1. Players 1, 2, and 4 take $(1_1, 1_2, 1_4)$ with probability one.

Off the equilibrium path, all players' deviation are ignored, except for deviations by player 2 in period 1 and deviations by player 4 in period $t \geq 2$. If player 2 deviates in period 1 (resp., player 4 deviates in period $t \geq 2$) then, in each period $\tau \geq 2$ (resp., each period $\tau \geq t+1$), players 1 and 2 play 1_1 and 1_2 with probability one, player 3 mixes 2_3 and 3_3 with probabilities $\frac{1}{2}, \frac{1}{2}$, independently across periods, and player 4 mixes 2_4 and 3_4 with probabilities $\frac{1}{2}, \frac{1}{2}$, independently across periods. Note that this action profile is a static Nash equilibrium.

It remains only to check incentive compatibility. Player 3 always plays a static best response. Players 1 and 2 play static best responses from period 2 onward, on and off the equilibrium path. Consider player 4's incentives. In period 1, player 4's payoff equals 0 regardless of her action. Hence, in period 2 player 4 believes that player 1 played $\frac{1}{2}1_1 + \frac{1}{2}2_1$ in period 1, and thus that $\Pr(2_3) = \Pr(3_3) = \frac{1}{2}$ in period 2. In addition, player 4's payoff remains constant at 0 as long as she plays 1_4, so in every period t she continues to believe that $\Pr(2_3) = \Pr(3_3) = \frac{1}{2}$. Hence, playing 1_4 is a static best response, and player 4 is minmaxed forever if she deviates, so following her recommendation is incentive compatible on path. Furthermore, if player 2 deviates in period 1 (or player 4 herself deviates in period $t \geq 2$), then player 4 sees this and anticipates that player 3 will thenceforth play $\frac{1}{2}2_3 + \frac{1}{2}3_3$ independently across periods, so $\frac{1}{2}2_4 + \frac{1}{2}3_4$ is a best response.

We are left to check players 1 and 2's incentives in period 1. For player 1, recalling that $\delta = \frac{3}{4}$, we have

$$\begin{aligned}\text{payoff of taking } 1_1 &= (1-\delta)\,0 + \delta \times \underbrace{0}_{\text{after } 1_1\text{, player 3 takes } 2_3 \text{ with probability one}} \\ &= (1-\delta)(-3) + \delta \times \underbrace{1}_{\text{after } 2_1\text{, player 3 takes } 3_3 \text{ with probability one}} \\ &= \text{payoff of taking } 2_1.\end{aligned}$$

Hence, player 1 is willing to mix between 1_1 and 2_1. For player 2,

$$\begin{aligned}\text{payoff of taking } 1_2 &= (1-\delta)\left(\frac{1}{2}\times 1 + \frac{1}{2}\times 0\right) + \delta \times \underbrace{1}_{\text{whether player 3 takes } 2_3 \text{ or } 3_3\text{, player 2 obtains 1}} \\ &= (1-\delta)\left(\frac{1}{2}\times 7\right) + \delta \times 0 \\ &= \text{payoff of taking } 2_2 \\ &= (1-\delta)\left(\frac{1}{2}\times 1 + \frac{1}{2}\times 6\right) + \delta \times 0 \\ &= \text{payoff of taking } 3_2.\end{aligned}$$

Hence, player 2 is willing to play 1_2.

11.10 Proof of Corollary 1

Since $u(A)$ is convex and $v > \underline{u}$, there exists $v' \in u(A)$ and $\varepsilon > 0$ such than $v - 2\varepsilon > v' > \underline{u}$. Taking $\bar{\delta}_1 = \max_{i\in I} \frac{d_i}{d_i + v'_i - \underline{u}_i}$ and applying Proposition 6 yields $v' \in \overline{E_{\text{med}}(\delta)}$ for all $\delta > \bar{\delta}_1$. Thus, for every $\delta > \bar{\delta}_1$, there exists $v(\delta) \in E_{\text{med}}(\delta)$ such that $v - \varepsilon > v(\delta) > \underline{u}$.

Let $\bar{\delta}_2 = \max_{i\in I} \frac{d_i}{d_i + \varepsilon}$ and let $\bar{\delta} = \max\left\{\bar{\delta}_1, \bar{\delta}_2\right\}$. Then, for every $\delta > \bar{\delta}$, the following is an equilibrium: On path, the mediator recommends a correlated action profile that attains v. If anyone deviates, play transitions to a sequential equilibrium yielding payoff $v(\delta)$. Such an equilibrium exists as $v(\delta) \in E_{\text{med}}(\delta)$ for $\delta > \bar{\delta} \geq \bar{\delta}_1$. Finally, since this punishment results in a loss of continuation payoff of at least ε for each player, the mediator's on-path recommendations are incentive compatible for $\delta > \bar{\delta} \geq \bar{\delta}_2$.

References

[1] Abreu, D., P.K. Dutta, and L. Smith (1994), "The Folk Theorem for Repeated Games: a NEU Condition," *Econometrica*, 62, 939-948.

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

[3] Amarante, M. (2003), "Recursive Structure and Equilibria in Games with Private Monitoring," *Economic Theory*, 22, 353-374.

[4] Athey, S. and K. Bagwell (2001), "Optimal Collusion with Private Information," *RAND Journal of Economics*, 32, 428-465.

[5] Awaya, Y. and V. Krishna (2014), "On Tacit versus Explicit Collusion," *mimeo.*

[6] Ben-Porath, E. and M. Kahneman (1996), "Communication in Repeated Games with Private Monitoring," *Journal of Economic Theory*, 70, 281-297.

[7] Bergemann, D. and S. Morris (2013), "Robust Predictions in Games with Incomplete Information," *Econometrica*, 81, 1251-1308.

[8] Cherry, J. and L. Smith (2011), "Unattainable Payoffs for Repeated Games of Private Monitoring," *mimeo.*

[9] Compte, O. (1998), "Communication in Repeated Games with Imperfect Private Monitoring," *Econometrica*, 66, 597-626.

[10] Dhillon, A. and J.-F. Mertens, (1996) "Perfect Correlated Equilibria," *Journal of Economic Theory*, 68, 279-302.

[11] Forges, F. (1986), "An Approach to Communication Equilibria," *Econometrica*, 54, 1375-1385.

[12] Forges, F. (2009), "Correlated Equilibrium and Communication in Games," *Encyclopedia of Complexity and Systems Science*, edited by R. Meyers, Springer: New York.

[13] Forges, F., J.-F. Mertens, and A. Neyman (1986), "A Counterexample to the Folk Theorem with Discounting," *Economics Letters*, 20, 7.

[14] Fuchs, W. (2007), "Contracting with Repeated Moral Hazard and Private Evaluations," *American Economic Review*, 97, 1432-1448.

[15] Fudenberg, D. and D. Levine (1983), "Subgame-Perfect Equilibria of Finite- and Infinite-Horizon Games," *Journal of Economic Theory*, 31, 251-268.

[16] Fudenberg, D., D. Levine, and E. Maskin (1994), "The Folk Theorem with Imperfect Public Information," *Econometrica*, 62, 997-1039.

[17] Gossner, O. (2000), "Comparison of Information Structures," *Games and Economic Behavior*, 30, 44-63.

[18] Green, E.J. and R.H. Porter (1984), "Noncooperative Collusion under Imperfect Price Information," *Econometrica*, 52, 87-100.

[19] Heller, Y., E. Solan, and T. Tomala, (2012) "Communication, Correlation and Cheap-Talk in Games with Public Information," *Games and Economic Behavior*, 74, 222-234.

[20] Kandori, M. (1991), "Cooperation in Finitely Repeated Games with Imperfect Private Information," *mimeo.*

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

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

[23] Kandori, M. and I. Obara (2010), "Towards a Belief-Based Theory of Repeated Games with Private Monitoring: An Application of POMDP," *mimeo.*

[24] Lehrer, E. (1992), "Correlated Equilibria in Two-Player Repeated Games with Nonobservable Actions," *Mathematics of Operations Research*, 17, 175-199.

[25] Lehrer, E., D. Rosenberg, and E. Shmaya (2010), "Signaling and Mediation in Games with Common Interests," *Games and Economic Behavior*, 68, 670-682.

[26] Levin, J. (2003), "Relational Incentive Contracts," *American Economic Review*, 93, 835-857.

[27] MacLeod, B.W. (2003), "Optimal Contracting with Subjective Evaluation," *American Economic Review*, 93, 216-240.

[28] Mailath, G.J., S.A. Matthews, and T. Sekiguchi (2002), "Private Strategies in Finitely Repeated Games with Imperfect Public Monitoring," *Contributions in Theoretical Economics*, 2.

[29] Mailath, G.J. and L. Samuelson (2006), *Repeated Games and Reputations*, Oxford University Press: Oxford.

[30] Mertens, J.-F., S. Sorin, and S. Zamir (1994), "Repeated Games. Part A," *CORE DP 9420.*

[31] Miyahara, Y. and T. Sekiguchi (2013), "Finitely Repeated Games with Monitoring Options," *Journal of Economic Theory*, 148, 1929-1952.

[32] Myerson, R.B. (1986), "Multistage Games with Communication," *Econometrica*, 54, 323-358.

[33] Pai, M.M, A. Roth, and J. Ullman (2014), "An Anti-Folk Theorem for Large Repeated Games with Imperfect Monitoring," *mimeo*.

[34] Phelan, C., and A. Skrzypacz (2012), "Beliefs and Private Monitoring," *Review of Economic Studies*, 79, 1637-1660.

[35] Phelan, C. and E. Stacchetti (2001), "Sequential Equilibria in a Ramsey Tax Model," *Econometrica*, 69, 1491-1518.

[36] Rahman, D. (2012), "But Who Will Monitor the Monitor?," *American Economic Review*, 102, 2767-2797.

[37] Rahman, D. (2014), "The Power of Communication," *American Economic Review*, 104, 3737-51.

[38] Rahman, D. and I. Obara (2010), "Mediated Partnerships," *Econometrica*, 78, 285-308.

[39] Renault, J. and T. Tomala (2004), "Communication Equilibrium Payoffs in Repeated Games with Imperfect Monitoring," *Games and Economic Behavior*, 49, 313-344.

[40] Sekiguchi, T. (2002), "Existence of Nontrivial Equilibria in Repeated Games with Imperfect Private Monitoring," *Games and Economic Behavior*, 40, 299-321.

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

[42] Sugaya, T. and A. Wolitzky (2014), "Perfect Versus Imperfect Monitoring in Repeated Games," *mimeo*.

[43] Tomala, T. (2009), "Perfect Communication Equilibria in Repeated Games with Imperfect Monitoring," *Games and Economic Behavior*, 67, 682-694.

Online Appendix: Proof of Proposition 4

We prove that $\overline{E_{\text{talk}}(\delta, p)} = \overline{E_{\text{med}}(\delta)}$. Since we consider two-player games, whenever we say players i and j, we assume that they are different players: $i \neq j$.

The structure of the proof is as follows: take any strategy of the mediator, $\tilde{\mu}$, that satisfies (2) (perfect monitoring incentive compatibility); and let $\tilde{v}$ be the value when the players follow $\tilde{\mu}$. Since each $\hat{v} \in E_{\text{med}}(\delta)$ has a corresponding $\hat{\mu}$ that satisfies (2), it suffices to show that, for each $\varepsilon > 0$, there exists a sequential equilibrium whose equilibrium payoff v satisfies $\|v - \tilde{v}\| < \varepsilon$ in the following environment:

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

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

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

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

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

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

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

Construction and Properties of μ

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

With mediated perfect monitoring, since μ^{strict} has full support, player i believes that the mediator's state is R_{perturb} with positive probability after any history. Therefore, by (2) and the fact that μ^{strict} is a strict equilibrium, it is always strictly optimal for each player i to follow her recommendation. This means that, for each period t, there exist $\varepsilon_t > 0$ and $T_t < \infty$ such that, for each player i and on-path history h_m^{t+1}, we have

$$\begin{aligned}
& (1-\delta)\mathbb{E}^{\mu}\left[u_i(r_t) \mid h_m^t, r_{i,t}\right] + \delta \mathbb{E}^{\mu}\left[(1-\delta)\sum_{\tau=t+1}^{\infty} \delta^{\tau-t-1} u_i(\mu(h_m^{\tau})) \mid h_m^t, r_{i,t}\right] \\
> \ & \max_{a_i \in A_i} (1-\delta)\mathbb{E}\left[u_i(a_i, r_{-i,t}) \mid h_m^t, r_{i,t}\right] \\
& +(\delta - \delta^{T_t})\left\{(1-\varepsilon_t)\max_{\hat{a}_i} u_i(\hat{a}_i, \alpha_j^{\varepsilon_t}) + \varepsilon_t \max_{a \in A} u_i(a)\right\} + \delta^{T_t}\max_{a\in A} u_i(a). \qquad (26)
\end{aligned}$$

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

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

$$\begin{aligned}
& (1-\delta)\mathbb{E}^{\mu}\left[u_i(r_t) \mid h_m^t, r_{i,t}\right] + \delta \mathbb{E}^{\mu}\left[(1-\delta)\sum_{\tau=t+1}^{\infty} \delta^{\tau-t-1} u_i(\mu(h_m^{\tau})) \mid h_m^t\right] \\
> \ & (1 - \delta^{T_t})\left\{(1-\varepsilon_t)\max_{\hat{a}_i} u_i(\hat{a}_i, \alpha_j^{\varepsilon_t}) + \varepsilon_t \max_{a \in A} u_i(a)\right\} + \delta^{T_t}\max_{a\in A} u_i(a). \qquad (27)
\end{aligned}$$

That is, if a deviation is punished with probability $1-\varepsilon_t$ for T_t periods including the current period, then player i believes that the deviation is strictly unprofitable.[22]

For each t, we fix $\varepsilon_t > 0$ and $T_t < \infty$ with (26) and (27). Without loss, we can take ε_t decreasing: $\varepsilon_t \geq \varepsilon_{t+1}$ for each t.

Construction and Properties of μ^*

In this subsection, we again consider mediated perfect monitoring. We further modify μ and create the following mediator's strategy μ^*: At the beginning of the game, for each i, t, and a^t, the mediator draws $r_{i,t}^{\text{punish}}(a^t)$ according to $\alpha_i^{\varepsilon_t}$. In addition, for each i and t, she draws $\omega_{i,t} \in \{R, P\}$ such that $\omega_{i,t} = R$ (regular) and P (punish) with probability $1 - p_t$ and p_t, respectively, independently across i and t. We will pin down $p_t > 0$ in Lemma 12. Moreover, given $\omega_t = (\omega_{1,t}, \omega_{2,t})$, the mediator chooses $r_t(a^t)$ for each a^t as follows: If $\omega_{1,t} = \omega_{2,t} = R$, then she draws $r_t(a^t)$ according to $\mu(a^t)(r)$. If $\omega_{i,t} = R$ and $\omega_{j,t} = P$, then she draws $r_{i,t}(a^t)$

[22] If the current on-path recommendation schedule $\Pr^{\mu}(r_{j,t} \mid h_m^t, r_{i,t})$ is very close to α_j^*, then (27) may be more restrictive than (26).

from $\Pr^{\mu}(r_i \mid r_{j,t}^{\text{punish}}(a^t))$ while she draws $r_{j,t}(a^t)$ randomly from $\sum_{a_j \in A_j} \frac{a_j}{|A_j|}$.[23] Finally, if $\omega_{1,t} = \omega_{2,t} = P$, then she draws $r_{i,t}(a^t)$ randomly from $\sum_{a_i \in A_i} \frac{a_i}{|A_i|}$ for each i independently. Since μ has full support, μ^* is well defined.

As will be seen, we will take p_t sufficiently small. In addition, recall that $\eta > 0$ (the perturbation of $\tilde{\mu}$ to μ) is arbitrarily. In the next subsection and onward, we construct an equilibrium with perfect monitoring with cheap talk that has the same equilibrium action distribution as μ^*. Since p_t is small and $\eta > 0$ is arbitrary, constructing such an equilibrium suffices to prove Proposition 4.

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

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

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

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

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

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

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

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

We say that player i has unilaterally deviated if there exist $\tau \leq t-1$ and a unique i such that (i) for each $\tau' < \tau$, we have $a_{n,\tau'} = r_{n,\tau'}$ for each $n \in \{1, 2\}$ (no deviation happened until period $\tau - 1$) and (ii) $a_{i,\tau} \neq r_{i,\tau}$ and $a_{j,\tau} = r_{j,\tau}$ (player i deviates in period τ and player j does not deviate).

Note that μ^* is close to μ on the equilibrium path for sufficiently small p_t. Hence, on-path strict incentive compatibility for player i follows from (26). Moreover, the incentive compatibility condition analogous to (27) also holds.

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

Lemma 12 *There exists $\{p_t\}_{t=1}^{\infty}$ with $p_t > 0$ for each t such that it is strictly optimal for each player i to follow her recommendation: For each player i and history*

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

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

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

$$\begin{aligned} & (1-\delta)\mathbb{E}^{\mu^*}\left[u_i(r_{i,t}, a_{j,t}) \mid h_i^t\right] + \delta \mathbb{E}^{\mu^*}\left[(1-\delta)\sum_{\tau=t+1}^{\infty} \delta^{\tau-t-1} u_i(r_{i,\tau}, a_{j,\tau}) \mid h_i^t, a_{i,t} = r_{i,t}\right] \\ > \; & \max_{a_i \in A_i} (1-\delta)\mathbb{E}^{\mu^*}\left[u_i(a_i, a_{j,t}) \mid h_i^t\right] \\ & + (\delta - \delta^{T_t})\left\{ \left(1 - \varepsilon_t - \textstyle\sum_{\tau=t}^{t+T_t-1} p_\tau\right) \max_{\hat{a}_i} u_i(\hat{a}_i, \alpha_j^{\varepsilon_t}) + \left(\varepsilon_t + \textstyle\sum_{\tau=t}^{t+T_t-1} p_\tau\right) \max_{a \in A} u_i(a) \right\} \\ & + \delta^{T_t} \max_{a \in A} u_i(a). \end{aligned} \tag{28}$$

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

$$\begin{aligned} & (1-\delta)\mathbb{E}^{\mu^*}\left[u_i(r_{i,t}, a_{j,t}) \mid h_i^t\right] + \delta \mathbb{E}^{\mu^*}\left[(1-\delta)\sum_{\tau=t+1}^{\infty} \delta^{\tau-t-1} u_i(r_{i,\tau}, a_{j,\tau}) \mid h_i^t, a_{i,t} = r_{i,t}\right] \\ > \; & (1 - \delta^{T_t})\left\{ \left(1 - \varepsilon_t - \textstyle\sum_{\tau=t}^{t+T_t-1} p_\tau\right) \max_{\hat{a}_i} u_i(\hat{a}_i, \alpha_j^{\varepsilon_t}) + \left(\varepsilon_t + \textstyle\sum_{\tau=t}^{t+T_t-1} p_\tau\right) \max_{a \in A} u_i(a) \right\} \\ & + \delta^{T_t} \max_{a \in A} u_i(a). \end{aligned} \tag{29}$$

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

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

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

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

we have that the difference

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

is small for small p_t.

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

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

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

Hence, (26) and (27) imply that, there exists $\bar{p}_t > 0$ such that, if $p_\tau \leq \bar{p}_t$ for each $\tau \geq t$, then the claims of the lemma hold. Hence, if we take $p_t \leq \min_{\tau \leq t} \bar{p}_\tau$, then the claims hold. ■

We fix $\{p_t\}_{t=1}^\infty$ so that Lemma 12 holds. This fully pins down μ^* with mediated perfect monitoring.

Construction with Perfect Monitoring with Cheap Talk

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

Intuitive Argument

As in μ^*, at the beginning of the game, for each i, t, and a^t, the mediator draws $r_{i,t}^{\text{punish}}(a^t)$ according to $\alpha_i^{\varepsilon_t}$. In addition, with $p_t > 0$ pinned down in Lemma 12, she draws $\omega_t \in \{R,P\}^2$ and $r_t(a^t)$ as in μ^* for each t and a^t. She then defines $\bar{\omega}_t(a^t)$ from a^t, $r_t(a^t)$, and ω_t as in μ^*.

Intuitively, the mediator sends all the information about

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

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

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

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

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

$$y^i_{i,t}(a^t) \equiv x^j_{i,t}(a^t) + r_{i,t}(a^t) \pmod{n_i}. \tag{30}$$

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

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

$$u^j_{i,t}(a^t) \equiv \alpha^i_{i,t}(a^t) \times x^j_{i,t}(a^t) + \beta^i_{i,t}(a^t) \pmod{p^t}. \tag{31}$$

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

$$u^j_{i,t}(a^t) \neq \alpha^i_{i,t}(a^t) \times x^j_{i,t}(a^t) + \beta^i_{i,t}(a^t) \pmod{p^t}, \tag{32}$$

then player i interprets that player j has deviated. For sufficiently large N^t, since player j does not know $\alpha^i_{i,t}(a^t)$ and $\beta^i_{i,t}(a^t)$, if player j tells a lie about $x^j_{i,t}(a^t)$, then with a high probability, player j creates a situation where (32) holds.

Since HST considers Nash equilibrium, they let player i minimax player j forever after (32) holds. On the other hand, since we consider sequential equilibrium, as in the proof of

Lemma 2, we will create a coordination mechanism such that, if player j tells a lie, then with high probability player i minimaxes player j for a long time and player i assigns probability zero to the event that player i punishes player j.

To this end, we consider the following coordination: First, if and only if $\bar{\omega}_{i,t}(a^t) = R$, the mediator defines $u^j_{i,t}(a^t)$ as (31). Otherwise, $u^j_{i,t}(a^t)$ is randomly drawn. That is,

$$u^j_{i,t}(a^t) \equiv \begin{cases} \alpha^i_{i,t}(a^t) \times x^j_{i,t}(a^t) + \beta^i_{i,t}(a^t) \pmod{p^t} & \text{if } \bar{\omega}_{i,t}(a^t) = R, \\ \text{uniformly distributed over } \mathbb{Z}_{p^t} & \text{if } \bar{\omega}_{i,t}(a^t) = P. \end{cases} \tag{33}$$

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

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

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

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

that is, $u^j_{i,t}(n, a^t)$ corresponds to $u^j_{i,t}(a^t)$ with (33) only if $n = n^i_{i,t}(a^t)$. For other n, $u^j_{i,t}(n, a^t)$ is completely random.

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

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

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

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

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

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

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

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

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

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

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

without knowing $\alpha^i_{i,\tau}(a^\tau)$ and $\beta^i_{i,\tau}(a^\tau)$ while the mediator's message does not tell her what is $\alpha^i_{i,t}(a^t) \times x^j_{i,t}(a^t) + \beta^i_{i,t}(a^t) \pmod{p^\tau}$ by (33). Hence, for sufficiently large N^τ, player j cannot avoid $\hat{\omega}_{i,\tau}(a^\tau) = P$ with a nonnegligible probability. Hence, player j will be minmaxed from the next period with a high probability.

The above argument in total shows that, if player j deviates, whether in communication or action, then she will be minmaxed for sufficiently long time. Lemma 12 ensures that player j does not want to tell a lie or take a deviant action.

Formal Construction

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

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

so that (30) holds.

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

As in (33), the mediator defines

$$u_{i,t}^{j}(a^t) \equiv \begin{cases} \alpha_{i,t}^{i}(a^t) \times x_{i,t}^{j}(a^t) + \beta_{i,t}^{i}(a^t) \pmod{p^t} & \text{if } \bar{\omega}_{i,t}(a^t) = R, \\ \tilde{u}_{i,t}^{j}(a^t) & \text{if } \bar{\omega}_{i,t}(a^t) = P. \end{cases}$$

In addition, the mediator defines

$$u_{i,t}^{j}(n, a^t) = \begin{cases} u_{i,t}^{j}(a^t) & \text{if } n = n_{i,t}^{i}(a^t), \\ v_{i,t}^{j}(n, a^t) & \text{if otherwise} \end{cases}$$

and

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

as explained above.

Let us now define the equilibrium:

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

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

   to each player i.

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

$$m_{j,t}^{\text{1st}} = \begin{cases} u_{i,t}^{j}(m_{i,t-1}^{\text{2nd}}, a^t), x_{i,t}^{j}(a^t) & \text{if } t \neq 1 \text{ and } m_{i,t-1}^{\text{2nd}} \neq \{\text{babble}\}, \\ u_{i,t}^{j}(n_{i,t}^{j}(a^t), a^t), x_{i,t}^{j}(a^t) & \text{if } t = 1 \text{ or } m_{i,t-1}^{\text{2nd}} = \{\text{babble}\} \end{cases} \quad (34)$$

   and

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

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

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

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

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

   If

$$m_{j,t}^{\text{1st}}(u) \neq \alpha_{i,t}^{i}(a^t) \times m_{j,t}^{\text{1st}}(x) + \beta_{i,t}^{i}(a^t) \pmod{p^t}, \quad (35)$$

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

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

$$r_{i,t} = \begin{cases} y_{i,t}^i(a^t) - m_{j,t}^{\text{1st}}(x) \pmod{n_i} & \text{if } \hat{\omega}_{i,t}(a^t) = R, \\ r_{i,t}^{\text{punish}}(a^t) & \text{if } \hat{\omega}_{i,t}(a^t) = P. \end{cases} \tag{36}$$

Recall that $y_{i,t}^i(a^t) \equiv x_{i,t}^j(a^t) + r_{i,t}(a^t) \pmod{n_i}$ by (30). By (34), therefore, player i takes $r_{i,t}^i(a^t)$ if $\bar{\omega}_{i,t}(a^t) = R$ and $r_{i,t}^{\text{punish}}(a^t)$ if $\bar{\omega}_{i,t}(a^t) = P$ on the equilibrium path, as in μ^*.

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

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

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

Incentive Compatibility

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

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

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

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

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

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

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

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

Hence, (35) (with indices i and j reversed) implies that $\hat{\omega}_{j,t}(a^t) = P$.

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

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

with probability $1-e$. Then, by (35) (with indices i and j reversed), player j will have $\hat{\omega}_{j,t+1}(a^{t+1}) = P$.

Recursively, if $\bar{\omega}_{j,\tau}(a^\tau) = R$ for each $\tau = t,..,t+T_t-1$, then player i will induce $\hat{\omega}_{j,\tau}(a^\tau) = P$ for each $\tau = t,..,t+T_t-1$ with a high probability. Hence, for $\varepsilon_t > 0$ and T_t fixed in (26) and (27), for sufficiently large $\bar{N}^t$, if $N^\tau \geq \bar{N}^t$ for each $\tau \geq t$, then player i will be punished for subsequent T_t periods regardless of player i's continuation strategy with probability no less than $1-\varepsilon_t-\sum_{\tau=t}^{t+T_t-1}p_\tau$. ($\sum_{\tau=t}^{t+T_t-1}p_\tau$ represents the maximum probability of having $\bar{\omega}_{i,\tau}(a^\tau) = P$ for some τ for subsequent T_t periods.) (29) implies that telling a lie gives strictly lower payoff than the equilibrium payoff. Therefore, it is optimal to tell the truth about $m_{i,t}^{\text{1st}}$. (In (29), we have shown interim incentive compatibility after knowing $\bar{\omega}_{i,t}(a^t)$ and $r_{i,t}$ while here, we consider h_i^t before $\bar{\omega}_{i,t}(a^t)$ and $r_{i,t}$. Taking the expectation with respect to $\bar{\omega}_{i,t}(a^t)$ and $r_{i,t}$, (29) ensures incentive compatibility before knowing $\bar{\omega}_{i,t}(a^t)$ and $r_{i,t}$.)

Second, consider player i's incentive to take the action $a_{i,t} = r_{i,t}$ according to (36) if player i follows the equilibrium strategy until she sends $m_{i,t}^{\text{1st}}$. If she follows the equilibrium strategy, then player i believes at the time when she takes an action that her equilibrium value is equal to

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

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

If player i deviates from $a_{i,t}$, then $\bar{\omega}_{j,\tau}(a^\tau) = P$ by definition for each $\tau \geq t+1$ and a^τ that is compatible with a^t (that is, $a^\tau = (a^t, a_t, ..., a_{\tau-1})$ for some $a_t, ..., a_{\tau-1}$). To avoid being minmaxed in period τ, player i needs to induce $\hat{\omega}_{j,\tau}(a^\tau) = R$ although $\bar{\omega}_{j,\tau}(a^\tau) = P$. Given $\bar{\omega}_{j,\tau}(a^\tau) = P$, since $\alpha_{i,t}^i(a^t)$, $\beta_{i,t}^i(a^t)$, $\tilde{u}_{i,t}^j(a^t)$, $v_{i,t}^j(n, a^t)$ $n_{i,t}^i(a^t)$, and $\tilde{n}_{i,t}^j(a^t)$ are uniform and independent (conditional on the other variables), regardless of player i's continuation strategy, by (35) (with indices i and j reversed), player i will send $m_{i,\tau}^{\text{1st}}$ with

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

with a high probability.

Hence, for sufficiently large $\bar{N}^t$, if $N^\tau \geq \bar{N}^t$ for each $\tau \geq t$, then player i will be punished for the next T_t periods regardless of player i's continuation strategy with probability no less than $1 - \varepsilon_t$. By (28), deviating from $r_{i,t}$ gives a strictly lower payoff than her equilibrium payoff. Therefore, it is optimal to take $a_{i,t} = r_{i,t}$.

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

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

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

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

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

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

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