Online Appendix

E Extending the Opening Example

In the extended example, there are four players (in addition to the mediator) and four periods. The roles of players 2 and 4 are similar to those of players 1 and 2 in the original example, respectively. The timing is as follows:

Period 1. No signals are observed. Player 1 takes an action $a_1 \in \{A_1, B_1\}$.

Period 2. The mediator observes a_1. Player 2 takes $a_2 \in \{A_2, B_2, C_2\}$ and player 3 takes $a_3 \in \{A_3, B_3\}$.

Period 3. Player 2 observes $\theta \in \{n, p\}$ such that $\theta = n$ with probability 3/4. The mediator takes $a_0 \in \{A_0, B_0\}$.

Period 4. The mediator and player 4 observe $s \in \{0, 1\}$, where $s = 0$ if $a_1 = A_1$ and either $a_0 = A_0 \wedge a_2 = A_2$ or $a_0 = B_0 \wedge a_2 = B_2$. Player 4 takes $a_4 \in \{N, P\}$.

Player 1's payoff equals $1_{\{a_1=B_1\}} - 1_{\{a_2=C_2 \wedge a_4=P\}}$. Player 2's payoff is given by

$$\begin{array}{ccc} & A_0 & B_0 \\ A_2 & 0 - 1_{\{a_4=P\}} & 1 - 1_{\{a_4=P\}} \\ B_2 & 1 - 1_{\{a_4=P\}} & 0 - 1_{\{a_4=P\}} \\ C_2 & -3 - 1_{\{a_4=P\}} & -3 - 1_{\{a_4=P\}} \\ & \multicolumn{2}{c}{a_1 = A_1} \end{array} \qquad \begin{array}{ccc} & A_0 & B_0 \\ A_2 & 1 - 1_{\{a_4=P\}} & 1 - 1_{\{a_4=P\}} \\ B_2 & 1 - 1_{\{a_4=P\}} & 1 - 1_{\{a_4=P\}} \\ C_2 & 0 & 0 \\ & \multicolumn{2}{c}{a_1 = B_1} \end{array} .$$

Player 3's payoff is constant. Player 4's payoff equals $-1_{\{(a_1,a_2,\theta) \neq (A_1,C_2,p)\}} 1_{\{a_4=P\}}$.

Consider the target outcome distribution where (i) $\frac{1}{2}A_1 + \frac{1}{2}B_1$ is played in period 1, (ii) when A_1 is played in period 1, $\frac{1}{2}(A_2, A_3, A_0) + \frac{1}{2}(B_2, B_3, B_0)$ is played in periods 2 and 3, (iii) when B_1 is played in period 1, (A_2, A_3, A_0) is played in periods 2 and 3, and (iv) N is played in period 4. We claim that this distribution is implementable in SE, but not with $\mathfrak{C} = \mathfrak{C}^*$.

Implementability with $\mathfrak{C} \neq \mathfrak{C}^*$ Again, it suffices to implement the target distribution in a canonical NE in which players avoid codominated actions. Consider the following mediator strategy:

The mediator draws $m_1 \in \{A_1, B_1\}$ with equal probability.

When $m_1 = a_1 = A_1$, the mediator draws $m_0 \in \{A_0, B_0\}$ with equal probability, and recommends $m_2 = A_2 \wedge m_3 = A_3$ if $m_0 = A_0$ and recommends $m_2 = B_2 \wedge m_3 = B_3$ if $m_0 = B_0$. If $s = 0$, he recommends $m_4 = N$; if $s = 1$, he recommends $m_4 = P$.

When $m_1 = A_1$ but $a_1 = B_1$, the mediator recommends $m_2 = C_2$, $m_3 = A_3$, and $m_4 = P$.

When $m_1 = B_1$ (regardless of a_1), the mediator recommends $m_0 = A_0$, $m_2 = A_2$, $m_3 = A_3$, and $m_4 = N$.

It is straightforward to check that this is a NE. Moreover, no codominated actions are recommended: For player 4, N is weakly dominant and hence never codominated, while P is recommended only following $s = 1$. Hence, we need only check that P is not codominated following $s = 1$. But this holds, because the event $(a_1, a_2, \theta) = (A_1, C_2, p)$ is compatible with $s = 1$, and in this event P is optimal.

Lemma 6 follows from another application of Bayes' rule. We again relegate the proof to the online appendix.

Given $\xi(h_i^{R,t}) = \left(t^*, \zeta_i^{t^*-1}\right)$, player i believes that the mediator and players $-i$ do not tremble after period t^*, and that recommendations are independent of θ and ζ after period t^*. Hence, by Lemma 6, (18) is equivalent to (14), and therefore follows from the definition of $\pi_{t^*}^k$. The proof for (19) is analogous.

This completes the proof that $(\sigma, \phi, J, K, \beta)$ is a quasi-SE.

Final Construction Fix any canonical NE (σ^*, π^*) in which codominated actions are never recommended.[45] The proof is completed by mixing the "motivating" quasi-SE $(\sigma, \phi, J, K, \beta)$ with this NE (with almost all weight on the latter) to create a quasi-SE that implements the same outcome.

We construct a sequence of quasi-strategy profiles $\left(\bar{\sigma}^k, \bar{\phi}^k, J, K\right)$ indexed by k that limit to a quasi-SE profile $\left(\bar{\sigma}, \bar{\phi}, J, K\right)$ (with the same sets J and K as in the motivating quasi-SE) satisfying $\rho^{\bar{\sigma},\bar{\phi}} = \rho^{\sigma^*,\pi^*}$.

Players' strategies σ^k*:* Players are faithful, and after receiving $m_{i,t} = \star$, with probability $1 - \sqrt{\varepsilon_k}$ player i takes $a_{i,t}$ according to the PE strategy $\hat{\sigma}_{i,t}(\mathring{h}_i^{R,t})$, and with probability $\sqrt{\varepsilon_k}$ she takes all actions with equal probability.

Mediator's strategy ϕ^k*:* At the beginning of the game, the mediator draws $f \in F^*$ according to π^* with probability $1 - \frac{1}{k}$ (and subsequently follows f), and the mediator follows quasi-strategy ϕ^k with probability $\frac{1}{k}$.

Letting $\left(\bar{\sigma}, \bar{\phi}\right) = \lim_{k\to\infty} \left(\bar{\sigma}^k, \bar{\phi}^k\right)$, we have $\rho^{\bar{\sigma},\bar{\phi}} = \rho^{\sigma^*,\pi^*}$.

Since J includes all faithful histories where no codominated actions have been recommended, $\left(\bar{\sigma}, \bar{\phi}, J, K\right)$ is valid. For each i, t, $h_i^{R,t} \in J_i^{R,t}$, and $h^{R,t}$ with i-component $h_i^{R,t}$, define

$$\bar{\beta}_{i,t}(h^{R,t}|h_i^{R,t}) = \lim_{k\to\infty} \frac{\Pr^{\bar{\sigma}^k,\bar{\phi}^k}\left(h^{R,t}\right)}{\Pr^{\bar{\sigma}^k,\bar{\phi}^k}(h_i^{R,t})}.$$

Define $\bar{\beta}_{i,t}(h^{A,t}|h_i^{A,t})$ analogously. Since $\Pr^{\bar{\sigma}^k,\bar{\phi}^k}(h_i^{R,t}) > 0$ for each $h_i^{R,t} \in J_i^{R,t}$ conditional on the mediator following ϕ^k, $\bar{\beta}$ is well-defined, and hence Kreps-Wilson consistent.

To prove that $\left(\bar{\sigma}, \bar{\phi}, J, K, \bar{\beta}\right)$ is a quasi-SE, it remains to verify sequential rationality. Under belief system $\bar{\beta}$, so long as a player i has been faithful and has not observed a signal or recommendation that occurs with probability 0 conditional on the mediator following π^*, she believes that with probability 1 the mediator is following π^* and other players have been faithful so far. At such a history, it is optimal for player i to be faithful, since (σ^*, π^*) is a NE. On the other hand, if player i has been faithful and does observe a signal or recommendation that occurs with probability 0 conditional on mediator strategy π^*, then she believes with probability 1 that the mediator is following ϕ^k and other players have been faithful. In this case, faithfulness is optimal by (18) and (19).

---

[45] Note that if (σ^{**}, π^*) is a canonical NE for some canonical (but possibly not fully canonical) player strategy profile σ^{**}, then (σ^*, π^*) is also a canonical NE, where σ^* denotes the fully canonical player strategy profile. One way of seeing this is to note that the strategy profile constructed in the proof of Proposition 2 is fully canonical.

Online Appendix

E Extending the Opening Example

In the extended example, there are four players (in addition to the mediator) and four periods. The roles of players 2 and 4 are similar to those of players 1 and 2 in the original example, respectively. The timing is as follows:

Period 1. No signals are observed. Player 1 takes an action $a_1 \in \{A_1, B_1\}$.

Period 2. The mediator observes a_1. Player 2 takes $a_2 \in \{A_2, B_2, C_2\}$ and player 3 takes $a_3 \in \{A_3, B_3\}$.

Period 3. Player 2 observes $\theta \in \{n, p\}$ such that $\theta = n$ with probability 3/4. The mediator takes $a_0 \in \{A_0, B_0\}$.

Period 4. The mediator and player 4 observe $s \in \{0, 1\}$, where $s = 0$ if $a_1 = A_1$ and either $a_0 = A_0 \wedge a_2 = A_2$ or $a_0 = B_0 \wedge a_2 = B_2$. Player 4 takes $a_4 \in \{N, P\}$.

Player 1's payoff equals $1_{\{a_1=B_1\}} - 1_{\{a_2=C_2 \wedge a_4=P\}}$. Player 2's payoff is given by

$$\begin{array}{ccc} & A_0 & B_0 \\ A_2 & 0-1_{\{a_4=P\}} & 1-1_{\{a_4=P\}} \\ B_2 & 1-1_{\{a_4=P\}} & 0-1_{\{a_4=P\}} \\ C_2 & -3-1_{\{a_4=P\}} & -3-1_{\{a_4=P\}} \\ \multicolumn{3}{c}{a_1 = A_1} \end{array} \qquad \begin{array}{ccc} & A_0 & B_0 \\ A_2 & 1-1_{\{a_4=P\}} & 1-1_{\{a_4=P\}} \\ B_2 & 1-1_{\{a_4=P\}} & 1-1_{\{a_4=P\}} \\ C_2 & 0 & 0 \\ \multicolumn{3}{c}{a_1 = B_1} \end{array} .$$

Player 3's payoff is constant. Player 4's payoff equals $-1_{\{(a_1,a_2,\theta)\neq(A_1,C_2,p)\}}1_{\{a_4=P\}}$.

Consider the target outcome distribution where (i) $\frac{1}{2}A_1 + \frac{1}{2}B_1$ is played in period 1, (ii) when A_1 is played in period 1, $\frac{1}{2}(A_2, A_3, A_0) + \frac{1}{2}(B_2, B_3, B_0)$ is played in periods 2 and 3, (iii) when B_1 is played in period 1, (A_2, A_3, A_0) is played in periods 2 and 3, and (iv) N is played in period 4. We claim that this distribution is implementable in SE, but not with $\mathfrak{C} = \mathfrak{C}^*$.

Implementability with $\mathfrak{C} \neq \mathfrak{C}^*$ Again, it suffices to implement the target distribution in a canonical NE in which players avoid codominated actions. Consider the following mediator strategy:

The mediator draws $m_1 \in \{A_1, B_1\}$ with equal probability.

When $m_1 = a_1 = A_1$, the mediator draws $m_0 \in \{A_0, B_0\}$ with equal probability, and recommends $m_2 = A_2 \wedge m_3 = A_3$ if $m_0 = A_0$ and recommends $m_2 = B_2 \wedge m_3 = B_3$ if $m_0 = B_0$. If $s = 0$, he recommends $m_4 = N$; if $s = 1$, he recommends $m_4 = P$.

When $m_1 = A_1$ but $a_1 = B_1$, the mediator recommends $m_2 = C_2$, $m_3 = A_3$, and $m_4 = P$.

When $m_1 = B_1$ (regardless of a_1), the mediator recommends $m_0 = A_0$, $m_2 = A_2$, $m_3 = A_3$, and $m_4 = N$.

It is straightforward to check that this is a NE. Moreover, no codominated actions are recommended: For player 4, N is weakly dominant and hence never codominated, while P is recommended only following $s = 1$. Hence, we need only check that P is not codominated following $s = 1$. But this holds, because the event $(a_1, a_2, \theta) = (A_1, C_2, p)$ is compatible with $s = 1$, and in this event P is optimal.

Given that P is not codominated for player 4 following $s = 1$, no action is codominated for player 2, as each $a_2 \in \{A_2, B_2\}$ can be optimal after $a_1 = A_1$, and $a_2 = C_2$ is optimal after $a_1 = B_1$ when $a_4 = P$ is anticipated. Finally, given that $a_2 = C_2$ and $a_4 = P$ are not codominated after $a_1 = B_1$, action $a_1 = A_1$ is not codominated for player 1.

Non-Implementability with $\mathfrak{C} = \mathfrak{C}^*$ Suppose towards a contradiction that such a SE exists. In what follows, each fraction p/q should be read as $\lim_{k\to\infty} p^k/q^k$, where $p^k, q^k > 0$ denote probabilities along a sequence of strategy profiles converging to the equilibrium.

For each player i and action a_i that is played with positive probability in the target outcome, assume without loss that a_i is played with positive probability after $m_i = a_i$. Moreover, since the on-path actions of players 2 and 3 must be perfectly correlated, it is without loss to assume that, for $i \in \{2, 3\}$, $a_i \in \{A_i, B_i\}$ is played with probability 1 after $m_i = a_i$. Further, to deter a deviation to $a_1 = B_1$ by player 1 following $m_1 = A_1$, player 2 must play $a_2 = C_2$ with probability 1 after some message, which without loss we take to be $m_2 = C_2$. Since player 3 is indifferent among all outcomes, we can also let $a_3 = m_3$ with probability 1. Finally, since player 4 moves last, the usual static revelation principle argument implies that we can let $a_4 = m_4$ with probability 1. We have thus established that, for players $i \in \{2, 3, 4\}$, $a_i = m_i$ with equilibrium probability 1 at every history.

Note that C_2 is strictly dominated conditional on $a_1 = A_1$ and weakly dominated conditional on $a_1 = B_1$. Since player 2 is willing to take C_2 after $m_2 = C_2$, we have $\Pr(a_1 = B_1 | m_2 = C_2) = 1$. Therefore,

$$\begin{aligned}
& \Pr(a_1 = B_1 | m_2 = C_2, a_2 = A_2) \\
= \; & \frac{\Pr(a_1 = B_1)\Pr(m_2 = C_2 | a_1 = B_1)\Pr(a_2 = A_2 | a_1 = B_1, m_2 = C_2)}{\Pr(m_2 = C_2)\Pr(a_2 = A_2 | m_2 = C_2)} \\
= \; & \frac{\Pr(a_1 = B_1)\Pr(m_2 = C_2 | a_1 = B_1)\Pr(a_2 = A_2 | m_2 = C_2)}{\Pr(m_2 = C_2)\Pr(a_2 = A_2 | m_2 = C_2)} \\
= \; & \frac{\Pr(a_1 = B_1)\Pr(m_2 = C_2 | a_1 = B_1)}{\Pr(m_2 = C_2)} \\
= \; & \Pr(a_1 = B_1 | m_2 = C_2) = 1.
\end{aligned}$$

Hence, if player 2 trembles to $a_2 = A_2$ after $m_2 = C_2$, she believes that $a_1 = B_1$ with probability 1, and she therefore chooses her report $\left(\hat{a}_2, \hat{\theta}\right)$ to minimize the probability that $a_1 = B_1$ and $a_4 = P$. Since $a_1 = B_1$ implies $s = 1$ and player 2 can always report as if she took $a_2 = C_2$, this implies that

$$\Pr(a_1 = B_1, a_4 = P | m_2 = C_2, a_2 = A_2) \le \Pr(a_1 = B_1, a_4 = P | m_2 = C_2, a_2 = C_2). \quad (28)$$

Note that if $\Pr(a_1 = B_1, a_4 = P | m_2 = C_2, a_2 = A_2) < 1$ then player 2 would deviate to A_2 after $m_2 = C_2$. So this probability must equal 1, and hence (28) implies

$$\Pr(a_1 = B_1, a_4 = P | m_2 = C_2, a_2 = C_2) = 1.$$

Given that P is not codominated for player 4 following $s = 1$, no action is codominated for player 2, as each $a_2 \in \{A_2, B_2\}$ can be optimal after $a_1 = A_1$, and $a_2 = C_2$ is optimal after $a_1 = B_1$ when $a_4 = P$ is anticipated. Finally, given that $a_2 = C_2$ and $a_4 = P$ are not codominated after $a_1 = B_1$, action $a_1 = A_1$ is not codominated for player 1.

Non-Implementability with $\mathfrak{C} = \mathfrak{C}^*$ Suppose towards a contradiction that such a SE exists. In what follows, each fraction p/q should be read as $\lim_{k\to\infty} p^k/q^k$, where $p^k, q^k > 0$ denote probabilities along a sequence of strategy profiles converging to the equilibrium.

For each player i and action a_i that is played with positive probability in the target outcome, assume without loss that a_i is played with positive probability after $m_i = a_i$. Moreover, since the on-path actions of players 2 and 3 must be perfectly correlated, it is without loss to assume that, for $i \in \{2, 3\}$, $a_i \in \{A_i, B_i\}$ is played with probability 1 after $m_i = a_i$. Further, to deter a deviation to $a_1 = B_1$ by player 1 following $m_1 = A_1$, player 2 must play $a_2 = C_2$ with probability 1 after some message, which without loss we take to be $m_2 = C_2$. Since player 3 is indifferent among all outcomes, we can also let $a_3 = m_3$ with probability 1. Finally, since player 4 moves last, the usual static revelation principle argument implies that we can let $a_4 = m_4$ with probability 1. We have thus established that, for players $i \in \{2, 3, 4\}$, $a_i = m_i$ with equilibrium probability 1 at every history.

Note that C_2 is strictly dominated conditional on $a_1 = A_1$ and weakly dominated conditional on $a_1 = B_1$. Since player 2 is willing to take C_2 after $m_2 = C_2$, we have $\Pr(a_1 = B_1 | m_2 = C_2) = 1$. Therefore,

$$\begin{aligned}
& \Pr(a_1 = B_1 | m_2 = C_2, a_2 = A_2) \\
= & \frac{\Pr(a_1 = B_1)\Pr(m_2 = C_2 | a_1 = B_1)\Pr(a_2 = A_2 | a_1 = B_1, m_2 = C_2)}{\Pr(m_2 = C_2)\Pr(a_2 = A_2 | m_2 = C_2)} \\
= & \frac{\Pr(a_1 = B_1)\Pr(m_2 = C_2 | a_1 = B_1)\Pr(a_2 = A_2 | m_2 = C_2)}{\Pr(m_2 = C_2)\Pr(a_2 = A_2 | m_2 = C_2)} \\
= & \frac{\Pr(a_1 = B_1)\Pr(m_2 = C_2 | a_1 = B_1)}{\Pr(m_2 = C_2)} \\
= & \Pr(a_1 = B_1 | m_2 = C_2) = 1.
\end{aligned}$$

Hence, if player 2 trembles to $a_2 = A_2$ after $m_2 = C_2$, she believes that $a_1 = B_1$ with probability 1, and she therefore chooses her report $\left(\hat{a}_2, \hat{\theta}\right)$ to minimize the probability that $a_1 = B_1$ and $a_4 = P$. Since $a_1 = B_1$ implies $s = 1$ and player 2 can always report as if she took $a_2 = C_2$, this implies that

$$\Pr(a_1 = B_1, a_4 = P | m_2 = C_2, a_2 = A_2) \leq \Pr(a_1 = B_1, a_4 = P | m_2 = C_2, a_2 = C_2). \quad (28)$$

Note that if $\Pr(a_1 = B_1, a_4 = P | m_2 = C_2, a_2 = A_2) < 1$ then player 2 would deviate to A_2 after $m_2 = C_2$. So this probability must equal 1, and hence (28) implies

$$\Pr(a_1 = B_1, a_4 = P | m_2 = C_2, a_2 = C_2) = 1.$$

Since $a_1 = B_1$ implies $s = 1$, we have

$$\Pr(a_1 = B_1, m_4 = P, s = 1 | m_2 = C_2, a_2 = C_2) = 1.$$

Finally, since $a_2 = C_2$ with probability 1 after $m_2 = C_2$, we have

$$\Pr(a_1 = B_1, a_2 = C_2, m_4 = P, s = 1 | m_2 = C_2) = 1. \tag{29}$$

On the other hand, since player 4 is willing to take P after $s = 1$ and $m_4 = P$, we have $\Pr(a_1 = A_1, a_2 = C_2, \theta = p | s = 1, m_4 = P) = 1$. In particular,

$$\frac{\left(\begin{array}{c} \Pr(m_2 = C_2) \Pr(a_1 = A_1, a_2 = C_2, \theta = p, m_4 = P, s = 1 | m_2 = C_2) \\ + \sum_{m_2 \neq C_2} \Pr(m_2) \Pr(a_1 = A_1, a_2 = C_2, \theta = p, m_4 = P, s = 1 | m_2) \end{array} \right)}{\left(\begin{array}{c} \Pr(m_2 = C_2) \sum_{a_1, a_2} \Pr(a_1, a_2, m_4 = P, s = 1 | m_2 = C_2) \\ + \sum_{m_2 \neq C_2} \Pr(m_2) \sum_{a_1, a_2} \Pr(a_1, a_2, m_4 = P, s = 1 | m_2) \end{array} \right)} = 1.$$

Since $(a + c) / (b + d) \leq (a/b) + (c/d)$ for all non-negative numbers a, b, c, d, the left-hand side is no more than

$$\begin{aligned} & \frac{\Pr(a_1 = A_1, a_2 = C_2, \theta = p, m_4 = P, s = 1, | m_2 = C_2)}{\sum_{a_1, a_2} \Pr(a_1, a_2, m_4 = P, s = 1 | m_2 = C_2)} \\ & + \sum_{m_2 \neq C_2} \frac{\Pr(a_1 = A_1, a_2 = C_2, \theta = p, m_4 = P, s = 1 | m_2)}{\sum_{a_1, a_2} \Pr(a_1, a_2, m_4 = P, s = 1 | m_2)}. \end{aligned}$$

Note that by (29),

$$\begin{aligned} & \Pr(a_1 = A_1, a_2 = C_2, \theta = p, m_4 = P, s = 1, | m_2 = C_2) \\ \leq \; & \Pr(a_1 = A_1, a_2 = C_2, m_4 = P, s = 1 | m_2 = C_2) = 0 \end{aligned}$$

and

$$\sum_{a_1, a_2} \Pr(a_1, a_2, m_4 = P, s = 1 | m_2 = C_2) \geq \Pr(a_1 = B_1, a_2 = C_2, m_4 = P, s = 1 | m_2 = C_2) = 1.$$

Hence,

$$\begin{aligned}
1 &\leq \sum_{m_2 \neq C_2} \frac{\Pr\left(a_1 = A_1, a_2 = C_2, \theta = p, m_4 = P, s = 1 | m_2\right)}{\sum_{a_1, a_2} \Pr\left(a_1, a_2, m_4 = P, s = 1 | m_2\right)} \\
&\leq \sum_{m_2 \neq C_2} \frac{\Pr\left(a_1 = A_1, a_2 = C_2, \theta = p, m_4 = P, s = 1 | m_2\right)}{\Pr\left(a_1 = A_1, a_2 = C_2, m_4 = P, s = 1 | m_2\right)} \\
&= \sum_{m_2 \neq C_2} \frac{\Pr\left(a_1 = A_1, a_2 = C_2, \theta = p, m_4 = P | m_2\right)}{\Pr\left(a_1 = A_1, a_2 = C_2, m_4 = P | m_2\right)} \\
&= \sum_{m_2 \neq C_2} \frac{\Pr\left(a_1 = A_1, a_2 = C_2 | m_2\right) \Pr\left(\theta = p, m_4 = P | a_1 = A_1, m_2, a_2 = C_2\right)}{\Pr\left(a_1 = A_1, a_2 = C_2 | m_2\right) \Pr\left(m_4 = P | a_1 = A_1, m_2, a_2 = C_2\right)} \\
&= \sum_{m_2 \neq C_2} \frac{\Pr\left(\theta = p, m_4 = P | a_1 = A_1, m_2, a_2 = C_2\right)}{\Pr\left(m_4 = P | a_1 = A_1, m_2, a_2 = C_2\right)}, \qquad (30)
\end{aligned}$$

where the second line drops the event $(a_1, a_2) \neq (A_1, C_2)$ from the denominator and the third line uses the fact that $a_2 = C_2$ implies $s = 1$.

Now, after $a_2 = C_2$, player 2 is strictly better off when player 4 takes N if $a_1 = A_1$, and player 2 is indifferent between player 4's actions if $a_1 = B_1$. Moreover, $\Pr\left(a_1 = A_1 | m_2\right) > 0$ for each $m_2 \neq C_2$. Hence, for each $m_2 \neq C_2$ and θ, after $(m_2, a_2 = C_2, \theta)$ player 2 chooses her report $\left(\hat{a}_2, \hat{\theta}\right)$ to minimize the conditional probability that $a_4 = P$ given $a_1 = A_1$, and hence to minimize the conditional probability that $m_4 = P$ given $a_1 = A_1$ (since $a_4 = m_4$ with probability 1). Therefore, for each $m_2 \neq C_2$,

$$\begin{aligned}
& \Pr\left(\theta = p, m_4 = P | a_1 = A_1, m_2, a_2 = C_2\right) \\
= \; & \Pr\left(\theta = p | a_1 = A_1, m_2, a_2 = C_2\right) \Pr\left(m_4 = P | a_1 = A_1, m_2, a_2 = C_2, \theta = p\right) \\
= \; & \Pr\left(\theta = p\right) \Pr\left(m_4 = P | a_1 = A_1, m_2, a_2 = C_2, \theta = p\right) \\
= \; & \Pr\left(\theta = p\right) \min_{\left(\hat{a}_2, \hat{\theta}\right)} \Pr\left(m_4 = P | a_1 = A_1, m_2, a_2 = C_2, \theta = p, \hat{a}_2, \hat{\theta}\right) \\
= \; & \Pr\left(\theta = p\right) \min_{\left(\hat{a}_2, \hat{\theta}\right)} \Pr\left(m_4 = P | a_1 = A_1, m_2, a_2 = C_2, \hat{a}_2, \hat{\theta}\right) \\
= \; & \Pr\left(\theta = p\right) \Pr\left(m_4 = P | a_1 = A_1, m_2, a_2 = C_2\right),
\end{aligned}$$

where the fourth equality follows since the distribution of m_4 is independent of θ conditional on $\left(\hat{a}_2, \hat{\theta}\right)$. Thus,

$$\begin{aligned}
& \sum_{m_2 \neq C_2} \frac{\Pr\left(\theta = p, m_4 = P | a_1 = A_1, m_2, a_2 = C_2\right)}{\Pr\left(m_4 = P | a_1 = A_1, m_2, a_2 = C_2\right)} \\
= & \sum_{m_2 \neq C_2} \frac{\Pr\left(\theta = p\right) \Pr\left(m_4 = P | a_1 = A_1, m_2, a_2 = C_2\right)}{\Pr\left(m_4 = P | a_1 = A_1, m_2, a_2 = C_2\right)} \\
= & \Pr\left(\theta = p\right) \sum_{m_2 \neq C_2} 1 = \frac{1}{4} \times 2 = \frac{1}{2}.
\end{aligned}$$

This contradicts (30).

F Proof of Lemma 5

We prove (24); the proof of (25) is analogous. We will prove the following: for each i, t, faithful history $h_i^{R,t}$ with $\delta^k\left(h_i^{R,t}\right) > 0$, ζ_i^t, and $y^t \in Y^t[\mathring{h}_i^{R,t}]$, there exist numbers $\varphi_k^R(h_i^{R,t}, \zeta_i^t) \geq 0$ and $e_k^R\left(h_i^{R,t}, \zeta_i^t, y^t\right) \geq 0$ such that

$$\begin{aligned}
& \delta^k(h_i^{R,t}, \zeta_i^t, y^t | \omega_T \in \Omega_0) = \varphi_k^R(h_i^{R,t}, \zeta_i^t) \left(\Pr^{\tilde{\sigma}^k, \tilde{\mu}}(\lambda(h_i^{R,t}), y^t) + e_k^R\left(h_i^{R,t}, \zeta_i^t, y^t\right)\right), \text{[46]} \\
& \lim_{k \to \infty} \frac{e_k^R\left(h_i^{R,t}, \zeta_i^t, y^t\right)}{\left(\frac{1}{k}\right)^{2(L+1)T}} \leq t.
\end{aligned} \tag{31}$$

(31) is sufficient for (24), since the former implies, for each $\zeta_i^t \in \{0,1\}^{t-1}$ and $y^t \in Y^t[\mathring{h}_i^{R,t}]$,

$$\begin{aligned}
& \lim_{k \to \infty} \delta^k\left(y^t | \zeta_i^t, h_i^{R,t}\right) \\
= & \lim_{k \to \infty} \delta^k\left(y^t | \zeta_i^t, h_i^{R,t}, \omega_T \in \Omega_0\right) \quad \left(\text{by } \xi(h_i^{R,t}) = \left(0, \zeta_i^t\right) \text{ and (17)}\right) \\
= & \lim_{k \to \infty} \frac{\varphi_k^R(h_i^{R,t}, \zeta_i^t)\left(\Pr^{\tilde{\sigma}^k, \tilde{\mu}}(\lambda(h_i^{R,t}), y^t) + e_k^R\left(h_i^{R,t}, \zeta_i^t, y^t\right)\right)}{\sum_{\tilde{y}^t \in Y^t[\mathring{h}_i^{R,t}]} \varphi_k^R(h_i^{R,t}, \zeta_i^t)\left(\Pr^{\tilde{\sigma}^k, \tilde{\mu}}(\lambda(h_i^{R,t}), \tilde{y}^t) + e_k^R\left(h_i^{R,t}, \zeta_i^t, \tilde{y}^t\right)\right)} \quad (\text{by (31)}) \\
= & \lim_{k \to \infty} \frac{\Pr^{\tilde{\sigma}^k, \tilde{\mu}}(\lambda(h_i^{R,t}), y^t) + e_k^R\left(h_i^{R,t}, \zeta_i^t, y^t\right)}{\sum_{\tilde{y}^t \in Y^t[\mathring{h}_i^{R,t}]} \Pr^{\tilde{\sigma}^k, \tilde{\mu}}(\lambda(h_i^{R,t}), \tilde{y}^t) + \sum_{\tilde{y}^t \in Y^t[\mathring{h}_i^{R,t}]} e_k^R\left(h_i^{R,t}, \zeta_i^t, \tilde{y}^t\right)} \\
= & \lim_{k \to \infty} \frac{\Pr^{\tilde{\sigma}^k, \tilde{\mu}}(\lambda(h_i^{R,t}), y^t)}{\sum_{\tilde{y}^t \in Y^t[\mathring{h}_i^{R,t}]} \Pr^{\tilde{\sigma}^k, \tilde{\mu}}(\lambda(h_i^{R,t}), \tilde{y}^t)} \\
= & \tilde{\beta}_i\left(\mathring{h}_{-i}^{R,t} | \lambda(h_i^{R,t})\right),
\end{aligned}$$

---

[46] There is a slight redundancy in this notation: the payoff-relevant part of $\lambda(h_i^{R,t})$ equals y_i^t, since the payoff-relevant component of $\lambda(h_i^{R,t})$ equals $\mathring{h}_i^{R,t}$ and $y^t \in Y^t[\mathring{h}_i^{R,t}]$.

where the second-to-last equality follows as $\Pr^{\tilde{\sigma}^k,\tilde{\mu}}(\lambda(h_i^{R,t}), y^t) \geq \underline{\varepsilon}_1 \underline{\varepsilon}_2 (\varepsilon_k)^{NT} / (T|A|^T)$, $e_k^R \left(h_i^{R,t}, \zeta_i^t, y^t\right) \leq \left(\frac{1}{k}\right)^{2(L+1)T} T$, and $k(\varepsilon_k)^{NT} \to \infty$.

We prove (31) by induction on t. Taking $\varphi_k^R(h_i^{R,1}, \zeta_i^1) = 1$ and $e_k^R \left(h_i^{R,1}, \zeta_i^1, y^1\right) = 0$, (31) holds for $t = 1$.

Suppose it holds for t. We prove it holds for $t+1$. For the rest of the proof, arbitrarily fix $h_i^{R,t+1} \in J_i^{R,t+1}$, $\zeta_i^t \in \{0,1\}^{t-1}$, $y^{t+1} \in Y^{t+1}[\mathring{h}_i^{R,t+1}]$, and $\zeta_{i,t} \in \{0,1\}$. Whenever we write (a_t, s_{t+1}), it means the components in y^{t+1}.

Since θ, ζ, and randomizations under $\left(\tilde{\sigma}^k, \tilde{\mu}\right)$ are independent across players, we have

$$\begin{aligned}
& \delta^k(h_i^{R,t+1}, \zeta_i^t, y^{t+1}, \zeta_{i,t} | \omega_T \in \Omega_0) \\
= \; & \delta^k(h_i^{R,t}, \zeta_i^t, y^t | \omega_T \in \Omega_0) \\
& \times \left(\begin{array}{c} 1_{\left\{\zeta_{i,t}=0, m_{i,t}=a_{i,t} \in A_{i,t} \setminus \hat{B}_{i,t}(\mathring{h}_i^{R,t})\right\}} \left(1 - \sqrt{\varepsilon_k}\right) \left(1 - \left(\frac{1}{k}\right)^{2(L+1)T}\right) \hat{\sigma}_{i,t}(\mathring{h}_i^{R,t})(a_{i,t}) \\ +1_{\left\{\zeta_{i,t}=0, m_{i,t}=\star\right\}} \sqrt{\varepsilon_k} \left(1 - \left(\frac{1}{k}\right)^{2(L+1)T}\right) \left(\left(1 - \sqrt{\varepsilon_k}\right) \hat{\sigma}_{i,t}(\mathring{h}_i^{R,t})(a_{i,t}) + \frac{\sqrt{\varepsilon_k}}{|A_{i,t}|}\right) \\ +1_{\left\{\zeta_{i,t}=1, m_{i,t}=a_{i,t} \in A_{i,t} \setminus \mathfrak{D}_{i,t}(\mathring{h}_i^{R,t})\right\}} \left(\frac{1}{k}\right)^{2(L+1)T} \frac{1}{|A_{i,t}| - \left|\mathfrak{D}_{i,t}(\mathring{h}_i^{R,t})\right|} \end{array} \right) \\
& \times \sum_{\theta_{-i,t}, \zeta_{-i,t}} \left(\begin{array}{l} \prod_{j \neq i : \theta_{j,t}=0, \zeta_{j,t}=0} \left(1 - \sqrt{\varepsilon_k}\right) \left(1 - \left(\frac{1}{k}\right)^{2(L+1)T}\right) \hat{\sigma}_{j,t}(y_j^t)(a_{j,t}) \\ \times \prod_{j \neq i : \theta_{j,t}=1, \zeta_{j,t}=0} \sqrt{\varepsilon_k} \left(1 - \left(\frac{1}{k}\right)^{2(L+1)T}\right) \left(\left(1 - \sqrt{\varepsilon_k}\right) \hat{\sigma}_{j,t}(y_j^t)(a_{j,t}) + \frac{\sqrt{\varepsilon_k}}{|A_{j,t}|}\right) \\ \times \prod_{j \neq i : \zeta_{j,t}=1, a_{j,t} \in A_{j,t} \setminus \mathfrak{D}_{j,t}(y_j^t)} \left(\frac{1}{k}\right)^{2(L+1)T} \frac{1}{|A_{j,t}| - \left|\mathfrak{D}_{j,t}(y_j^t)\right|} \end{array} \right) \\
& \times p(s_{t+1} | y^t, a_t). \qquad (32)
\end{aligned}$$

By the inductive hypothesis, the first line of (32) equals

$$\varphi_k^R(h_i^{R,t}, \zeta_i^t) \left(\Pr^{\tilde{\sigma}^k,\tilde{\mu}}(\lambda(h_i^{R,t}), y^t) + e_k^R \left(h_i^{R,t}, \zeta_i^t, y^t\right)\right).$$

Note that

$$\Pr^{\tilde{\sigma}^k,\tilde{\mu}}(a_{-i,t} | \lambda(h_i^{R,t}), y^t) = \prod_{j \neq i} \left((1 - \varepsilon_k) \hat{\sigma}_{j,t}(y_j^t)(a_{j,t}) + \frac{\varepsilon_k}{|A_{j,t}|}\right).$$

Define

$$\tilde{\varphi}_k\left(\mathring{h}_i^{R,t}, m_{i,t}, a_{i,t}, \zeta_{i,t}\right) = \left(\begin{array}{c} 1_{\left\{\zeta_{i,t}=0, m_{i,t}=a_{i,t} \in A_{i,t} \setminus \hat{B}_{i,t}(\mathring{h}_i^{R,t})\right\}} \left(1 - \sqrt{\varepsilon_k}\right) \left(1 - \left(\frac{1}{k}\right)^{2(L+1)T}\right) \hat{\sigma}_{i,t}(\mathring{h}_i^{R,t})(a_{i,t}) \\ +1_{\left\{\zeta_{i,t}=0, m_{i,t}=\star\right\}} \sqrt{\varepsilon_k} \left(1 - \left(\frac{1}{k}\right)^{2(L+1)T}\right) \left(\left(1 - \sqrt{\varepsilon_k}\right) \hat{\sigma}_{i,t}(\mathring{h}_i^{R,t})(a_{i,t}) + \frac{\sqrt{\varepsilon_k}}{|A_{i,t}|}\right) \\ +1_{\left\{\zeta_{i,t}=1, m_{i,t}=a_{i,t} \in A_{i,t} \setminus \mathfrak{D}_{i,t}(\mathring{h}_i^{R,t})\right\}} \left(\frac{1}{k}\right)^{2(L+1)T} \frac{1}{|A_{i,t}| - \left|\mathfrak{D}_{i,t}(\mathring{h}_i^{R,t})\right|} \end{array} \right)$$

and

$$\begin{aligned}
\tilde{e}_k\left(y^{t+1}\right) &= \sum_{\theta_{-i,t},\zeta_{-i,t}} \left(\begin{array}{l} \prod_{j\neq i:\theta_{j,t}=0,\zeta_{j,t}=0}\left(1-\sqrt{\varepsilon_k}\right)\left(1-\left(\frac{1}{k}\right)^{2(L+1)T}\right)\hat{\sigma}_{j,t}(y_j^t)(a_{j,t}) \\ \times\prod_{j\neq i:\theta_{j,t}=1,\zeta_{j,t}=0}\sqrt{\varepsilon_k}\left(1-\left(\frac{1}{k}\right)^{2(L+1)T}\right)\left(\left(1-\sqrt{\varepsilon_k}\right)\hat{\sigma}_{j,t}(y_j^t)(a_{j,t})+\frac{\sqrt{\varepsilon_k}}{|A_{j,t}|}\right) \\ \times\prod_{j\neq i:\zeta_{j,t}=1,a_{j,t}\in A_{j,t}\setminus\mathfrak{D}_{j,t}(y_j^t)}\left(\frac{1}{k}\right)^{2(L+1)T}\frac{1}{|A_{j,t}|-\left|\mathfrak{D}_{j,t}(y_j^t)\right|} \end{array} \right) \\
&\quad -\Pr^{\tilde{\sigma}^k,\tilde{\mu}}(a_{-i,t}|\lambda(h_i^{R,t}),y^t) \\
&= \left(\frac{1}{k}\right)^{2(L+1)T}\prod_{j\neq i}\left(\frac{1_{\left\{a_{j,t}\in A_{j,t}\setminus\mathfrak{D}_{j,t}(y_j^t)\right\}}}{|A_{j,t}|-\left|\mathfrak{D}_{j,t}(y_j^t)\right|}-\left(1-\varepsilon_k\right)\hat{\sigma}_{j,t}(y_j^t)(a_{j,t})-\frac{\varepsilon_k}{|A_{j,t}|}\right).
\end{aligned}$$

Substituting these into (32), we have

$$\begin{aligned}
&\delta^k(h_i^{R,t+1},\zeta_i^t,y^{t+1},\zeta_{i,t}|\omega_T\in\Omega_0) \\
=\;& \varphi_k^R(h_i^{R,t},\zeta_i^t)\left(\Pr^{\tilde{\sigma}^k,\tilde{\mu}}(\lambda(h_i^{R,t}),y^t)+e_k^R\left(h_i^{R,t},\zeta_i^t,y^t\right)\right) \\
&\times\tilde{\varphi}_k\left(\mathring{h}_i^{R,t},m_{i,t},a_{i,t},\zeta_{i,t}\right) \\
&\times\left(\Pr^{\tilde{\sigma}^k,\tilde{\mu}}(a_{-i,t}|\lambda(h_i^{R,t}),y^t)+\tilde{e}_k\left(y^{t+1}\right)\right) \\
&\times p(s_{t+1}|y^t,a_t). \qquad (33)
\end{aligned}$$

Next, define

$$\varphi_k^R(h_i^{R,t+1},\zeta_i^{t+1})=\varphi_k^R(h_i^{R,t},\zeta_i^t)\times\frac{\tilde{\varphi}_k\left(\mathring{h}_i^{R,t},m_{i,t},a_{i,t},\zeta_{i,t}\right)}{\Pr^{\tilde{\sigma}^k,\tilde{\mu}}\left(m_{i,t}^*(h_i^{R,t+1}),a_{i,t}|\mathring{h}_i^{R,t}\right)}.$$

We can write

$$\begin{aligned}
&\delta^k(h_i^{R,t+1},\zeta_i^{t+1},y^{t+1}|\omega_T\in\Omega_0) \qquad (34) \\
=\;& \varphi_k^R(h_i^{R,t+1},\zeta_i^{t+1})\left(\begin{array}{c} \Pr^{\tilde{\sigma}^k,\tilde{\mu}}(\lambda(h_i^{R,t}),y^t)\times\Pr^{\tilde{\sigma}^k,\tilde{\mu}}(a_{-i,t}|\lambda(h_i^{R,t}),y^t) \\ \times\Pr^{\tilde{\sigma}^k,\tilde{\mu}}\left(m_{i,t}^*(h_i^{R,t+1}),a_{i,t}|\mathring{h}_i^{R,t}\right)\times p(s_{t+1}|y^t,a_t) \\ +e_k^R\left(h_i^{R,t+1},\zeta_i^t,y^{t+1}\right) \end{array}\right),
\end{aligned}$$

where $e_k^R\left(h_i^{R,t+1},\zeta_i^{t+1},y^{t+1}\right)$ is defined to satisfy this equality given (33): that is,

$$\begin{aligned}
&e_k^R\left(h_i^{R,t+1},\zeta_i^{t+1},y^{t+1}\right) \\
=\;& \left(\begin{array}{c} \tilde{e}_k\left(y^{t+1}\right)\Pr^{\tilde{\sigma}^k,\tilde{\mu}}(\lambda(h_i^{R,t}),y^t) \\ +e_k^R\left(h_i^{R,t},\zeta_i^t,y^t\right)\left(\Pr^{\tilde{\sigma}^k,\tilde{\mu}}(a_{-i,t}|\lambda(h_i^{R,t}),y^t)+\tilde{e}_k\left(y^{t+1}\right)\right) \end{array}\right) \\
&\times\Pr^{\tilde{\sigma}^k,\tilde{\mu}}\left(m_{i,t}^*(h_i^{R,t+1}),a_{i,t}|y^t\right)\times p(s_{t+1}|y^t,a_t).
\end{aligned}$$

Since $h_i^{R,t}$ is faithful, the distribution of player i's message and action in period t is fully determined by her own payoff-relevant history $\mathring{h}_i^{R,t}$. Hence,

$$\Pr{}^{\tilde{\sigma}^k,\tilde{\mu}}\left(m_{i,t}^*(h_i^{R,t+1}),a_{i,t}|\mathring{h}_i^{R,t}\right)=\Pr{}^{\tilde{\sigma}^k,\tilde{\mu}}(m_{i,t}^*(h_i^{R,t+1}),a_{i,t}|\lambda(h_i^{R,t}),y^t,a_{-i,t}).$$

Given this equality, we have

$$\begin{aligned}&\Pr{}^{\tilde{\sigma}^k,\tilde{\mu}}(\lambda(h_i^{R,t}),y^t)\times\Pr{}^{\tilde{\sigma}^k,\tilde{\mu}}(a_{-i,t}|\lambda(h_i^{R,t}),y^t)\times\Pr{}^{\tilde{\sigma}^k,\tilde{\mu}}\left(m_{i,t}^*(h_i^{R,t+1}),a_{i,t}|\mathring{h}_i^{R,t}\right)\times p(s_{t+1}|y^t,a_t)\\ =\;&\Pr{}^{\tilde{\sigma}^k,\tilde{\mu}}(\lambda(h_i^{R,t+1}),y^{t+1}).\end{aligned}$$

Substituting this into (34), we have

$$\begin{aligned}&\delta^k(h_i^{R,t+1},\zeta_i^{t+1},y^{t+1}|\omega_T\in\Omega_0)\\ =\;&\varphi_k^R(h_i^{R,t+1},\zeta_i^{t+1})\left(\Pr{}^{\tilde{\sigma}^k,\tilde{\mu}}(\lambda(h_i^{R,t+1}),y^{t+1})+e_k^R\left(h_i^{R,t+1},\zeta_i^t,y^{t+1}\right)\right).\end{aligned}$$

Finally, we have

$$\begin{aligned}\lim_{k\to\infty}\frac{e_k^R\left(h_i^{R,t+1},\zeta_i^{t+1},y^{t+1}\right)}{\left(\frac{1}{k}\right)^{2(L+1)T}}&\leq\lim_{k\to\infty}\frac{\tilde{e}_k\left(y^{t+1}\right)}{\left(\frac{1}{k}\right)^{2(L+1)T}}+\frac{e_k^R\left(h_i^{R,t},\zeta_i^t,y^t\right)}{\left(\frac{1}{k}\right)^{2(L+1)T}}\left(1+\tilde{e}_k\left(y^{t+1}\right)\right)\\ &\leq 1+t,\end{aligned}$$

where the last line uses $\lim_{k\to\infty}\frac{\tilde{e}_k\left(y^{t+1}\right)}{\left(\frac{1}{k}\right)^{2(L+1)T}}\leq 1$ (and hence $\tilde{e}_k\left(y^{t+1}\right)\to 0$ by (16)) and the inductive hypothesis that $\lim_{k\to\infty}\frac{e_k^R\left(h_i^{R,t},\zeta_i^t,y^t\right)}{\left(\frac{1}{k}\right)^{2(L+1)T}}\leq t$. Hence, (31) holds for $t+1$, as desired.

G Proof of Lemma 6

We prove (26); the proof of (27) is analogous. Let $y_i^t=\mathring{h}_i^{R,t}$. By definition of $\beta_{i,t}^{\sigma^*,\pi_{t^*}}$, (26) is equivalent to

$$\lim_{k\to\infty}\delta^k\left(f^{\geq t^*},y^t|h_i^{R,t}\right)=\lim_{k\to\infty}\frac{\pi_{t^*}^k(f^{\geq t^*},y^{t^*})\Pr{}^{\sigma^*}\left(y^t|f^{\geq t^*},y^{t^*}\right)}{\sum_{\tilde{y}^t\in Y^t[y_i^t],\tilde{f}^{\geq t^*}}\pi_{t^*}^k\left(\tilde{f}^{\geq t^*},\tilde{y}^{t^*}\right)\Pr{}^{\sigma^*}\left(\tilde{y}^t|\tilde{f}^{\geq t^*},\tilde{y}^{t^*}\right)}.\tag{35}$$

From the definition of δ^k, $\delta^k\left(f^{\geq t^*},y^{R,t}|h_i^{R,t},t^*,\zeta_i^{t^*}\right)$ equals

$$\frac{A\sum_{f_i^{<t^*},\theta_i^{t^*},\left(f_j^{<t^*},\theta_j^{t^*},m_j^{t^*}\right)_{j\neq i}}\pi_{t^*}^k\left(f^{\geq t^*},y^{t^*}\right)\Pr{}^{\sigma^*}\left(y^t|f^{\geq t^*},y^{t^*}\right)B_i\prod_{j\neq i}C_jD}{A\sum_{\tilde{y}^t\in Y^t[y_i^t],\tilde{f}^{\geq t^*}}\sum_{f_i^{<t^*},\theta_i^{t^*},\left(f_j^{<t^*},\theta_j^{t^*},m_j^{t^*}\right)_{j\neq i}}\pi_{t^*}^k\left(\tilde{f}^{\geq t^*},\tilde{y}^{t^*}\right)\Pr{}^{\sigma^*}\left(\tilde{y}^t|\tilde{f}^{\geq t^*},\tilde{y}^{t^*}\right)B_i\prod_{j\neq i}\tilde{C}_j\tilde{D}},\tag{36}$$

where the summation is taken over $f_i^{<t^*} \in \operatorname{supp} \tilde{\mu}_i^{<t^*}$, $\theta_i^{t^*} \in \{0,1\}^{t^*-1}$, and $(f_j^{<t^*}, \theta_j^{t^*}, m_j^{t^*}) \in \operatorname{supp} \tilde{\mu}_j^{<t^*} \times \{0,1\}^{t^*-1} \times \prod_{\tau=1}^{t^*-1} (A_{j,\tau} \cup \{\star\})\ \forall j$, and we define

$$\begin{array}{ll} A = \left(\frac{1}{k}\right)^{(L+1)t^*+2(L+1)T\left|\zeta_i^{t^*}\right|}, & B_i = \frac{1}{\# M_i(y_i^{t^*})}, \\ C_j = \frac{1}{\# M_j(y_j^{t^*})}, & \tilde{C}_j = \frac{1}{\# M_j(\tilde{y}_j^{t^*})}, \\ D = 1_{\left\{m^{t^*} \in M^{t^*}(y^{t^*}, f^{<t^*}, \theta^{t^*}, \zeta_i^{t^*})\right\}}, & \tilde{D} = 1_{\left\{m^{t^*} \in M^{t^*}(\tilde{y}^{t^*}, f^{<t^*}, \theta^{t^*}, \zeta_i^{t^*})\right\}}. \end{array}$$

Note that A and B_i cancel in (36). Moreover, we have

$$\begin{aligned} D &= D_i^{00} \times D_i^{01} \times D_i^{\bullet 1} \times \textstyle\prod_{j \neq i} \left(D_j^0 \times D_j^1\right), \\ \tilde{D} &= D_i^{00} \times D_i^{01} \times D_i^{\bullet 1} \times \textstyle\prod_{j \neq i} \left(\tilde{D}_j^0 \times \tilde{D}_j^1\right), \end{aligned}$$

where

$$\begin{array}{ll} D_i^{00} = 1_{\left\{m_{i,\tau} = f_{i,\tau}^{<t^*}(y_i^\tau)\ \forall \tau \leq t^*-1 \text{ s.t. } \zeta_{i,\tau} = \theta_{i,\tau} = 0\right\}}, & D_i^{01} = 1_{\left\{m_{i,\tau} = \star\ \forall \tau \leq t^*-1 \text{ s.t. } \zeta_{i,\tau} = 0 \text{ and } \theta_{i,\tau} = 1\right\}}, \\ D_i^{\bullet 1} = 1_{\left\{m_{i,\tau} \in A_{i,\tau} \setminus \mathfrak{D}_{i,\tau}(y_i^\tau)\ \forall \tau \leq t^*-1 \text{ s.t. } \zeta_{i,\tau} = 1\right\}}, & D_j^0 = 1_{\left\{m_{j,\tau} = f_{j,\tau}^{<t^*}(y_j^\tau)\ \forall \tau \leq t^*-1 \text{ s.t. } \theta_{j,\tau} = 0\right\}}, \\ D_j^1 = 1_{\{m_{j,\tau} = \star\ \forall \tau \leq t^*-1 \text{ s.t. } \theta_{j,\tau} = 1\}}, & \tilde{D}_j^0 = 1_{\left\{m_{j,\tau} = \tilde{f}_{j,\tau}^{<t^*}(\tilde{y}_j^\tau)\ \forall \tau \leq t^*-1 \text{ s.t. } \theta_{j,\tau} = 0\right\}}, \\ \tilde{D}_j^1 = 1_{\{m_{j,\tau} = \star\ \forall \tau \leq t^*-1 \text{ s.t. } \theta_{j,\tau} = 1\}}. & \end{array}$$

(The pneumonic here is that the first superscript of D_i indicates the value of $\theta_i \in \{0,1\}$, where $\bullet$ indicates that θ_i is not specified, and the second superscript indicates the value of $\zeta_i \in \{0,1\}$. For player $j \neq i$, the superscript of D_j indicates the value of $\theta_j \in \{0,1\}$.)

Having cancelled A and B_i, since (i) the term $D_i^{00} D_i^{01} D_i^{\bullet 1}$ does not depend on $\left(f_j^{<t^*}, \theta_j^{t^*}, m_j^{t^*}\right)_{j \neq i}$, (ii) $D_i^{00} D_i^{01} D_i^{\bullet 1}$ is the only term that depends on $(f_i^{<t^*}, \theta_i^{t^*})$ in the numerator of (36), and (iii) $\prod_{j \neq i} \left(C_j D_j^0 D_j^1\right)$ is the only term that depends on $\left(f_j^{<t^*}, \theta_j^{t^*}, m_j^{t^*}\right)_{j \neq i}$, the numerator of (36) equals

$$\left(\sum_{f_i^{<t^*}, \theta_i^{t^*}} D_i^{00} D_i^{01} D_i^{\bullet 1} \right) \times \left(\pi_{t^*}^k \left(f^{\geq t^*}, y^{t^*}\right) \Pr^{\sigma^*}\left(y^t | f^{\geq t^*}, y^{t^*}\right) \sum_{\left(f_j^{<t^*}, \theta_j^{t^*}, m_j^{t^*}\right)_{j \neq i}} \textstyle\prod_{j \neq i} \left(C_j D_j^0 D_j^1\right) \right).$$

Since $D_i^{00} D_i^{01} D_i^{\bullet 1}$ also does not depend on $f^{\geq t}$ and y_{-i}^t, the denominator of (36) equals

$$\left(\sum_{f_i^{<t^*}, \theta_i^{t^*}} D_i^{00} D_i^{01} D_i^{\bullet 1} \right) \times \left(\sum_{\tilde{y}^t \in Y^t[y_i^t], \tilde{f}^{\geq t^*}} \pi_{t^*}^k \left(\tilde{f}^{\geq t}, \tilde{y}^{t^*}\right) \Pr^{\sigma^*}\left(\tilde{y}^t | \tilde{f}^{\geq t^*}, \tilde{y}^{t^*}\right) \sum_{\left(f_j^{<t^*}, \theta_j^{t^*}, m_j^{t^*}\right)_{j \neq i}} \textstyle\prod_{j \neq i} \left(C_j D_j^0 D_j^1\right) \right).$$

Moreover, $\sum_{f_j^{<t^*},\theta_j^{t^*},m_j^{t^*}} C_j D_j^0 D_j^1 = 1$ by (15). Hence, (36) equals

$$\frac{\pi^k_{t^*}\left(f^{\geq t^*}, y^{t^*}\right) \Pr^{\sigma^*}\left(y^t | f^{\geq t^*}, y^{t^*}\right)}{\sum_{\tilde{y}^t \in Y^t[y_i^t], \tilde{f}^{\geq t^*}} \pi^k_{t^*}\left(\tilde{f}^{\geq t^*}, \tilde{y}^{t^*}\right) \Pr^{\sigma^*}\left(\tilde{y}^t | \tilde{f}^{\geq t^*}, \tilde{y}^{t^*}\right)}.$$

Taking the limit $k \to \infty$, we obtain (35).

H Proof of Claims 1–4 of Proposition 4

We postpone the proof of Claim 5 to Appendix J, since it relies on results proved in Appendix I.

H.1 Proof of Claim 1

We first show that, if an outcome can be implemented in a NE in which no player can detect another's unilateral deviation, it can also be implemented in a canonical NE in which no player can detect another's unilateral deviation.

Lemma 7 *For any game $G = (\Gamma, \mathfrak{C})$ and NE (σ, ϕ) with $\operatorname{supp} \rho_i^{\sigma,\phi} = \bigcup_{j \neq i,0} \bigcup_{\sigma'_j \in \Sigma_j} \operatorname{supp} \rho_i^{\sigma'_j,\sigma_{-j},\phi}$ for all $i \neq 0$, there exists a canonical NE $\left(\tilde{\sigma}, \tilde{\phi}\right)$ in game $G^* = (\Gamma, \mathfrak{C}^*)$ such that $\rho^{\sigma,\phi} = \rho^{\tilde{\sigma},\tilde{\phi}}$ and $\operatorname{supp} \rho_i^{\tilde{\sigma},\tilde{\phi}} = \bigcup_{j \neq i,0} \bigcup_{\tilde{\sigma}'_j \in \Sigma_j^*} \operatorname{supp} \rho_i^{\tilde{\sigma}'_j,\tilde{\sigma}_{-j},\tilde{\phi}}$ for all $i \neq 0$.*

Proof. Fix such a G and (σ, ϕ). Let $\left(\tilde{\sigma}, \tilde{\phi}\right)$ be the profile in G^* constructed in the proof of Proposition 2. Recall that $\rho^{\tilde{\sigma},\tilde{\phi}} = \rho^{\sigma,\phi}$. By Lemma 1, for each $i \neq 0$, $j \neq i$, and $\tilde{\sigma}'_j \in \Sigma_j^*$, there exists a strategy $\sigma'_j \in \Sigma_j$ such that $\rho_i^{\sigma'_j,\sigma_{-j},\phi} = \rho_i^{\tilde{\sigma}'_j,\tilde{\sigma}_{-j},\tilde{\phi}}$. Hence, we have

$$\operatorname{supp} \rho_i^{\tilde{\sigma},\tilde{\phi}} = \operatorname{supp} \rho_i^{\sigma,\phi} = \bigcup_{j \neq i,0} \bigcup_{\sigma'_j \in \Sigma_j} \operatorname{supp} \rho_i^{\sigma'_j,\sigma_{-j},\phi} \supset \bigcup_{j \neq i,0} \bigcup_{\tilde{\sigma}'_j \in \Sigma_j^*} \operatorname{supp} \rho_i^{\tilde{\sigma}'_j,\tilde{\sigma}_{-j},\tilde{\phi}} \supset \operatorname{supp} \rho_i^{\tilde{\sigma},\tilde{\phi}}.$$

■

Thus, let (σ, ϕ) be a canonical NE in game $G^* = (\Gamma, \mathfrak{C}^*)$ such that $\operatorname{supp} \rho_i^{\sigma,\phi} = \bigcup_{j \neq i,0} \bigcup_{\sigma'_j \in \Sigma_j^*} \operatorname{supp} \rho_i^{\sigma'_j,\sigma_{-j},\phi}$ for all $i \neq 0$. We first construct a (possibly non-canonical) SE in G^* with outcome $\rho^{\sigma,\phi}$. Then we construct a canonical SE with the same outcome.

Non-canonical SE construction: Denote the set of on-path histories for player i by $\hat{H}_i = \{h_i \in H_i : \Pr^{\sigma,\phi}(h_i) > 0\}$. Since (σ, ϕ) is canonical, $h_i \in \hat{H}_i$ if and only if $\mathring{h}_i \in \operatorname{supp} \rho_i^{\sigma,\phi}$ and $r_{i,t} = (a_{i,t-1}, s_{i,t})$ and $m_{i,t} = a_{i,t}$ for all t.

For each k, let σ_i^k denote the perturbation of σ_i where player i trembles uniformly with probability $|R_{i,t}|/k$ at each reporting history $h_i^{R,t} \in H_i^{R,t}$, and trembles uniformly with probability $|A_{i,t}|/k$ at each acting history $h_i^{A,t} \in H_i^{A,t}$.

For each k, let $\left(\Gamma^k, \mathfrak{C}^*\right)$ denote the constrained game where the mediator follows strategy ϕ while each player i is required to play $\sigma_{i,t}^{R,k}\left(h_i^{R,t}\right)$ at each on-path history $h_i^{R,t} \in \hat{H}_i^{R,t}$, is required to play $\sigma_{i,t}^{A,k}\left(h_i^{A,t}\right)$ at each on-path history $h_i^{A,t} \in \hat{H}_i^{A,t}$, is required to send each report with probability no less than $1/k$ at each off-path history $h_i^{R,t} \in H_i^{R,t} \setminus \hat{H}_i^{R,t}$, and is required to take each action with probability no less than $1/k$ at each off-path history $h_i^{A,t} \in H_i^{A,t} \setminus \hat{H}_i^{A,t}$. By standard arguments, this game admits a NE $\left(\bar{\sigma}^k, \phi\right)$. Taking a convergent subsequence if necessary, let $(\bar{\sigma}, \phi) = \lim_{k\to\infty}\left(\bar{\sigma}^k, \phi\right)$. Clearly, $\rho^{\bar{\sigma},\phi} = \rho^{\sigma,\phi}$.

Let $M_{i,t}\left(h_i^{R,t}, r_{i,t}\right)$ denote the set of messages $m_{i,t}$ that player i receives with positive probability at history $\left(h_i^{R,t}, r_{i,t}\right)$ under profile $\left(\bar{\sigma}^k, \phi\right)$. Since $\bar{\sigma}^k$ has full support, this set depends only on player i's reports and messages $\left(r_i^{t+1}, m_i^t\right)$ at history $\left(h_i^{R,t}, r_{i,t}\right)$. Therefore, we can define a mediation range Q by $Q_{i,t}\left(r_i^{t+1}, m_i^t\right) = M_{i,t}\left(h_i^{R,t}, r_{i,t}\right)$ for all i, t, r_i^{t+1}, m_i^t, and $h_i^{R,t}$ such that $\left(r_i^{t+1}, m_i^t\right)$ equals i's reports and messages at $\left(h_i^{R,t}, r_{i,t}\right)$.

For each k, let ϕ^k denote the perturbation of ϕ where the mediator trembles uniformly with probability $\left|Q_{i,t}\left(r_i^{t+1}, m_i^t\right)\right| / k^{N|Z|}$ over messages $m_{i,t} \in Q_{i,t}\left(r_i^{t+1}, m_i^t\right)$ at each $\left(r_i^{t+1}, m_i^t\right)$, independently across i and t. Define a belief system β as $\lim_{k\to\infty} \Pr^{\bar{\sigma}^k,\phi^k}$. By construction of the relative tremble probabilities for players and the mediator, for each i, t, and $h^{R,t} \in H^{R,t}|_Q$, we have

$$\beta_{i,t}(h^{R,t}|h_i^{R,t}) = \lim_{k\to\infty} \Pr^{\bar{\sigma}^k,\phi^k}(h^{R,t}|h_i^{R,t}) = \lim_{k\to\infty} \Pr^{\bar{\sigma}^k,\phi}(h^{R,t}|h_i^{R,t}),$$

and similarly for $\beta_{i,t}(h^{A,t}|h_i^{A,t})$.

Let J be the set of histories compatible with the mediation range Q, and let K be the set of the mediator's history compatible with the mediation range Q.

We show that $(\bar{\sigma}, \phi, J, K, \beta)$ is a quasi-SE in G^*. The two conditions for validity hold since the messages are within the mediation range as long as the mediator follows ϕ. Consistency of β is by construction. Sequential rationality at off-path histories $h_i \in J_i \setminus \hat{H}_i$ follows from a standard upper hemi-continuity argument. To verify sequential rationality at on-path histories, fix $h_i^{R,t} \in \hat{H}_i^{R,t}$, note that

$$\beta_{i,t}(h^{R,t}|h_i^{R,t}) = \lim_{k\to\infty} \Pr^{\bar{\sigma}^k,\phi}(h^{R,t}|h_i^{R,t}) = \Pr^{\bar{\sigma},\phi}\left(h^{R,t}|h_i^{R,t}\right) = \Pr^{\sigma,\phi}\left(h^{R,t}|h_i^{R,t}\right), \tag{37}$$

where the second equality follows because $\Pr^{\bar{\sigma},\phi}\left(h_i^{R,t}\right) = \Pr^{\sigma,\phi}\left(h_i^{R,t}\right) > 0$. Let $\hat{H} = \{h \in J : \Pr^{\sigma,\phi}(h) > 0\}$. (Note that $\hat{H}$ is not necessarily equal to $\prod_i \hat{H}_i$.) Since $\bigcup_{\sigma_i' \in \Sigma_i} \operatorname{supp} \rho_j^{\sigma_i',\sigma_{-i},\phi} = \operatorname{supp} \rho_j^{\sigma,\phi}$ for each $j \neq i$, we have

$$\Pr^{\sigma_i',\sigma_{-i},\phi}\left(\mathring{h}_j^{T+1} \in \operatorname{supp} \rho_j^{\sigma,\phi}\ \forall j \neq i | h^{R,t}\right) = 1$$

for all $h^{R,t} \in \hat{H}^{R,t}$ and $\sigma_i' \in \Sigma_i^*$.[47] Since (σ, ϕ) is canonical, with probability 1 conditional on $h^{R,t} \in \hat{H}^{R,t}$, $a_{j,\tau} = m_{j,\tau}$ and $r_{j,\tau} = (a_{j,\tau-1}, s_{j,\tau})$ for each $\tau \geq t$. Hence,

$$\Pr^{\sigma_i',\sigma_{-i},\phi}\left(h_j^{T+1} \in \hat{H}_j\ \forall j \neq i | h^{R,t}\right) = 1.$$

---

[47] Note that $\Pr^{\sigma_i',\sigma_{-i},\phi}$ is well-defined since $(\bar{\sigma}, \phi, J, K)$ is valid.

Finally, since σ_{-i} and $\bar{\sigma}_{-i}$ coincide at all on-path histories, we have

$$\Pr^{\sigma_i',\bar{\sigma}_{-i},\phi}\left(h_j^{T+1} \in \hat{H}_j \ \forall j \neq i | h^{R,t}\right) = 1. \tag{38}$$

By (37), player i's belief over $h^{R,t} \in J^{R,t}$ at $h_i^{R,t}$ under β is the same as the conditional probability distribution over $h^{R,t} \in J^{R,t}$ at $h_i^{R,t}$ under (σ, ϕ); and for each $h^{R,t}$, by (38), the conditional probability distribution over Z induced by $(\sigma_i', \bar{\sigma}_{-i}, \phi)$ is the same as that induced by $(\sigma_i', \sigma_{-i}, \phi)$. Hence, since (σ, ϕ) is a NE, σ_i is sequentially rational at $h_i^{R,t}$. The argument for acting histories $h_i^{A,t} \in \hat{H}_i^{A,t}$ is analogous.

Canonical SE construction: First, construct a canonical strategy profile $(\tilde{\sigma}, \tilde{\phi})$ from $(\bar{\sigma}, \phi)$ as in the proof of Proposition 2. As in that proof, we let r and m denote the mediator's fictitious reports and messages, and let $\tilde{r}$ and $\tilde{m}$ denote actual reports and messages. Here we also let $\tilde{h}^{R,t} \in \tilde{H}^{R,t}$ and $\tilde{h}^{A,t} \in \tilde{H}^{A,t}$ denote the reporting and acting histories without fictitious reports or messages. Thus, $\left(\tilde{\sigma}, \tilde{\phi}\right)$ is a canonical NE satisfying $\rho^{\tilde{\sigma},\tilde{\phi}} = \rho^{\sigma,\phi}$. To complete the proof, we construct beliefs $\tilde{\beta}$, subsets of histories $\tilde{J}$ and $\tilde{K}$, and a mediation range $\tilde{Q}$ such that $(\tilde{\sigma}, \tilde{\phi}, \tilde{J}, \tilde{K}, \tilde{\beta})$ is a quasi-SE and, for each i, $\tilde{J}_i$ includes all histories where player i has not lied to the mediator or received a message outside the mediation range.

For each k, let $(\tilde{\sigma}^k, \tilde{\phi}^k)$ denote the perturbation of $(\tilde{\sigma}, \tilde{\phi})$ where (i) players report honestly with probability 1 at all reporting histories $\tilde{h}_i^{R,t}$, (ii) players tremble uniformly over actions with probability $|A_{i,t}|/k$ at all acting histories $\tilde{h}_i^{A,t}$, and (iii) the mediator trembles uniformly with probability $|R_{i,t}|/k$ when she draws fictitious report $r_{i,t}$ at history $(\tilde{r}^{t+1}, r^t, m^t, \tilde{m}^t)$ (but does not tremble when he draws fictitious messages m_t or recommendations $\tilde{m}_t$). By construction, for each s^{T+1}, r^{T+1}, m^{T+1}, and a^{T+1}, we have

$$\Pr^{\tilde{\sigma}^k,\tilde{\phi}^k}\left(s^{T+1}, r^{T+1}, m^{T+1}, a^{T+1}\right) = \Pr^{\bar{\sigma}^k,\phi}\left(s^{T+1}, r^{T+1}, m^{T+1}, a^{T+1}\right). \tag{39}$$

Let $\tilde{M}_{i,t}\left(\tilde{h}_i^{R,t}, \tilde{r}_{i,t}\right)$ denote the set of messages $\tilde{m}_{i,t}$ that player i receives with positive probability at history $\left(\tilde{h}_i^{R,t}, \tilde{r}_{i,t}\right)$ under profile $(\tilde{\sigma}^k, \tilde{\phi}^k)$. Since players $-i$ take all actions with positive probability and the mediator selects each fictitious report with positive probability, this set depends only on player i's reports and messages $\left(\tilde{r}_i^{t+1}, \tilde{m}_i^t\right)$ at $\left(\tilde{h}_i^{R,t}, \tilde{r}_{i,t}\right)$. Therefore, we can define a mediation range $\tilde{Q}$ by $\tilde{Q}_{i,t}\left(\tilde{r}_i^{t+1}, \tilde{m}_i^t\right) = \tilde{M}_{i,t}\left(\tilde{h}_i^{R,t}, \tilde{r}_{i,t}\right)$ for all i, t, $\tilde{r}_i^{t+1}$, $\tilde{m}_i^t$, and $\tilde{h}_i^{R,t}$ such that $\left(\tilde{r}_i^{t+1}, \tilde{m}_i^t\right)$ equals player i's reports and messages at $\left(\tilde{h}_i^{R,t}, \tilde{r}_{i,t}\right)$. By construction, $\Pr^{\tilde{\sigma}^k,\tilde{\phi}^k}\left(\tilde{m}_{i,t} | \tilde{h}_i^{R,t}, \tilde{r}_{i,t}\right) > 0$ if and only if $\tilde{m}_{i,t} \in \tilde{Q}_{i,t}\left(\tilde{r}_i^{t+1}, \tilde{m}_i^t\right)$.

For each i, let $\tilde{J}_i^{A,T+}$ equal the set of histories $\tilde{h}_i^{T+1}$ such that $\Pr^{\tilde{\sigma}^k,\tilde{\phi}^k}\left(\tilde{h}_i^{T+1}\right) > 0$. Let $\tilde{K}^{T+}$ equal the set of mediator histories $\left(\tilde{r}^{T+1}, \tilde{m}^{T+1}\right)$ such that there exists $\hat{\sigma} \in \Sigma$ such that $\Pr^{\hat{\sigma},\tilde{\phi}^k}\left(\tilde{r}^{T+1}, \tilde{m}^{T+1}\right) > 0$. Let the other elements of $\tilde{J}$ and $\tilde{K}$ include all truncations of histories in $\tilde{J}^{A,T+}$ and $\tilde{K}^{T+}$. Since $\tilde{J}_i$ includes all histories where player i has not lied to the mediator or received a message outside the mediation range, $(\tilde{\sigma}, \tilde{\phi}, \tilde{J}, \tilde{K})$ is valid.

Define a belief system $\tilde{\beta}$ on $\left(\tilde{J}, \tilde{K}\right)$ as $\lim_{k\to\infty} \Pr^{\tilde{\sigma}^k, \tilde{\phi}^k}$. This is well defined because $\Pr^{\tilde{\sigma}^k,\tilde{\phi}^k}(\tilde{h}_i^{R,t}) > 0$ and $\Pr^{\tilde{\sigma}^k,\tilde{\phi}^k}(\tilde{h}_i^{A,t}) > 0$ for all i, t, $\tilde{h}_i^{R,t} \in \tilde{J}_i^{R,t}$, and $\tilde{h}_i^{A,t} \in \tilde{J}_i^{A,t}$. By construction, $\tilde{\beta}$ is consistent, and for each $\tilde{h}_i^{R,t} \in \tilde{J}_i^{R,t}$ and $\tilde{h}^{R,t}$ with $\tilde{\beta}_{i,t}\left(\tilde{h}^{R,t}|\tilde{h}_i^{R,t}\right) > 0$, all players have been truthful at $\tilde{h}^{R,t}$.

It remains to show that $(\tilde{\sigma}, \tilde{\phi}, \tilde{J}, \tilde{K}, \tilde{\beta})$ is sequentially rational. We prove this for reporting histories; the argument for acting histories is analogous. Suppose towards a contradiction that there exist $i \neq 0$, t, $\tilde{h}_i^{R,t} \in \tilde{J}_i^{R,t}$, and a strategy $\tilde{\sigma}_i' \in \Sigma_i^*$ such that

$$\begin{aligned} &\sum_{\tilde{h}^{R,t} \in \tilde{H}^{R,t}[\tilde{h}_i^{R,t}]|_{\tilde{J},\tilde{K}}, r^t, m^t} \tilde{\beta}_{i,t}\left(\tilde{h}^{R,t}, r^t, m^t|\tilde{h}_i^{R,t}\right) \bar{u}_i\left(\tilde{\sigma}_i', \tilde{\sigma}_{-i}, \tilde{\phi}|\tilde{h}^{R,t}, r^t, m^t\right) \\ > &\sum_{\tilde{h}^{R,t} \in \tilde{H}^{R,t}[\tilde{h}_i^{R,t}]|_{\tilde{J},\tilde{K}}, r^t, m^t} \tilde{\beta}_{i,t}\left(\tilde{h}^{R,t}, r^t, m^t|\tilde{h}_i^{R,t}\right) \bar{u}_i\left(\tilde{\sigma}, \tilde{\phi}|\tilde{h}^{R,t}, r^t, m^t\right). \end{aligned}$$

Here, $\tilde{\beta}$ naturally extends to the belief about the profile of the history $\tilde{h}^{R,t}$ and fictitious reports and messages (r^t, m^t). This implies that there exists (r_i^t, m_i^t) with $\tilde{\beta}_{i,t}\left(r_i^t, m_i^t|\tilde{h}_i^{R,t}\right) > 0$ such that

$$\begin{aligned} &\sum_{\tilde{h}^{R,t} \in \tilde{H}^{R,t}[\tilde{h}_i^{R,t}]|_{\tilde{J},\tilde{K}}, r^t, m^t} \tilde{\beta}_{i,t}\left(\tilde{h}^{R,t}, r^t, m^t|\tilde{h}_i^{R,t}, r_i^t, m_i^t\right) \bar{u}_i\left(\tilde{\sigma}_i', \tilde{\sigma}_{-i}, \tilde{\phi}|\tilde{h}^{R,t}, r^t, m^t\right) \\ > &\sum_{\tilde{h}^{R,t} \in \tilde{H}^{R,t}[\tilde{h}_i^{R,t}]|_{\tilde{J},\tilde{K}}, r^t, m^t} \tilde{\beta}_{i,t}\left(\tilde{h}^{R,t}, r^t, m^t|\tilde{h}_i^{R,t}, r_i^t, m_i^t\right) \bar{u}_i\left(\tilde{\sigma}, \tilde{\phi}|\tilde{h}^{R,t}, r^t, m^t\right), \end{aligned}$$

where the summation is taken over (r^t, m^t) whose i-component equals (r_i^t, m_i^t).

Since player i believes that nobody has lied to the mediator, by the same construction as in the proof of Lemma 1, there exists $\sigma_i' \in \Sigma_i^*$ such that, for each $\left(\tilde{h}^{R,t}, r^t, m^t\right) = (s^{t+1}, \tilde{r}^t, r^t, m^t, \tilde{m}^t, a^t)$ with $\tilde{\beta}_{i,t}\left(\tilde{h}^{R,t}, r^t, m^t|\tilde{h}_i^{R,t}, r_i^t, m_i^t\right) > 0$, we have

$$\bar{u}_i\left(\tilde{\sigma}_i', \tilde{\sigma}_{-i}, \tilde{\phi}|\tilde{h}^{R,t}, r^t, m^t\right) = \bar{u}_i\left(\sigma_i', \sigma_{-i}, \phi|s^{t+1}, r^t, m^t, a^t\right).$$

Similarly, by construction of $\left(\tilde{\sigma}, \tilde{\phi}\right)$, we have

$$\bar{u}_i\left(\tilde{\sigma}, \tilde{\phi}|\tilde{h}^{R,t}, r^t, m^t\right) = \bar{u}_i\left(\sigma, \phi|s^{t+1}, r^t, m^t, a^t\right).$$

In total, there exist $i \neq 0$, t, $\tilde{h}_i^{R,t} \in \tilde{J}_i^{R,t}$, (r_i^t, m_i^t) with $\tilde{\beta}_{i,t}\left(r_i^t, m_i^t | \tilde{h}_i^{R,t}\right) > 0$, and a strategy $\sigma_i' \in \Sigma_i^*$ such that

$$\begin{aligned}
& \sum_{s^{t+1}, r^t, m^t, a^t} \tilde{\beta}_{i,t}\left(s^{t+1}, r^t, m^t, a^t | \tilde{h}_i^{R,t}, r_i^t, m_i^t\right) \bar{u}_i\left(\sigma_i', \sigma_{-i}, \phi | s^{t+1}, r^t, m^t, a^t\right) \\
> & \sum_{s^{t+1}, r^t, m^t, a^t} \tilde{\beta}_{i,t}\left(s^{t+1}, r^t, m^t, a^t | \tilde{h}_i^{R,t}, r_i^t, m_i^t\right) \bar{u}_i\left(\sigma, \phi | s^{t+1}, r^t, m^t, a^t\right),
\end{aligned}$$

where the summation is take over (s^{t+1}, r^t, m^t, a^t) whose i-component corresponds to the counterpart of $\left(\tilde{h}_i^{R,t}, r_i^t, m_i^t\right)$.

Since $(\bar{\sigma}, \phi, J, K, \beta)$ is a quasi-SE in G^*, to derive a contradiction, it remains to show that, for each $i \neq 0$, t, $\tilde{h}_i^{R,t} = (s_i^{t+1}, \tilde{r}_i^t, \tilde{m}_i^t, a_i^t) \in \tilde{J}_i^{R,t}$, (r_i^t, m_i^t) with $\tilde{\beta}_{i,t}\left(r_i^t, m_i^t | \tilde{h}_i^{R,t}\right) > 0$, and (s^{t+1}, r^t, m^t, a^t) whose i-component equals $\left(s_i^{t+1}, r_i^t, m_i^t, a_i^t\right)$, we have

$$\tilde{\beta}_{i,t}\left(s^{t+1}, r^t, m^t, a^t | s_i^{t+1}, \tilde{r}_i^t, r_i^t, m_i^t, \tilde{m}_i^t, a_i^t\right) = \beta_{i,t}\left(s^{t+1}, r^t, m^t, a^t | s_i^{t+1}, r_i^t, m_i^t, a_i^t\right). \tag{40}$$

By construction, given $\tilde{\beta}_{i,t}\left(r_i^t, m_i^t | \tilde{h}_i^{R,t}\right) > 0$, we have $\left(s_i^{t+1}, r_i^t, m_i^t, a_i^t\right) \in J_i^{R,t}$.

To prove (40), note that, for all k we have

$$\begin{aligned}
& \Pr{}^{\tilde{\sigma}^k, \tilde{\phi}^k}\left(s^{t+1}, r^t, m^t, a^t | s_i^{t+1}, \tilde{r}_i^t, r_i^t, m_i^t, \tilde{m}_i^t, a_i^t\right) \\
= & \Pr{}^{\tilde{\sigma}^k, \tilde{\phi}^k}\left(s^{t+1}, r^t, m^t, a^t | s_i^{t+1}, r_i^t, m_i^t, a_i^t\right) \\
= & \Pr{}^{\bar{\sigma}^k, \phi}\left(s^{t+1}, r^t, m^t, a^t | s_i^{t+1}, r_i^t, m_i^t, a_i^t\right),
\end{aligned}$$

where the first equality follows because (i) given $\tilde{h}_i^{R,t} \in \tilde{J}_i^{R,t}$, we have $\tilde{r}_{i,\tau} = (a_{i,\tau-1}, s_{i,\tau})$ for all $\tau \leq t$ and (ii) the distribution of $\tilde{m}_i^t$ is fully determined by $(\tilde{r}_i^t, r_i^t, m_i^t)$, and the second equality follows from (39). Therefore,

$$\begin{aligned}
& \tilde{\beta}_{i,t}\left(s^{t+1}, r^t, m^t, a^t | s_i^{t+1}, \tilde{r}_i^t, r_i^t, m_i^t, \tilde{m}_i^t, a_i^t\right) \\
= & \lim_{k \to \infty} \Pr{}^{\tilde{\sigma}^k, \tilde{\phi}^k}\left(s^{t+1}, r^t, m^t, a^t | s_i^{t+1}, \tilde{r}_i^t, r_i^t, m_i^t, \tilde{m}_i^t, a_i^t\right) \\
= & \lim_{k \to \infty} \Pr{}^{\bar{\sigma}^k, \phi}\left(s^{t+1}, r^t, m^t, a^t | s_i^{t+1}, r_i^t, m_i^t, a_i^t\right) \\
= & \beta_{i,t}\left(s^{t+1}, r^t, m^t, a^t | s_i^{t+1}, r_i^t, m_i^t, a_i^t\right).
\end{aligned}$$

H.2 Proof of Claims 2, 3, and 4

Claim 2: Note that the condition $\operatorname{supp} \rho_i^\sigma = \bigcup_{j \neq i,0} \bigcup_{\sigma_j' \in \Sigma_j} \operatorname{supp} \rho_i^{\sigma_j', \sigma_{-j}, \phi}$ is vacuous when $N = 1$. Hence, the result follows from Claim 1.

Claim 3: The only difference from the proof of Claim 1 is in the verification that the non-canonical assessment $(\bar{\sigma}, \phi, J, K, \beta)$ in G^* is sequentially rational at on-path histories $h_i^{R,t} \in \hat{H}_i^{R,t}$ and $h_i^{A,t} \in \hat{H}_i^{A,t}$. There, this followed from equations (37) and (38). Here, note that in game G^* player i faces sequential rationality constraints only in periods $t \geq t_i$.

Sequential rationality at on-path histories in period t_i now follows from (37) and the fact that player i's payoff does not depend on actions taken in periods $\tau \geq t_i$ (other than her own action a_{i,t_i}). The latter fact also implies sequential rationality in periods $t > t_i$.

Claim 4: Claim 4 follows from the fact that players' reports are always canonical in the equilibrium that will be constructed in the proof of Proposition 5.

I Results for CPPBE

This section contains our analysis of CPPBE, culminating in the proofs of Propositions 7 and 8.

I.1 Quasi-Strategies and Quasi-CPPBE

As in Appendix D.1, we begin by introducing notions of "quasi-strategy," which is simply a partially defined strategy, and "quasi-equilibrium,"which is a profile of quasi-strategies where incentive constraints are satisfied wherever strategies are defined.

Fix a game $G = (\Gamma, \mathfrak{C})$. A *quasi-strategy* (χ_i, J_i) for each player i is defined exactly as in Appendix D.1.

Intuitively, a quasi-strategy $(\psi, P, F|_P)$ for the mediator consists of a subset of reports P, a set of mediation plans $F|_P$ that specify messages only after reports in P, and a probability distribution ψ over $F|_P$. Formally, a *quasi-strategy* $(\psi, P, F|_P)$ for the mediator consists of

1. A set of reports $P = \prod_{t=1}^{T+1} P^t$ with $P^t \subset R^t$ for each t, such that (i) for every $r^t \in P^t$ there exists $r^{T+1} \in P^{T+1}$ that coincides with r^t up to period t, and (ii) for every $r^{T+1} \in P^{T+1}$ and t, the period-t truncation of r^{T+1}, denoted r^{t+1}, satisfies $r^{t+1} \in P^{t+1}$.

2. A set $F|_P$, where each $f = (f_t)_t \in F|_P$ consists of, for each $t = 1, \ldots, T$, a function $f_t : P^{t+1} \to M_t$.

3. A probability distribution $\psi \in \Delta(F|_P)$.

Let $\mathcal{Z}|_{J,P}$ be the set of $\left(f, h^{T+1}\right)$ such that $h^{T+1} = \left(s^{T+1}, r^{T+1}, m^{T+1}, a^{T+1}\right) \in H^{T+1}$, $h_i^{T+1} \in J_i^{T+1}$ for each i, and $f_{i,t}\left(r^{t+1}\right) = m_{i,t}$ for each i and t. For each i and $h_i^{T+1} \in J_i^{T+1}$, let $\mathcal{Z}[h_i^{T+1}]|_{J,P} = \left\{\left(f, h^{T+1}\right) \in \mathcal{Z}|_{J,P} : h^{T+1} \in H^{T+1}[h_i^{T+1}]\right\}$. Define $\mathcal{Z}^{R,t}|_{J,P}$, $\mathcal{Z}^{R,t+}|_{J,P}$, $\mathcal{Z}^{A,t}|_{J,P}$ and $\mathcal{Z}^{R,t+1}|_{J,P}$ as the projections of $\mathcal{Z}|_{J,P}$ to $F \times H^{R,t}$, $F \times H^{R,t} \times R_t$, $F \times H^{A,t}$, and $F \times H^{A,t} \times A_t$, respectively. Define $\mathcal{Z}^{R,t}[h_i^{R,t}]|_{J,P}$ and $\mathcal{Z}^{A,t}[h_i^{A,t}]|_{J,P}$ analogously.

We say a quasi-strategy profile $(\chi, \psi, J, P, F|_P)$ is *valid* if

1. $J^{R,1} = S_1$. For each $t \geq 1$, $f \in F|_P$, $h^{R,t}$ with $\left(f, h^{R,t}\right) \in \mathcal{Z}^{R,t}|_{J,P}$, $i \neq 0$, σ_i, $\tau \geq t$, and $h^{R,\tau}$ with $\Pr^{\sigma_i, \chi_{-i}, f}\left(h^{R,\tau}|h^{R,t}\right) > 0$, we have $h_j^{R,\tau} \in J_j^{R,\tau}$ for each $j \neq i$ and the report-component r^τ of $h^{R,\tau}$ lies in P^τ.[48] Similarly, for each r_τ with $\Pr^{\sigma_i, \chi_{-i}, f}\left(h^{R,\tau}, r_\tau|h^{R,t}\right) > 0$, we have $(r^\tau, r_\tau) \in P^{\tau+1}$, where r^τ is the report in $h^{R,\tau}$; and for each m_τ with

---

[48] If $h_j^{A,t-1} \notin J_j^{A,t-1}$ for some $j \neq i$ or $r^t \notin P^t$, then $\Pr^{\sigma_i, \chi_{-i}, f}\left(h^{R,\tau}|h^{R,t}\right)$ is not well-defined. In this case, the condition vacuously holds. The same caution applies to the following conditions.

$\Pr^{\sigma_i,\chi_{-i},f}\left(h^{R,\tau}, r_\tau, m_\tau | h^{R,t}\right) > 0$, we have $\left(h_j^{R,\tau}, r_{j,\tau}, m_{j,\tau}\right) \in J_j^{A,\tau}$ for each $j \neq i$. The same condition holds when we replace $h^{R,t}$ with $\left(f, h^{R,t}\right) \in \mathcal{Z}^{R,t}|_{J,P}$ by $h^{A,t}$ with $\left(f, h^{A,t}\right) \in \mathcal{Z}^{A,t}|_{J,P}$. That is, no unilateral player-deviation leads to a history where either the mediator's or another player's quasi-strategy is undefined.

2. For each i and t, if $h_i^{R,t} \in J_i^{R,t}$ then there exist f and $h_{-i}^{R,t} \in J_{-i}^{R,t}$ such that $(f, h_i^{R,t}, h_{-i}^{R,t}) \in \mathcal{Z}^{R,t}|_{J,P}$. Similarly, for each i and t, if $h_i^{A,t} \in J_i^{A,t}$ then there exist f and $h_{-i}^{A,t} \in J_{-i}^{A,t}$ such that $(f, h_i^{A,t}, h_{-i}^{A,t}) \in \mathcal{Z}^{A,t}|_{J,P}$.

The first requirement implies that, for every valid quasi-strategy profile (χ, ψ, J, K), every mediation plan f and terminal history h^{T+1} with $\Pr^{\chi,\psi}\left(f, h^{T+1}\right) > 0$ lies in $\mathcal{Z}|_{J,P}$. The second requirement implies that the projection of $\mathcal{Z}|_{J,P}$ on H_i^{T+1} includes all histories $h_i^{T+1} \in J_i^{T+1}$.

Finally, a *quasi-CPPBE* $\left(\chi, \psi, J, P, F|_P, \bar{\psi}\right)$ is a valid quasi-strategy profile $(\chi, \psi, J, P, F|_P)$ together with a CPS $\bar{\psi}$ on $\mathcal{Z}|_{J,P}$ such that no player has a profitable deviation at any history in $J_i^{R,t}$ and $J_i^{A,t}$: that is, we have

1. *[CPS consistency]* For all f, t, $h^{R,t}$, r_t, m_t, a_t, and s_{t+1} such that $\left(f, h^{R,t}, r_t, m_t, a_t, s_{t+1}\right) \in \mathcal{Z}^{R,t+1}|_{J,P}$, we have

$$\bar{\psi}(f) = \psi(f), \qquad \bar{\psi}\left(r_t | f, h^{R,t}\right) = \textstyle\prod_{i=0}^N \chi_{i,t}^R\left(r_{i,t} | h_i^{R,t}\right),$$
$$\bar{\psi}\left(m_t | f, h^{R,t}, r_t\right) = 1_{\{m_t = f_t(r^t, r_t)\}}, \qquad \bar{\psi}\left(a_t | f, h^{R,t}, r_t, m_t\right) = \textstyle\prod_{i=0}^N \chi_{i,t}^A\left(a_{i,t} | h_i^{R,t}, r_{i,t}, m_{i,t}\right),$$
$$\bar{\psi}\left(s_{t+1} | f, h^{R,t}, r_t, a_t\right) = p\left(s_{t+1} | \mathring{h}^{A,t}, a_t\right),$$

2. *[Sequential rationality of reports]* For all $i \neq 0$, t, $\sigma_i' \in \Sigma_i$, and $h_i^{R,t} \in J_i^{R,t}$, we have

$$\sum_{(f,h^{R,t}) \in \mathcal{Z}^{R,t}[h_i^{R,t}]|_{J,P}} \bar{\psi}\left(f, h^{R,t} | h_i^{R,t}\right) \bar{u}_i\left(\chi, f | h^{R,t}\right) \geq \sum_{(f,h^{R,t}) \in \mathcal{Z}^{R,t}[h_i^{R,t}]|_{J,P}} \bar{\psi}\left(f, h^{R,t} | h_i^{R,t}\right) \bar{u}_i\left(\sigma_i', \chi_{-i}, f | h^{R,t}\right) \tag{41}$$

3. *[Sequential rationality of actions]* For all $i \neq 0$, t, $\sigma_i' \in \Sigma_i$, and $h_i^{A,t} \in J_i^{A,t}$, we have

$$\sum_{(f,h^{A,t}) \in \mathcal{Z}^{A,t}[h_i^{A,t}]|_{J,P}} \bar{\psi}\left(f, h^{A,t} | h_i^{A,t}\right) \bar{u}_i\left(\chi, f | h^{A,t}\right) \geq \sum_{(f,h^{A,t}) \in \mathcal{Z}^{A,t}[h_i^{A,t}]|_{J,P}} \bar{\psi}\left(f, h^{A,t} | h_i^{A,t}\right) \bar{u}_i\left(\sigma_i', \chi_{-i}, f | h^{A,t}\right) \tag{42}$$

Let $\rho^{\chi,\psi} \in \Delta(X)$ denote the outcome distribution induced by valid quasi-strategy profile (χ, ψ). The following lemma says that it is without loss to consider quasi-CPPBE rather than fully specified CPPBE.

Lemma 8 *For any game G and outcome $\rho \in \Delta(X)$, ρ is a CPPBE outcome if and only if $\rho = \rho^{\chi,\psi}$ for some quasi-CPPBE profile $\left(\chi, \psi, J, P, F|_P, \bar{\psi}\right)$ in G. Moreover, given a quasi-CPPBE profile $\left(\chi, \psi, J, P, F|_P, \bar{\psi}\right)$, for each Q such that $J \subset Z|_Q$, there exists a CPPBE $(\sigma, \mu, Q, \bar{\mu})$ such that (σ, μ) and (χ, ψ) coincide on (J, P).*

Proof. Fix a game G. If $(\sigma, \mu, Q, \bar{\mu})$ is a CPPBE, then let J be the set of histories compatible with the mediation range: for each t, define

$$\begin{aligned} J_i^{R,t} &= \left\{ h_i^{R,t} \in H_i^{R,t} : m_{i,\tau} \in Q_{i,\tau}\left(r_i^\tau, m_i^\tau, r_{i,\tau}\right) \quad \forall \tau < t \right\}, \\ J_i^{A,t} &= \left\{ h_i^{A,t} \in H_i^{A,t} : m_{i,\tau} \in Q_{i,\tau}\left(r_i^\tau, m_i^\tau, r_{i,\tau}\right) \quad \forall \tau \le t \right\}, \\ J_i^{R,t+} &= \left\{ \left(h_i^{R,t}, r_{i,t}\right) : h_i^{R,t} \in J_i^{R,t} \text{ and } r_{i,t} \in R_{i,t} \right\}, \text{ and} \\ J_i^{A,t+} &= \left\{ \left(h_i^{A,t}, a_{i,t}\right) : h_i^{A,t} \in J_i^{A,t} \text{ and } a_{i,t} \in A_{i,t} \right\}. \end{aligned}$$

Let $P = R$, and let $F|_P$ be the set of mediation plans compatible with the mediation range:

$$F|_P = \prod_{t=1}^{T} \left\{ f_t : \prod_{\tau=1}^{t} R_\tau \to Q_t\left(r^t, \left(f_\tau\left(r^\tau, r_\tau\right)\right)_{\tau=1}^{t-1}, r_t\right) \right\}.$$

Given this definition, we now show that $(\sigma, \mu, J, P, F|_P, \bar{\mu})$ is a quasi-CPPBE. The two defining conditions for validity holds, since (i) $J^{R,1} = S^1$ by definition, (ii) histories outside J_i cannot arise as long as the mediator follows μ, and (iii) every message history in J can arise for some mediation plan in $F|_P$. CPS consistency and sequential rationality follow from the fact that $(\sigma, \mu, Q, \bar{\mu})$ is a CPPBE.

For the converse, fix a quasi-CPPBE $\left(\chi, \psi, J, P, F|_P, \bar{\psi}\right)$ and a mediation range Q with $J \subset Z|_Q$. We say that a *move distribution on* $\mathcal{Z}|_{J,P}$ is a triple $\left(\alpha^F, \alpha^R, \alpha^A\right)$, where $\alpha^F \in \Delta(F|_P)$, $\alpha^R = \left(\alpha^{R,t}\right)_{t=1}^{T}$ with $\alpha^{R,t} : \mathcal{Z}^{R,t}|_{J,P} \to \Delta(R_t)$, and $\alpha^A = \left(\alpha^{A,t}\right)_{t=1}^{T}$ with $\alpha^{A,t} : \mathcal{Z}^{A,t}|_{J,P} \to \Delta(A_t)$. A move distribution on $\mathcal{Z}|_{J,P}$ has *full support* if we have (i) for each $f \in F|_P$, $\alpha^F(f) > 0$, (ii) for each $\left(f, h^{R,t}\right) \in \mathcal{Z}^{R,t}|_{J,P}$, $\alpha^{R,t}(r_t|f, h^{R,t}) > 0$ if and only if $(h^{R,t}, r_t) \in \mathcal{Z}^{R,t+}|_{J,P}$, and (iii) for each $\left(f, h^{A,t}\right) \in \mathcal{Z}^{A,t}|_{J,P}$, $\alpha^{A,t}(a_t|f, h^{A,t}) > 0$ if and only if $(f, h^{A,t}, a_t) \in \mathcal{Z}^{t+1}|_{J,P}$.

By Theorem 1 of Myerson (1986), every CPS is the limit of conditional probabilities derived from a sequence of full support move distributions. Thus, there exists a sequence of move distributions $\left(\alpha^{F,k}, \alpha^{R,k}, \alpha^{A,k}\right)_k$ with full support on $\mathcal{Z}|_{J,P}$ such that (i) $\alpha^{F,k}(f) \to \psi(f)$ for all $f \in F|_P$, (ii) $\alpha^{R,k}(r_t|f, h^{R,t}) \to \prod_{i=0}^{N} \chi_{i,t}^{R}\left(r_{i,t}|h_i^{R,t}\right)$ for all $\left(f, h^{R,t}, r_t\right) \in \mathcal{Z}^{R,t+}|_{J,P}$, and (iii) $\alpha^{A,k}(a_t|f, h^{A,t}) \to \prod_{i=0}^{N} \chi_{i,t}^{A}(a_{i,t}|h_i^{A,t})$ for all $\left(f, h^{A,t}, a_t\right) \in \mathcal{Z}^{t+1}|_{J,P}$. For each k, let

$$\varepsilon_k = \min_{t,(f,h^{R,t},r_t,m_t,a_t) \in \mathcal{Z}^{t+1}|_{J,P}} \min\{\alpha^{F,k,t}(f), \alpha^{R,k,t}(r_t|f, h^{R,t}), \alpha^{A,k,t}(a_t|f, h^{R,t}, r_t, m_t)\} > 0. \quad (43)$$

Let $\mathfrak{R}_t^k(f, h^{R,t}) = \operatorname{supp} \alpha^{R,k,t}(r_t|f, h^{R,t})$ and $\mathfrak{A}_t^k(f, h^{A,t}) = \operatorname{supp} \alpha^{A,k,t}(a_t|f, h^{R,t}, r_t, m_t)$.

Given $f \in F|_P$, denote the set of mediation plans that coincide with f after history J^t by

$$F(f) = \left\{ \begin{array}{c} f' \in F|_Q : f'_{t-1}(r^t) = f_{t-1}\left(r^t\right) \quad \text{for all } t \text{ and } r^t \\ \text{s.t. there exists } h^{R,t} \in J^{R,t} \text{ with report component equal to } r^t \end{array} \right\}.$$

Note that, given $J \subset Z|_Q$, validity implies that the mediator sends messages compatible with mediation range Q for each r^t such that there exists $h^{R,t} \in J^{R,t}$ with report component equal to r^t. Hence, $F(f)$ is non-empty.

For each k, define an auxiliary game $\left(\Gamma^k, \mathfrak{C}\right)$ as follows:

1. The mediator uses the mixed mediation plan $\mu^k \in \Delta\left(F|_Q\right)$ defined as follows: (i) with probability $1 - \frac{\varepsilon_k}{k}$, draw $f \in F|_P$ according to $\alpha^{F,k} \in \Delta(F|_P)$, and then draw $f' \in F|_Q$ uniformly at random from $F\left(f\right)$; (ii) with probability $\frac{\varepsilon_k}{k}$, draw $f' \in F|_Q$ uniformly at random from $F|_Q$.
2. Each player i chooses probability distributions $\sigma_{i,t}^{R,k}(\cdot|h_i^{R,t}) \in \Delta(R_{i,t})$ and $\sigma_{i,t}^{A,k}(\cdot|h_i^{A,t}) \in \Delta(A_{i,t})$ for each t, $h_i^{R,t} \in H_i^{R,t} \backslash J_i^{R,t}$, and $h_i^{A,t} \in H_i^{A,t} \backslash J_i^{A,t}$. At histories $h_i^{R,t} \in J_i^{R,t}$ and $h_i^{A,t} \in J_i^{A,t}$, player i is required to choose $\sigma_{i,t}^{R,k}(\cdot|h_i^{R,t}) = \chi_{i,t}^{R,t}\left(\cdot|h_i^{R,t}\right)$ and $\sigma_{i,t}^{A,k}(\cdot|h_i^{A,t}) = \chi_{i,t}^{A,t}\left(\cdot|h_i^{A,t}\right)$.
3. Given σ^k and f, the distribution of terminal histories H^{T+1} is determined recursively as follows:

   Given $f \in F|_Q$ and $h^{R,t} \in H^{R,t}$, each $r_t \in R_t$ is drawn with probability

$$\begin{array}{ll} \left(1 - \frac{\varepsilon_k}{k}\left|R_t \backslash \mathfrak{R}_t^k(f, h^{R,t})\right|\right) \alpha^{R,k}(r_t|f, h^{R,t}) & \text{if } \left(f, h^{R,t}\right) \in \mathcal{Z}^{R,t}|_{J,P} \wedge r_t \in \mathfrak{R}_t^k(f, h^{R,t}), \\ \frac{\varepsilon_k}{k} & \text{if } \left(f, h^{R,t}\right) \in \mathcal{Z}^{R,t}|_{J,P} \wedge r_t \notin \mathfrak{R}_t^k(f, h^{R,t}), \\ \prod_{i=0}^N \left(\left(1 - \frac{\varepsilon_k}{k}\left|R_{i,t}\right|\right) \sigma_{i,t}^{R,k}(r_{i,t}|h_i^{R,t}) + \frac{\varepsilon_k}{k}\right) & \text{if } \left(f, h^{R,t}\right) \notin \mathcal{Z}^{R,t}. \end{array}$$

   Given $f \in F|_Q$ and $h^{A,t} \in H^{A,t}$, each $a_t \in A_t$ is drawn with probability

$$\begin{array}{ll} \left(1 - \frac{\varepsilon_k}{k}\left|A_t \backslash \mathfrak{A}_t^k(f, h^{A,t})\right|\right) \alpha^{A,k}(a_t|f, h^{A,t}) & \text{if } \left(f, h^{A,t}\right) \in \mathcal{Z}^{A,t}|_{J,P} \wedge a_t \in \mathfrak{A}_t^k(f, h^{A,t}), \\ \frac{\varepsilon_k}{k} & \text{if } \left(f, h^{A,t}\right) \in \mathcal{Z}^{A,t}|_{J,P} \wedge a_t \notin \mathfrak{A}_t^k(f, h^{A,t}), \\ \prod_{i=0}^N \left(\left(1 - \frac{\varepsilon_k}{k}\left|A_{i,t}\right|\right) \sigma_{i,t}^{A,k}(a_{i,t}|h_i^{A,t}) + \frac{\varepsilon_k}{k}\right) & \text{if } \left(f, h^{A,t}\right) \notin \mathcal{Z}^{A,t}|_{J,P}. \end{array}$$

   Given $h^{A,t} \in H_t^{A,t}$ and $a_t \in A_t$, each $s_{t+1} \in S_{t+1}$ is drawn with probability $p\left(s_{t+1}|\mathring{h}^{A,t}, a_t\right)$.
4. Player i's payoff at terminal history h^{T+1} is $u_i(\mathring{h}^{T+1})$.

As in the proof of Lemma 3, $\left(\Gamma^k, \mathfrak{C}\right)$ admits a NE $(\bar{\sigma}^k, \mu^k)$. Moreover, for any σ^k, $\left(\sigma^k, \mu^k\right)$ has full support on $F|_Q \times Z|_Q$ in $\left(\Gamma^k, \mathfrak{C}\right)$. Hence, $(\bar{\sigma}^k, \mu^k)$ induces a CPS $\bar{\mu}^k$ on $F|_Q \times Z|_Q$ by Bayes' rule.

Let $(\bar{\sigma}^k, \mu^k, \bar{\mu}^k)_k$ denote a sequence of NE $(\bar{\sigma}^k, \mu^k)$ and corresponding CPS's $\bar{\mu}^k$ in $\left(\Gamma^k, \mathfrak{C}\right)$. Taking a convergent subsequence if necessary, let $(\bar{\sigma}, \mu, \bar{\mu}) = \lim_{k\to\infty}(\bar{\sigma}^k, \mu^k, \bar{\mu}^k)$. Note that $(\bar{\sigma}, \mu)$ and (χ, ψ) coincide on (J, P). We claim that $(\bar{\sigma}, \mu, Q, \bar{\mu})$ is a CPPBE in $(\Gamma, \mathfrak{C})$. Since $\bar{\mu}$ is a CPS as the limit of conditional probabilities, it remains to verify sequential rationality. The proof is exactly parallel to the corresponding part of the proof of Lemma 3. We include it for completeness.

We consider reporting histories $h_i^{R,t}$; the argument for acting histories $h_i^{A,t}$ is analogous. There are two cases, depending on whether or not $h_i^{R,t} \in J_i^{R,t}$. If $h_i^{R,t} \notin J_i^{R,t}$, then $h_i^{T+1} \notin J_i^{T+1}$ for all h_i^{T+1} that follow $h_i^{R,t}$, so by inspection the outcome distribution (and hence player

i's expected payoff) conditional on $h^{R,t}$ is continuous in σ^k, μ^k, ε_k, and k. Since $\bar{\sigma}^k_{i,t}\left(\cdot|h_i^{R,t}\right)$ is sequentially rational in $\left(\Gamma^k, \mathfrak{C}\right)$ (as $\left(\bar{\sigma}^k, \mu^k\right)$ is a NE in $\left(\Gamma^k, \mathfrak{C}\right)$ where the distribution over h^{T+1} has full support), it follows that $\sigma^R_{i,t}\left(\cdot|h_i^{R,t}\right)$ is sequentially rational in $(\Gamma, \mathfrak{C})$.

Now consider the case where $h_i^{R,t} \in J_i^{R,t}$. We show that player i believes that $h^{R,t} \in \mathcal{Z}^{R,t}[h_i^{R,t}]|_{J,P}$ with probability 1. Note that, for each $h_i^{T+1} \in J_i^{T+1}$ and $\left(f, h^{T+1}\right) \notin \mathcal{Z}^{R,t}[h_i^{R,t}]|_{J,P}$, there exists $\left(\tilde{f}, \tilde{h}^{T+1}\right) \in \mathcal{Z}^{R,t}[h_i^{R,t}]|_{J,P}$ such that

$$\lim_{k\to\infty} \frac{\bar{\mu}^k(f, h^{T+1})}{\bar{\mu}^k(\tilde{f}, \tilde{h}^{T+1})} = 0.$$

This follows because in $\left(\Gamma^k, \mathfrak{C}\right)$ each "tremble" leading to a history outside J occurs with probability at most ε_k/k, while every history $h_i^{T+1} \in J_i^{T+1}$ occurs with positive probability given move distribution $\left(\alpha^{F,k}, \alpha^{R,k}, \alpha^{A,k}\right)$ (this is an implication of the third condition in the definition of a valid quasi-strategy profile), and with this distribution each move occurs with probability at least ε_k.

Therefore, for each $h_i^{R,t} \in J_i^{R,t}$ and $\left(f, h^{R,t}\right) \in \mathcal{Z}^{R,t}[h_i^{R,t}]|_{J,P}$, we have $\bar{\mu}(f, h^{R,t}|h_i^{R,t}) = \bar{\psi}(f, h^{R,t}|h_i^{R,t})$, and the conditional probability that $\left(f, h^{R,t}\right) \in \mathcal{Z}^{R,t}[h_i^{R,t}]|_{J,P}$ equals 1. Hence, the fact that (41) holds with CPS $\bar{\psi}$ implies that $\sigma^R_{i,t}(\cdot|h_i^{R,t}) = \chi_{i,t}\left(\cdot|h_i^{R,t}\right)$ is sequentially rational in $(\Gamma, \mathfrak{C})$. ■

I.2 SCE Implies CPPBE

Lemma 9 *For any base game Γ, mediation range Q, and SCE $(\mu, Q, \bar{\mu})$, there exists a canonical strategy profile σ and CPS $\bar{\mu}'$ such that $(\sigma, \mu, Q, \bar{\mu}')$ is a CPPBE in $(\Gamma, \mathfrak{C}^*)$ with the same outcome distribution.*

Proof. In the direct-communication game $G^* = (\Gamma, \mathfrak{C}^*)$, let J be the set of truthful histories compatible with the mediation range: for each t, define

$$\begin{aligned}
J_i^{R,t} &= \left\{ \begin{array}{c} h_i^{R,t} \in H_i^{R,t} : \\ r_{i,\tau} = (a_{i,\tau-1}, s_{i,\tau}) \text{ and } m_{i,\tau} \in Q_{i,\tau}\left(r_i^\tau, m_i^\tau, r_{i,\tau}\right) \quad \forall \tau < t \end{array} \right\}, \\
J_i^{A,t} &= \left\{ \begin{array}{c} h_i^{A,t} \in H_i^{A,t} : \\ r_{i,\tau} = (a_{i,\tau-1}, s_{i,\tau}) \text{ and } m_{i,\tau} \in Q_{i,\tau}\left(r_i^\tau, m_i^\tau, r_{i,\tau}\right) \quad \forall \tau \leq t \end{array} \right\}, \\
J_i^{R,t+} &= \left\{ \left(h_i^{R,t}, r_{i,t}\right) : h_i^{R,t} \in J_i^{R,t} \text{ and } r_{i,t} = (a_{i,t-1}, s_{i,t}) \right\}, \text{ and} \\
J_i^{A,t+} &= \left\{ \left(h_i^{A,t}, a_{i,t}\right) : h_i^{A,t} \in J_i^{A,t} \text{ and } a_{i,t} \in A_{i,t} \right\}.
\end{aligned}$$

Let $P = R$, and let $F|_P$ be the set of mediation plans compatible with the mediation range:

$$F|_P = \prod_{t=1}^{T} \left\{ f_t : \prod_{\tau=1}^{t} R_\tau \to Q_t\left(r^t, \left(f_\tau\left(r^\tau, r_\tau\right)\right)_{\tau=1}^{t-1}, r_t\right) \right\}.$$

Consider the quasi-strategy profile $(\chi, \mu, J, P, F|_P)$ where, for each i, χ_i is honest and obedient at each $h_i \in J_i$. This quasi-strategy profile is valid, since (i) histories outside J_i can arise only if player i is dishonest or the mediator uses a mediation plan outside $F|_P$, and (ii) every message history in J can arise for some mediation plan in $F|_P$. Moreover, by inspection, $(\chi, \mu, J, P, F|_P, \bar{\mu})$ is a quasi-CPPBE in G^* if $(\mu, Q, \bar{\mu})$ is a SCE. Hence, the former is a quasi-CPPBE in G^*. Moreover, J_i includes all histories at which player i has been honest and the mediator's messages lie in the mediation range. Hence, Lemma 8 implies that there exists a canonical CPPBE $(\sigma, \mu, Q, \bar{\mu}')$ in G^* with the same mediation range and outcome as $(\mu, Q, \bar{\mu})$. ■

I.3 CPPBE and Codominated Actions

We now show that, in any CPPBE, players do not take codominated actions at any history.

Lemma 10 *For any game G, mediation range Q, and CPPBE $(\sigma, \mu, Q, \bar{\mu})$, $\operatorname{supp} \sigma_{i,t}^{A}(h_i^{A,t}) \cap \mathfrak{D}_{i,t}(\mathring{h}_i^{A,t}) = \emptyset$ for all i, t, and $h_i^{A,t} \in H_i^{A,t}$.*

I.3.1 Proof of Lemma 10

Fix a game G, mediation range Q, CPPBE $(\sigma, \mu, Q, \bar{\mu})$, and sequence of full-support CPS's $\left(\bar{\mu}^k\right)_k$ converging to $\bar{\mu}$. For each i and t, the sequential rationality condition at history $h_i^{A,t}$ is

$$\sum_{(f, h^{A,t}) \in \mathcal{Z}^{A,t}[h_i^{A,t}]|_Q} \bar{\mu}(f, h^{A,t}|h_i^{A,t}) \bar{u}_i\left(\sigma, f|h^{A,t}\right) = \max_{\sigma_i' \in \Sigma_i} \sum_{(f, h^{A,t}) \in \mathcal{Z}^{A,t}[h_i^{A,t}]|_Q} \bar{\mu}(f, h^{A,t}|h_i^{A,t}) \bar{u}_i\left(\sigma_i', \sigma_{-i}, f|h^{A,t}\right). \tag{44}$$

We wish to prove the following lemma, which establishes the corresponding sequential rationality condition in the direct-communication game G^*. Let $\tilde{F}$ be the set of mediation plans in game G^*. For each i, t, and $y_i^t \in Y_i^t$, let $\tilde{M}_{i,t}\left(y_i^t\right) = \bigcup_{h_i^{A,t} : \mathring{h}_i^{A,t} = y_i^t} \operatorname{supp} \sigma_{i,t}^{A}(h_i^{A,t})$; and, for each $\tilde{r}_i^{t+1} \in R_i^{*t+1}$, let

$$\tilde{Q}_{i,t}(\tilde{r}_i^{t+1}) = \begin{cases} \tilde{M}_{i,t}\left(\tilde{r}_i^{t+1}\right) & \text{if } \tilde{r}_i^{t+1} \in Y_i^t \\ A_{i,t} & \text{otherwise} \end{cases}. \tag{45}$$

Given $\tilde{Q}$, let $\tilde{F}|_{\tilde{Q}}$ be the set of mediation plans in game G^* with mediation range $\tilde{Q}$.

Lemma 11 *In game G^*, for each t, there exists a CPS $\tilde{\mu}_t$ on $\tilde{F}|_{\tilde{Q}} \times Y^t$ such that, for each i, $y_i^t \in Y_i^t$, $\tilde{m}_{i,t} \in \tilde{M}_{i,t}\left(y_i^t\right)$, and $\sigma_i' \in \Sigma_i^*$,*

$$\begin{aligned} & \sum_{(\tilde{f}, y^t) \in \tilde{F}|_{\tilde{Q}} \times Y^t[y_i^t]} \tilde{\mu}_t(\tilde{f}, y^t|y_i^t, \tilde{m}_{i,t}) \bar{u}_i\left(\sigma^*, \tilde{f}|\bar{h}\left(\tilde{f}, y^t\right), \tilde{m}_{i,t}\right) \\ \geq & \sum_{(\tilde{f}, y^t) \in \tilde{F}|_{\tilde{Q}} \times Y^t[y_i^t]} \tilde{\mu}_t(\tilde{f}, y^t|y_i^t, \tilde{m}_{i,t}) \bar{u}_i\left(\sigma_i', \sigma_{-i}^*, \tilde{f}|\bar{h}\left(\tilde{f}, y^t\right), \tilde{m}_{i,t}\right). \end{aligned} \tag{46}$$

Before proving Lemma 11, we first show how it implies Lemma 10. Toward a contradiction, suppose that there exists a period t such that $\tilde{M}_{i,t}\left(y_i^t\right) \cap \mathfrak{D}_{i,t}(y_i^t) \neq \emptyset$ for some i and $y_i^t \in Y_i^t$. Let t^* be the last such period. Note that, for all $\tilde{f} \in \tilde{F}|_{\tilde{Q}}$, we have $\tilde{f}_{i,t}\left(\tilde{r}^{t+1}\right) \cap \mathfrak{D}_{i,t}(\tilde{r}_i^{t+1}) = \emptyset$ for all $t > t^*$, i, and $\tilde{r}^{t+1}$ such that $\tilde{r}_i^{t+1} \in Y_i^t$: that is, recommendations after period t^* exclude codominated actions.

Let $E \subset \tilde{F}|_{\tilde{Q}} \times Y^{t^*}$ denote the set of pairs $\left(\tilde{f}, y^{t^*}\right)$ such that $\tilde{f}_{i,t^*}(y^{t^*}) \in \mathfrak{D}_{i,t^*}(y_i^{t^*})$ for some i. Define $\tilde{\mu}_{t^*}^{E}\left(\tilde{f}, y^{t^*}\right) = \tilde{\mu}_{t^*}(\tilde{f}, y^{t^*}|E)$. For each $\tilde{m}_{i,t^*} \in \mathfrak{D}_{i,t^*}(y_i^{t^*})$, the conditioning event $\tilde{m}_{i,t^*}$ in (46) implies that the realized pair $\left(\tilde{f}, y^{t^*}\right)$ lies in E. Hence, (46) implies that, for each i and $\tilde{m}_{i,t^*} \in \tilde{M}_{i,t^*}\left(y_i^{t^*}\right) \cap \mathfrak{D}_{i,t^*}(y_i^{t^*})$, we have

$$\sum_{\substack{\left(\tilde{f}, y^{t^*}\right) \in \tilde{F}|_{\tilde{Q}} \times Y^{t^*}: \\ \tilde{f}_{i,t^*}(y^{t^*}) = \tilde{m}_{i,t^*}}} \tilde{\mu}_{t^*}^{E}(\tilde{f}, y^{t^*}) \bar{u}_i\left(\sigma^*, \tilde{f} | \bar{h}\left(\tilde{f}, y^{t^*}\right)\right) \geq \sum_{\substack{\left(\tilde{f}, y^{t^*}\right) \in \tilde{F}|_{\tilde{Q}} \times Y^{t^*}: \\ \tilde{f}_{i,t^*}(y^{t^*}) = \tilde{m}_{i,t^*}}} \tilde{\mu}_{t^*}^{E}(\tilde{f}, y^{t^*}) \bar{u}_i\left(\sigma_i', \sigma_{-i}^*, \tilde{f} | \bar{h}\left(\tilde{f}, y^{t^*}\right)\right).$$

This contradicts the hypothesis that $\tilde{m}_{i,t^*} \in \mathfrak{D}_{i,t^*}(y_i^{t^*})$, which completes the proof of Lemma 10.

We now prove Lemma 11. Let $\phi^{\geq t}$ denote a collection of functions $(\phi_\tau)_{\tau \geq t}$, where $\phi_t : R^{*t+1} \to \Delta\left(M_t^* \times \mathcal{Z}^{A,t}|_Q\right)$ (where here $\mathcal{Z}^{A,t}|_Q \subset F|_Q \times H^{A,t}$ denotes the set of mediation plans and period-t acting histories in G with mediation range Q), and for $\tau > t$, $\phi_\tau : \mathcal{Z}^{A,\tau-1}|_Q \times R_\tau^* \to \Delta\left(M_\tau^* \times H^{A,\tau}\right)$. Intuitively, $\phi^{\geq t}$ may be viewed as a continuation strategy for the mediator in the direct-communication game starting with arbitrary past reports $\tilde{r}^{t+1} \in R^{*t+1}$ in period t.

Lemma 12 *For each t, in game G^*, there exists $((\phi_\tau^k)_{\tau \geq t})_k$ with limit $\phi_\tau = \lim_{k\to\infty} \phi_\tau^k$ for each $\tau \geq t$ such that, for each i, $y_i^t \in Y_i^t$, and $\tilde{m}_{i,t} \in \tilde{M}_{i,t}\left(y_i^t\right)$, we have*

$$\sum_{y^t \in Y^t[y_i^t]} \bar{\mu}^k(y^t | y_i^t) \phi_t^k\left(\tilde{m}_{i,t} | y^t\right) > 0 \text{ for all } k, \tag{47}$$

and, for all $\sigma_i' \in \Sigma_i^*$,

$$\begin{aligned} &\sum_{y^t \in Y^t[y_i^t]} \bar{\mu}(y^t | y_i^t, \tilde{m}_{i,t}) \bar{u}_i\left(\sigma^*, (\phi_\tau)_{\tau \geq t} | y^t, \left\{\tilde{r}^{t+1} = y^t\right\}, \tilde{m}_{i,t}\right) \qquad (48) \\ \geq &\sum_{y^t \in Y^t[y_i^t]} \bar{\mu}(y^t | y_i^t, \tilde{m}_{i,t}) \bar{u}_i\left(\sigma_i', \sigma_{-i}^*, (\phi_\tau)_{\tau \geq t} | y^t, \left\{\tilde{r}^{t+1} = y^t\right\}, \tilde{m}_{i,t}\right), \end{aligned}$$

where $\bar{\mu}$ is defined by

$$\bar{\mu}(y^t | y_i^t, \tilde{m}_{i,t}) = \lim_{k\to\infty} \frac{\bar{\mu}^k(y^t | y_i^t) \phi_t^k(\tilde{m}_{i,t} | y^t)}{\sum_{\tilde{y}^t \in Y^t[y_i^t]} \bar{\mu}^k(\tilde{y}^t | y_i^t) \phi_t^k(\tilde{m}_{i,t} | \tilde{y}^t)}.$$

Proof. Construction of $\left(\phi_\tau^k\right)_{\tau \geq t}$: This is similar to the proof of Proposition 2. For each k, first define $\left(\phi_\tau^k\right)_{\tau \geq t}$ recursively in τ, then define $\phi_\tau = \lim_{k\to\infty} \phi_\tau^k$ for all $\tau \geq t$.

For each canonical $\tilde{r}^{t+1} \in R^{*t+1}$, the mediator draws a mediation plan and "fictitious history" $\left(f, h^{A,t}\right) \in \mathcal{Z}^{A,t}|_Q$ according to $\bar{\mu}^k(f, h^{A,t}|y^t)$ for $y^t = \tilde{r}^{t+1}$.[49] Then, he recommends $\tilde{m}_{i,t} \in A_{i,t}$ to player i according to $\sigma^A_{i,t}(h^{A,t})$. This defines ϕ^k_t.

For each $\tau > t$, we now define ϕ^k_τ as a function of $\left(f, h^{A,\tau-1}, a_{i,\tau-1}, s_{i,\tau}\right)$ with $(a_{i,\tau-1}, s_{i,\tau}) = \tilde{r}_{i,\tau}$. For each i, the mediator draws a "fictitious report" $r_{i,\tau} \in R_{i,\tau}$ according to $\sigma^R_{i,\tau}(h_i^{A,\tau-1}, \tilde{r}_{i,\tau})$, independently across players. Next, given r_τ, the mediator calculates the vector of "fictitious messages" $m_\tau = f_\tau(r^\tau, r_\tau)$, where r^τ is the report component of $h^{A,\tau-1}$. Finally, the mediator draws recommendation $\tilde{m}_{i,\tau} \in A_{i,\tau}$ according to $\sigma^A_{i,\tau}(h_i^{A,\tau-1}, a_{i,\tau-1}, s_{i,\tau}, r_{i,\tau}, m_{i,\tau})$ with $(a_{i,\tau-1}, s_{i,\tau}) = \tilde{r}_{i,\tau}$, independently across players. This then defines $h_i^{A,\tau} = (h_i^{A,\tau-1}, a_{i,\tau-1}, s_{i,\tau}, r_{i,\tau}, m_{i,\tau})$ with $(a_{i,\tau-1}, s_{i,\tau}) = \tilde{r}_{i,\tau}$.

Proof of (47): For each $y^t \in Y^t$, since $\bar{\mu}^k(f, h^{A,t}|y^t)$ has full support over $\left(f, h^{A,t}\right) \in \mathcal{Z}^{A,t}\left[y^t\right]|_Q$, we have $\phi^k_t(\tilde{m}_{i,t}|y^t) > 0$ for each i and $\tilde{m}_{i,t} \in \tilde{M}_{i,t}\left(y^t_i\right)$. Hence, (47) is satisfied.

Proof of (48): Toward a contradiction, suppose (48) is violated for some i, $y^t_i \in Y^t_i$, $\tilde{m}_{i,t} \in \tilde{M}_{i,t}\left(y^t_i\right)$, and $\sigma'_i \in \Sigma^*_i$. Denote the conditional probability of $\left(f, h^{A,t}\right) \in \mathcal{Z}^{A,t}|_Q$ and $x \in X$ given y^t, $\tilde{r}^{t+1} = y^t$, and $\tilde{m}_{i,t}$ by

$$\Pr{}^{\sigma^*,(\phi_\tau)_{\tau\geq t}}\left(f, h^{A,t}, x|y^t, \left\{\tilde{r}^{t+1} = y^t\right\}, \tilde{m}_{i,t}\right).$$

By construction, for each $\tilde{\sigma} \in \Sigma^*$,

$$\begin{aligned}
& \Pr{}^{\tilde{\sigma},(\phi_\tau)_{\tau\geq t}}\left(f, h^{A,t}, x|y^t, \left\{\tilde{r}^{t+1} = y^t\right\}, \tilde{m}_{i,t}\right) \\
= \; & \lim_{k\to\infty} \Pr{}^{\phi^k_t}\left(f, h^{A,t}|y^t, \left\{\tilde{r}^{t+1} = y^t\right\}, \tilde{m}_{i,t}\right) \Pr{}^{\tilde{\sigma},(\phi_\tau)_{\tau\geq t}}\left(x|y^t, \left\{\tilde{r}^{t+1} = y^t\right\}, f, h^{A,t}, \tilde{m}_{i,t}\right) \\
= \; & \bar{\mu}(f, h^{A,t}|y^t, \tilde{m}_{i,t}) \Pr{}^{\tilde{\sigma},(\phi_\tau)_{\tau\geq t}}\left(x|y^t, \left\{\tilde{r}^{t+1} = y^t\right\}, f, h^{A,t}, \tilde{m}_{i,t}\right) \\
= \; & \bar{\mu}(f, h^{A,t}|y^t, \tilde{m}_{i,t}) \Pr{}^{\tilde{\sigma},(\phi_\tau)_{\tau\geq t}}\left(x|f, h^{A,t}, \tilde{m}_{i,t}\right).
\end{aligned}$$

In the last line, we omit y^t since $\Pr^{\phi^k_t}\left(\cdot|y^t, \tilde{m}_{i,t}\right)$ assigns probability 1 to $\mathring{h}^{A,t} = y^t$, and we also omit $\{\tilde{r}^{t+1} = y^t\}$ by defining $\Pr^{\tilde{\sigma},(\phi_\tau)_{\tau\geq t}}\left(x|f, h^{A,t}, \tilde{m}_{i,t}\right)$ as follows: conditional on $h^{A,t}$, define $\tilde{r}^{t+1} = \mathring{h}^{A,t}$ and calculate the conditional distribution of x given $\tilde{\sigma}$ and $\left(f, h^{A,t}, \tilde{r}^{t+1}, \tilde{m}_{i,t}\right)$. By definition, for each $y^t \in Y^t[y^t_i]$,

$$\bar{\mu}(y^t|y^t_i, \tilde{m}_{i,t})\bar{\mu}(f, h^{A,t}|y^t, \tilde{m}_{i,t}) = \bar{\mu}(f, h^{A,t}|y^t_i, \tilde{m}_{i,t}) 1_{\left\{\mathring{h}^{A,t} = y^t\right\}}.$$

Hence, the violation of (48) implies

$$\begin{aligned}
& \sum_{(f,h^{A,t})\in\mathcal{Z}^{A,t}|_Q : \mathring{h}^{A,t}\in Y^t[y^t_i]} \bar{\mu}(f, h^{A,t}|y^t_i, \tilde{m}_{i,t})\bar{u}_i\left(\sigma^*, (\phi_\tau)_{\tau\geq t}|f, h^{A,t}, \tilde{m}_{i,t}\right) \\
< \; & \sum_{(f,h^{A,t})\in\mathcal{Z}^{A,t}|_Q : \mathring{h}^{A,t}\in Y^t[y^t_i]} \bar{\mu}(f, h^{A,t}|y^t_i, \tilde{m}_{i,t})\bar{u}_i\left(\sigma'_i, \sigma^*_{-i}, (\phi_\tau)_{\tau\geq t}|f, h^{A,t}, \tilde{m}_{i,t}\right).
\end{aligned}$$

---

[49] As in the proof of Proposition 2, $\left(f, h^{A,t}\right)$ can be chosen arbitrarily if $\tilde{r}^{t+1}$ is not a feasible payoff-relevant history. A similar comment applies to $r_{i,\tau}$ and $h_i^{A,\tau}$ in the next paragraph.

Therefore, there must exist $h_i^{A,t}$ with $\bar{\mu}(h_i^{A,t}|y_i^t, \tilde{m}_{i,t}) > 0$ such that

$$\begin{aligned}
& \sum_{(f,h^{A,t})\in \mathcal{Z}^{A,t}[h_i^{A,t}]|_Q} \bar{\mu}(f, h^{A,t}|h_i^{A,t}, \tilde{m}_{i,t})\bar{u}_i\left(\sigma^*, (\phi_\tau)_{\tau\geq t}\,|f, h^{A,t}, \tilde{m}_{i,t}\right) \\
< & \sum_{(f,h^{A,t})\in \mathcal{Z}^{A,t}[h_i^{A,t}]|_Q} \bar{\mu}(f, h^{A,t}|h_i^{A,t}, \tilde{m}_{i,t})\bar{u}_i\left(\sigma_i', \sigma_{-i}^*, (\phi_\tau)_{\tau\geq t}\,|f, h^{A,t}, \tilde{m}_{i,t}\right). \qquad (49)
\end{aligned}$$

Note that $(\phi_\tau)_{\tau\geq t}$ is constructed so that the conditional distribution of x given $\left(f, h^{A,t}\right)$ and $\tilde{m}_{i,t}$ under $\left(\sigma^*, (\phi_\tau)_{\tau\geq t}\right)$ in game G^* is the same as the conditional distribution of x given $\left(f, h^{A,t}\right)$ and $a_{i,t} = \tilde{m}_{i,t} \in \operatorname{supp} \sigma_{i,t}^A(h_i^{A,t})$ in game G with mediation range Q under (σ, f):

$$\begin{aligned}
& \sum_{(f,h^{A,t})\in \mathcal{Z}^{A,t}[h_i^{A,t}]|_Q} \bar{\mu}(f, h^{A,t}, \tilde{m}_t|h_i^{A,t}, \tilde{m}_{i,t})\bar{u}_i\left(\sigma^*, (\phi_\tau)_{\tau\geq t}\,|f, h^{A,t}, \tilde{m}_{i,t}\right) \\
= & \sum_{(f,h^{A,t})\in \mathcal{Z}^{A,t}[h_i^{A,t}]|_Q} \bar{\mu}(f, h^{A,t}, \tilde{m}_t|h_i^{A,t}, \tilde{m}_{i,t})\bar{u}_i\left(\sigma|f, h^{A,t}, \tilde{m}_{i,t}\right).
\end{aligned}$$

To derive a contradiction, it suffices to find a strategy $\hat{\sigma}_i'$ in G with mediation range Q that attains the expected payoff in the second line of (49),

$$\begin{aligned}
& \sum_{(f,h^{A,t})\in \mathcal{Z}^{A,t}[h_i^{A,t}]|_Q} \bar{\mu}(f, h^{A,t}|h_i^{A,t}, \tilde{m}_{i,t})\bar{u}_i\left(\sigma_i', \sigma_{-i}^*, (\phi_\tau)_{\tau\geq t}\,|f, h^{A,t}, \tilde{m}_{i,t}\right) \\
= & \sum_{(f,h^{A,t})\in \mathcal{Z}^{A,t}[h_i^{A,t}]|_Q} \bar{\mu}(f, h^{A,t}|h_i^{A,t}, \tilde{m}_{i,t})\bar{u}_i\left(\hat{\sigma}_i', \sigma_{-i}|f, h^{A,t}, \tilde{m}_{i,t}\right),
\end{aligned}$$

since the existence of such a strategy contradicts (44). The same construction as in the proof of Lemma 1 defines such a strategy $\hat{\sigma}_i'$. ■

Proof of Lemma 11. We now prove (46). For each k, by Kuhn's theorem, there exists a collection of mixed mediation plan $\left(\tilde{\mu}_{\tilde{r}^{t+1}}^k\right)_{\tilde{r}^{t+1}}$ in G^*, with $\tilde{\mu}_{\tilde{r}^{t+1}}^k \in \Delta\left(\tilde{F}\right)$ for each $\tilde{r}^{t+1}$, that satisfies the following condition: for each $y^t \in Y^t$, $\tilde{r}^{t+1} \in R^{*t+1}$, strategy $\sigma' \in \Sigma^*$, and vector $(\tilde{m}_t, a_t, (s_\tau, \tilde{r}_\tau, \tilde{m}_\tau, a_\tau)_{\tau=t+1}^T)$, we have

$$\begin{aligned}
& \Pr{}^{\sigma', \left(\phi_\tau^k\right)_{\tau\geq t}}\left(\tilde{m}_t, a_t, (s_\tau, \tilde{r}_\tau, \tilde{m}_\tau, a_\tau)_{\tau=t+1}^T\,|y^t, \tilde{r}^{t+1}\right) \\
= & \sum_{\tilde{f}\in\Delta(\tilde{F})} \tilde{\mu}_{\tilde{r}^{t+1}}^k(\tilde{f}) \Pr{}^{\sigma'}\left(\tilde{m}_t, a_t, (s_\tau, \tilde{r}_\tau, \tilde{m}_\tau, a_\tau)_{\tau=t+1}^T\,|f, y^t, \tilde{r}^{t+1}\right).
\end{aligned}$$

(That is, $\left(\sigma', \left(\phi_\tau^k\right)_{\tau\geq t}\right)$ and $\left(\sigma', \tilde{\mu}_{\tilde{r}^{t+1}}^k\right)$ give rise to the same distribution of histories in G^*

conditional on y^t and $\tilde{r}^{t+1}$.) In particular, if $\tilde{r}^{t+1} = y^t$, we have

$$\begin{aligned}
&\Pr^{\sigma', (\phi^k_\tau)_{\tau \geq t}} \left(\tilde{m}_t, a_t, (s_\tau, \tilde{r}_\tau, \tilde{m}_\tau, a_\tau)_{\tau=t+1}^T \, | y^t \right) \\
= &\sum_{\tilde{f} \in \Delta(\tilde{F})} \tilde{\mu}^k_{y^t}(\tilde{f}) \Pr^{\sigma'} \left(\tilde{m}_t, a_t, (s_\tau, \tilde{r}_\tau, \tilde{m}_\tau, a_\tau)_{\tau=t+1}^T \, | \tilde{f}, y^t \right).
\end{aligned}$$

Define

$$\tilde{\mu}^k(y^t, \tilde{f}, \tilde{m}_t) = \bar{\mu}^k(y^t) \times \tilde{\mu}^k_{y^t}(\tilde{f}) \times 1_{\left\{\tilde{m}_t = \tilde{f}(y^t)\right\}}.$$

Let $\tilde{\mu} = \lim_{k \to \infty} \tilde{\mu}^k$. Since each $\tilde{f} \in \operatorname{supp} \tilde{\mu}$ satisfies $\tilde{f}_{i,\tau}(\tilde{r}^{\tau+1}) \in \tilde{Q}_{i,\tau}(\tilde{r}^{t+1}_\tau)$ for all $\tau \geq t$, (48) implies

$$\begin{aligned}
&\sum_{\left(\tilde{f}, y^t\right) \in \tilde{F}|_{\tilde{Q}} \times Y^t[y^t_i]} \tilde{\mu}_t(\tilde{f}, y^t | y^t_i, \tilde{m}_{i,t}) \bar{u}_i \left(\sigma^*, \tilde{f} | y^t, \left\{ \tilde{r}^{t+1} = y^t \right\}, \tilde{m}_{i,t} \right) \\
\geq &\sum_{\left(\tilde{f}, y^t\right) \in \tilde{F}|_{\tilde{Q}} \times Y^t[y^t_i]} \tilde{\mu}_t(\tilde{f}, y^t | y^t_i, \tilde{m}_{i,t}) \bar{u}_i \left(\sigma'_i, \sigma^*_{-i}, \tilde{f} | y^t, \left\{ \tilde{r}^{t+1} = y^t \right\}, \tilde{m}_{i,t} \right). \qquad (50)
\end{aligned}$$

For each $\tilde{f}$, we can write $\tilde{f} = (\tilde{f}^{<t}, \tilde{f}^{\geq t})$ with $\tilde{f}^{<t} = (\tilde{f}_\tau)_{\tau=1}^{t-1}$ and $\tilde{f}^{\geq t} = (\tilde{f}_\tau)_{\tau=t}^T$. Since the past recommendations $\tilde{m}^t$ do not affect the continuation strategy, there exists $\hat{\mu}_t$ such that $\hat{\mu}_t(\tilde{f}) = \hat{\mu}^{<t}_t(\tilde{f}^{<t}) \times \hat{\mu}^{\geq t}_t(\tilde{f}^{\geq t})$ and (50) holds with $\hat{\mu}_t$ in place of $\tilde{\mu}_t$. Since under $\hat{\mu}_t$ recommendations prior to period t are independent of those after period t, this yields (46). ■

I.4 Proof of Propositions 7 and 8

Proposition 7: By Lemma 9, for each SCE $(\mu, Q, \bar{\mu})$, there exists a canonical CPPBE $(\sigma, \mu, Q, \bar{\mu}')$ in G^* with $\rho^{\sigma,\mu} = \rho^{\sigma^*,\mu}$. Conversely, take a CPPBE $(\sigma, \mu, Q, \bar{\mu})$ in G^* with canonical σ and outcome ρ. As in the proof of Lemma 9, let J be the set of histories such that players are honest and messages are compatible with mediation range Q, let $P = R$, and let $F|_P$ be the set of mediation plans compatible with mediation range Q. Since σ is canonical, the quasi-CPPBE $(\sigma, \mu, J, P, F|_P)$ is SCE.

Proposition 8: By Lemma 10, players do not take codominated actions at any history for any $\mathfrak{C}$. Since every CPPBE $(\sigma, \mu, Q, \bar{\mu})$ is a NE, there exists a NE with outcome ρ where players do not take codominated actions at any history. Hence, by Proposition 1, there exists a SCE $(\mu', Q', \bar{\mu}')$ with $\rho = \rho^{\sigma^*,\mu}$ and $Q_{i,t}\left(r^{t+1}_i, m^t_i\right) = A_{i,t} \setminus \mathfrak{D}_{i,t}(r^{t+1}_i)$. Hence, by Lemma 9, there exists a canonical CPPBE $(\sigma, \mu', Q', \bar{\mu}'')$ with outcome ρ.

J Proof of Claim 5 of Proposition 4

We first establish a preliminary result. Fix a game G, mediation range Q, and CPPBE $(\sigma, \mu, Q, \bar{\mu})$. For each i, let $\Sigma_i\left(\sigma_{-i}, \mu, \bar{\mu}\right) \subset \Sigma_i$ denote the set of sequentially rational strategies for player i against (σ_{-i}, μ) under CPS $\bar{\mu}$: that is, the set of strategies $\hat{\sigma}_i$ such that, for each

t and $h_i^{R,t}$,

$$\sum_{(f,h^{R,t})\in\mathcal{Z}^{R,t}[h_i^{R,t}]|_Q} \bar{\mu}(f,h^{R,t}|h_i^{R,t})\bar{u}_i\left(\hat{\sigma}_i,\sigma_{-i},f|h^{R,t}\right) = \max_{\sigma_i'\in\Sigma_i} \sum_{(f,h^{R,t})\in\mathcal{Z}^{R,t}[h_i^{R,t}]|_Q} \bar{\mu}(f,h^{R,t}|h_i^{R,t})\bar{u}_i\left(\sigma_i',\sigma_{-i},f|h^{R,t}\right)$$

and, for each $h_i^{A,t}$,

$$\sum_{(f,h^{A,t})\in\mathcal{Z}^{A,t}[h_i^{A,t}]|_Q} \bar{\mu}(f,h^{A,t}|h_i^{A,t})\bar{u}_i\left(\hat{\sigma}_i,\sigma_{-i},f|h^{A,t}\right) = \max_{\sigma_i'\in\Sigma_i} \sum_{(f,h^{A,t})\in\mathcal{Z}^{A,t}[h_i^{A,t}]|_Q} \bar{\mu}(f,h^{A,t}|h_i^{A,t})\bar{u}_i\left(\sigma_i',\sigma_{-i},f|h^{A,t}\right).$$

Let $\hat{M}_{i,t}\left(y_i^t\right) = \bigcup_{h_i^{A,t}:\mathring{h}_i^{A,t}=y_i^t}\bigcup_{\hat{\sigma}_i\in\Sigma_i^*(\sigma_{-i},\mu,\bar{\mu})} \operatorname{supp}\hat{\sigma}_{i,t}^A(h_i^{A,t})$. The following lemma shows that codominated actions are never taken by any sequentially rational strategy $\hat{\sigma}_i\in\Sigma_i\left(\sigma_{-i},\mu,\bar{\mu}\right)$.

Lemma 13 *For any game G, mediation range Q, and CPPBE $(\sigma,\mu,\bar{\mu})$, $\hat{M}_{i,t}\left(y_i^t\right)\cap\mathfrak{D}_{i,t}\left(y_i^t\right)=\emptyset$ for all i, t, and $y_i^t\in Y_i^t$.*

Proof. Suppose otherwise that there exists t such that $\hat{M}_{i,t}\left(y_i^t\right)\cap\mathfrak{D}_{i,t}\left(y_i^t\right)\neq\emptyset$ for some i and $y_i^t\in Y_i^t$. Let t^* be the last such period, and fix i, $\hat{\sigma}_i\in\Sigma_i\left(\sigma_{-i},\mu,\bar{\mu}\right)$, $y_i^{t^*}$, and action $\tilde{m}_{i,t^*}$ such that $\tilde{m}_{i,t^*}\in\bigcup_{h_i^{A,t^*}:\mathring{h}_i^{A,t^*}=y_i^{t^*}}\operatorname{supp}\hat{\sigma}_{i,t^*}^A(h_i^{A,t^*})\cap\mathfrak{D}_{i,t^*}(y_i^{t^*})$.

For each t, $y_i^t\in Y_i^t$, and $\tilde{r}_i^{t+1}\in R_i^{*t+1}$, let

$$\hat{Q}_{i,t}(\tilde{r}_i^{t+1}) = \begin{cases} \hat{M}_{i,t}\left(\tilde{r}_i^{t+1}\right) & \text{if } \tilde{r}_i^{t+1}\in Y_i^t \\ A_{i,t} & \text{otherwise} \end{cases}$$

and let $\hat{Q}=(\hat{Q}_i,\tilde{Q}_{-i})$, where $\tilde{Q}_j$ is defined in (45) for $j\neq i$. By Lemma 10, for each t the range of $\tilde{Q}_{-i,t}$ excludes all codominated actions; and by definition of t^*, for each $t>t^*$, the range of $\hat{Q}_{i,t}$ excludes all codominated actions as well.

Applying the same construction as in the proof of Lemma 10 with $\hat{\sigma}_i$ in place of σ_i yields a CPS $\tilde{\mu}_{t^*}$ on $\tilde{F}|_{\hat{Q}}\times Y^{t^*}$ such that, for each $\sigma_i'\in\Sigma_i^*$,

$$\begin{aligned} &\sum_{\left(\tilde{f},y^t\right)\in\tilde{F}|_{\hat{Q}}\times Y^t[y_i^t]} \tilde{\mu}_t(\tilde{f},y^t|y_i^t,\tilde{m}_{i,t})\bar{u}_i\left(\hat{\sigma}_i,\sigma_{-i},\tilde{f}|\bar{h}\left(\tilde{f},y^t\right),\tilde{m}_{i,t}\right) \\ \geq &\sum_{\left(\tilde{f},y^t\right)\in\tilde{F}|_{\hat{Q}}\times Y^t[y_i^t]} \tilde{\mu}_t(\tilde{f},y^t|y_i^t,\tilde{m}_{i,t})\bar{u}_i\left(\sigma_i',\sigma_{-i},\tilde{f}|\bar{h}\left(\tilde{f},y^t\right),\tilde{m}_{i,t}\right). \end{aligned}$$

Let $E\subset\tilde{F}|_{\tilde{Q}}\times Y^{t^*}$ denote the set of pairs $\left(\tilde{f},y^{t^*}\right)$ such that $\tilde{f}_{i,t^*}(y^{t^*})=\tilde{m}_{i,t^*}$. Letting $\tilde{\mu}_{t^*}^E\left(\tilde{f},y^{t^*}\right)=\tilde{\mu}_{t^*}(\tilde{f},y^{t^*}|E)$, we have

$$\sum_{\substack{\left(\tilde{f},y^{t^*}\right)\in F|_{\hat{Q}}\times Y^{t^*}: \\ \tilde{f}_{i,t^*}(y^{t^*})=\tilde{m}_{i,t^*}}} \tilde{\mu}_{t^*}^E(\tilde{f},y^{t^*})\bar{u}_i\left(\sigma^*,\tilde{f}|\bar{h}\left(\tilde{f},y^{t^*}\right)\right) \geq \sum_{\substack{\left(\tilde{f},y^{t^*}\right)\in F|_{\hat{Q}}\times Y^{t^*}: \\ \tilde{f}_{i,t^*}(y^{t^*})=\tilde{m}_{i,t^*}}} \tilde{\mu}_{t^*}^E(\tilde{f},y^{t^*})\bar{u}_i\left(\sigma_i',\sigma_{-i}^*,\tilde{f}|\bar{h}\left(\tilde{f},y^{t^*}\right)\right).$$

This contradicts the hypothesis that $\tilde{m}_{i,t^*}\in\mathfrak{D}_{i,t^*}(y_i^{t^*})$. ■

Turning to the proof of the claim, note that in games of pure moral hazard, for each i and t, the set of codominated actions for player i in period t does not depend on the realized payoff-relevant history y_i^t: that is, there exists a set $D_{i,t} \subset A_{i,t}$ such that $\mathfrak{D}_{i,t}\left(y_i^t\right) = D_{i,t}$ for all $y_i^t \in Y_i^t$. Let $\bar{Y}_i^t$ be the set of payoff-relevant histories that can be reached without any player taking a codominated action:

$$\left\{y_i^t \in Y_i^t : \exists y_{-i}^t \text{ s.t. } \left(y_i^t, y_{-i}^t\right) \in Y^t \text{ and } a_{n,\tau} \in A_{n,\tau} \setminus D_{n,\tau} \text{ for each } n \text{ and } \tau \leq t-1\right\}.$$

Define L and π_t^k as in the proof of Proposition 5, where now each $\pi_t^{[l]} \in \Delta\left(F^{*\geq t} \times Y^t\right)$ depends only on its first component.

Fix a canonical NE (σ^*, μ^*) in G^* where players never take codominated actions at any history.[50] We first construct a (possibly non-canonical) SE $(\tilde{\sigma}, \tilde{\mu})$ in G^* with outcome ρ^{σ^*,μ^*}, and then construct a canonical SE $(\tilde{\sigma}^*, \tilde{\mu}^*)$ with the same outcome.

Construction of $\left(\tilde{\sigma}^k, \tilde{\mu}^k\right)$ We first construct a sequence of profiles $\left(\tilde{\sigma}^k, \tilde{\mu}^k\right)$ indexed by k that limits to a (quasi) SE in G^*.

Mediator's Strategy: At the beginning of the game, the mediator draws $\zeta_t \in \{0,1\}$ for each $t = 0, ..., T$ such that $\zeta_t = 1$ with probability $\left(\frac{1}{k}\right)^{2(L+1)T}$, independently across periods. Given $(\zeta_t)_{t=0}^T$, the mediator draws $(\omega_t)_{t=0}^T$ as follows: For $t = 0$, $\omega_0 = (0, f)$, where f is distributed according to $\mu^*(f)$. For each $t \geq 1$, given $\left(\omega_{t-1}, \zeta_{t-1}\right)$, ω_t is determined as follows:

1. If $\zeta_{t-1} = 0$, then $\omega_t = \omega_{t-1}$ with probability $1 - \frac{1}{k}$, and $\omega_t = \left(t, f^{\geq t}\right)$ with probability $\frac{1}{k}\pi_t^k\left(f^{\geq t}\right)$.
2. If $\zeta_{t-1} = 1$, then $\omega_t = \left(t, f^{\geq t}\right)$ with probability $\pi_t^k\left(f^{\geq t}\right)$.

Given (ζ_t, ω_t), the mediator's recommendation is determined as follows: If $\zeta_t = 0$, then $m_t = f\left(r^{t+1}\right)$ if $\omega_t = \omega_0$ and $m_t = f^{\geq \tau}\left(r^{t+1}\right)$ if $\omega_t = \left(\tau, f^{\geq \tau}\right)$. If $\zeta_t = 1$, then the mediator draws each $m_{i,t} \in A_{i,t} \setminus D_{i,t}$ with equal probability $1/\left(|A_{i,t}| - |D_{i,t}|\right)$, independently across i and t.

Each realization of the mediator's randomization defines the realization of the first element of ω_t, for each t. Let $\mathbf{t}^*(t)$ be the corresponding random variable, which takes values in $\{0, ..., t\}$.

Player i's Strategy: We say that player i is *honest and obedient* at history $h_i^{R,t}$ if $r_{i,\tau} = (a_{i,\tau-1}, s_{i,\tau})$ and $a_{i,\tau} = m_{i,\tau}$ for all $\tau < t$, and is *honest and obedient* at history $h_i^{A,t}$ if $r_{i,\tau} = (a_{i,\tau-1}, s_{i,\tau})$ for all $\tau \leq t$ and $a_{i,\tau} = m_{i,\tau}$ for all $\tau < t$. Let $\hat{J}_i^{R,t}$ (resp., $\hat{J}_i^{A,t}$) be the set of histories $h_i^{R,t}$ (resp., $h_i^{A,t}$) such that player i has been honest and obedient and $\mathring{h}_i^{R,t} \in \bar{Y}_i^t$.

For each k, let $\left(\Gamma^k, \mathfrak{C}^*\right)$ denote the following auxiliary game:

1. The mediator follows $\tilde{\mu}^k$.
2. Each player i chooses probability distributions $\sigma_{i,t}^{R,k}(\cdot|h_i^{R,t}) \in \Delta(R_{i,t})$ and $\sigma_{i,t}^{A,k}(\cdot|h_i^{A,t}) \in \Delta(A_{i,t} \setminus D_{i,t})$ for each t, $h_i^{R,t} \in H_i^{R,t} \backslash \hat{J}_i^{R,t}$ and $h_i^{A,t} \in H_i^{A,t} \backslash \hat{J}_i^{A,t}$. Note that player i is

---

[50] As in footnote 45, it is without loss to consider here the fully canonical strategy profile σ^*.

required not to take a codominated action. At histories $h_i^{R,t} \in \hat{J}_i^{R,t}$ and $h_i^{A,t} \in \hat{J}_i^{A,t}$, player i is required to report $r_{i,t} = (a_{i,t-1}, s_{i,t})$ and take $a_{i,t} = m_{i,t}$, respectively.

3. Given $\left(\sigma^k, \tilde{\mu}^k\right)$, the distribution of terminal nodes H^{T+1} is determined recursively as follows:

   Given $h^{R,t} \in H^{R,t}$, each $r_{i,t} \in R_{i,t}$ is drawn independently across players with probability $\sigma_{i,t}^{R,k}\left(r_{i,t}|h_i^{R,t}\right)$.

   Given $h^{A,t} \in H^{A,t}$, each $a_{i,t} \in R_{i,t}$ is drawn independently across players with probability
$$\left(1 - |A_{i,t}|\left(\frac{1}{k}\right)^{3(L+1)T^2}\right)\sigma_{i,t}^{A,k}(a_{i,t}|h_i^{A,t}) + \left(\frac{1}{k}\right)^{3(L+1)T^2}.$$

   Given $h^{A,t} \in H_t^{A,t}$ and $a_t \in A_t$, each $s_{t+1} \in S_{t+1}$ is drawn with probability $p\left(s_{t+1}|\mathring{h}^{A,t}, a_t\right)$.

4. Player i's payoff at terminal history $h^{t+1} \in H^{T+1}$ is $u_i\left(\mathring{h}^{t+1}\right)$.

As in the proof of Lemma 2, $\left(\Gamma^k, \mathfrak{C}^*\right)$ admits a NE $\left(\hat{\sigma}^k, \tilde{\mu}^k\right)$.

Now define strategy $\tilde{\sigma}^k$ by $\tilde{\sigma}_i^{R,k} = \hat{\sigma}_i^{R,k}$ and
$$\tilde{\sigma}_{i,t}^{A,k}(a_{i,t}|h_i^{A,t}) = \left(1 - |A_{i,t}|\left(\frac{1}{k}\right)^{3(L+1)T^2}\right)\hat{\sigma}_{i,t}^{A,k}(a_{i,t}|h_i^{A,t}) + \left(\frac{1}{k}\right)^{3(L+1)T^2}.$$

Note that the distribution over terminal histories under $\left(\hat{\sigma}^k, \tilde{\mu}^k\right)$ in game $\left(\Gamma^k, \mathfrak{C}^*\right)$ is the same as that under $\left(\tilde{\sigma}^k, \tilde{\mu}^k\right)$ in game $(\Gamma, \mathfrak{C}^*)$. Let $(\tilde{\sigma}, \tilde{\mu}) = \lim_{k\to\infty}\left(\tilde{\sigma}^k, \tilde{\mu}^k\right)$. Note that $(\tilde{\sigma}, \tilde{\mu})$ is a profile in $(\Gamma, \mathfrak{C}^*)$.

We next claim that, for each i and $\sigma_i' \in \Sigma_i^*$ such that $\operatorname{supp}\left(\sigma_i'^{A,t}(h_i^{A,t})\right) \cap D_{i,t} = \emptyset$ for all t and $h_i^{A,t}$, we have $\bar{u}(\tilde{\sigma}, \tilde{\mu}) \geq \bar{u}(\sigma_i', \tilde{\sigma}_{-i}, \tilde{\mu})$: that is, no player i has a profitable deviation that avoid codominated actions. To see this, let $\tilde{\phi}$ denote the behavioral mediation plan induced by $\tilde{\mu}$, and let ϕ^* denote the behavioral mediation plan induced by μ^*. Then $\tilde{\phi}_t\left(\cdot|r^{t+1}, m^t\right) = \phi_t^*\left(\cdot|r^{t+1}, m^t\right)$ for all (r^{t+1}, m^t) satisfying $\Pr^{\hat{\sigma},\tilde{\phi}}\left(r^{t+1}, m^t\right) > 0$ for some player strategy $\hat{\sigma}$. Similarly, for each i, $\tilde{\sigma}_{i,t}^R\left(\cdot|h_i^{R,t}\right) = \sigma_{i,t}^{*R}\left(\cdot|h_i^{R,t}\right)$ for all $h_i^{R,t}$ where player i has been honest and obedient and $\mathring{h}_i^{R,t} \in \bar{Y}_i^t$, and similarly for acting histories $h_i^{A,t}$. Hence, for any on-path history $h_i^{R,t}$ under $(\tilde{\sigma}, \tilde{\mu})$, the continuation play of i's opponents differs from that under (σ^*, μ^*) only following a deviation by player i to a codominated action. Therefore, for any $\sigma_i' \in \Sigma_i^*$ such that $\operatorname{supp}\left(\sigma_i'^{A,t}(h_i^{A,t})\right) \cap D_{i,t} = \emptyset$ for all t and $h_i^{A,t}$, the fact that (σ^*, μ^*) is a NE implies that $\bar{u}(\tilde{\sigma}, \tilde{\mu}) \geq \bar{u}(\sigma_i', \tilde{\sigma}_{-i}, \tilde{\mu})$.

Construction of Quasi-SE $(\tilde{\sigma}, \tilde{\mu}, \tilde{J}, \tilde{K}, \tilde{\beta})$**.** We now define subsets of histories $\tilde{J}$ and $\tilde{K}$ along with beliefs $\tilde{\beta}$ and a mediation range $\tilde{Q}$ such that $(\tilde{\sigma}, \tilde{\mu}, \tilde{J}, \tilde{K}, \tilde{\beta})$ is a quasi-SE in game $G^*|_{\tilde{Q}}$.

Define $\tilde{Q}_{i,t}(r_i^{t+1}, m_i^t) = A_{i,t} \setminus D_{i,t}$ for each i, t, and (r_i^{t+1}, m_i^t). For each i, let $\tilde{J}_i^{A,T+}$ equal the set of histories h_i^{T+1} such that $\Pr^{\tilde{\sigma}^k,\tilde{\mu}^k}\left(h_i^{T+1}\right) > 0$. Let $\tilde{K}^{T+}$ equal the set of mediator histories $\left(\tilde{r}^{T+1}, \tilde{m}^{T+1}\right)$ such that there exists $\sigma' \in \Sigma$ satisfying $\Pr^{\sigma',\tilde{\mu}^k}\left(r^{T+1}, m^{T+1}\right) > 0$. Let the other elements of $\tilde{J}$ and $\tilde{K}$ include all truncations of histories in $\tilde{J}^{A,T+}$ and $\tilde{K}^{T+}$. Since $\tilde{J}_i$ includes all histories where player i has not lied to the mediator or received a message outside the mediation range $\tilde{Q}$, $(\tilde{\sigma}, \tilde{\mu}, \tilde{J}, \tilde{K})$ is valid. Define belief system $\tilde{\beta}$ by

$$\tilde{\beta}_{i,t}\left(h^{R,t}|h_i^{R,t}\right) = \lim_{k\to\infty} \frac{\Pr^{\tilde{\sigma}^k,\tilde{\mu}^k}\left(h^{R,t}\right)}{\Pr^{\tilde{\sigma}^k,\tilde{\mu}^k}\left(h_i^{R,t}\right)}$$

for each $h^{R,t} \in H^{R,t}[h_i^{R,t}]|_{\tilde{J},\tilde{K}}$. Since $\Pr^{\tilde{\sigma}^k,\tilde{\mu}^k}\left(h_i^{R,t}\right) > 0$ for each $h_i^{R,t} \in \tilde{J}_i^{R,t}$, this is well-defined. Since $\tilde{\beta}$ is consistent by construction, it remains to verify sequential rationality of $(\tilde{\sigma}, \tilde{\mu}, \tilde{J}, \tilde{K}, \tilde{\beta})$.

As in the proof of Claim 1 of Proposition 4, sequential rationality at histories outside $\hat{J}_i^{R,t}$ or $\hat{J}_i^{A,t}$ follows from a standard upper hemi-continuity argument.

Fix any $h_i^{R,t} \in \hat{J}_i^{R,t}$. By Lemma 13, it suffices to show that player i never has a profitable deviation to a strategy that avoids codominated actions. By definition of $\hat{J}_i^{R,t}$, there exists $\left(\omega^t, \zeta^t\right) = (\omega_1, ..., \omega_{t-1}, \zeta_1, ..., \zeta_{t-1})$ such that $\Pr^{\sigma^*,\tilde{\mu}^k}\left(h_i^{R,t}|\omega^t, \zeta^t\right) > 0$, where as usual σ^* denotes the fully canonical strategy profile. Since under $\tilde{\mu}^k$ any sequence $\left(\omega^t, \zeta^t\right)$ occurs with probability at least $\left(\frac{1}{k}\right)^{2(L+1)T+2(L+1)T^2}$, while under $\tilde{\sigma}^k$ players' actions differ from those under $\hat{\sigma}^k$ with probability at most $\left(\frac{1}{k}\right)^{3(L+1)T^2}$, player i assesses that other players are honest and obedient with probability 1. Hence, it suffices to verify that, for each i, t, $\tau \leq t$, and $h_i^{R,t} \in \hat{J}_i^{R,t}$, $\tilde{\sigma}_i$ is sequentially rational conditional on the event that all players are honest and obedient and $\mathbf{t}^*(t) = \tau$. There are three mutually exclusive events to which player i assigns positive probability:

1. If $\mathbf{t}^*(t) = 0$ then $\Pr^{\tilde{\sigma},\tilde{\mu}}\left(h_i^{R,t}\right) > 0$. Sequential rationality follows since $\bar{u}(\tilde{\sigma}, \tilde{\mu}) \geq \bar{u}(\sigma_i', \tilde{\sigma}_{-i}, \tilde{\mu})$ for any strategy σ_i' that avoids codominated actions.

2. If $\mathbf{t}^*(t) = \tau < t$ then $\omega_t = \left(\tau, f^{\geq\tau}\right)$. Since the mediator draws $f^{\geq\tau}$ from π_τ^k, by the same argument as in the $\xi\left(h_i^{R,t}\right) = \left(t^*, \zeta_i^{t^*}\right)$ case of the proof of Proposition 5 (i.e., the discussion immediately following Lemma 6), honesty is optimal.

3. If $\mathbf{t}^*(t) = t$ then with probability 1 future recommendations are independent of the current report. Hence, honesty is optimal.

Next fix any $h_i^{A,t} \in \hat{J}_i^{A,t}$. Again, player i assesses that other players are honest and obedient with probability 1. She also assesses that $\zeta_t = 1$ with probability 0, since (i) for any $\left(h_i^{R,t}, r_{i,t}\right)$, each $m_{i,t} \in A_{i,t} \setminus D_{i,t}$ occurs with positive probability conditional on $\zeta_t = 0$ and $\mathbf{t}^*(t) = t$, (ii) $\zeta_t = 0$ and $\mathbf{t}^*(t) = t$ occur with probability at least $\left(1 - \left(\frac{1}{k}\right)^{2(L+1)T}\right)\left(\frac{1}{k}\right)^L$, while $\zeta_t = 1$ occurs with probability at most $\left(\frac{1}{k}\right)^{2(L+1)T}$. Hence, we again consider three mutually exclusive events: $\mathbf{t}^*(t) = 0$, $\mathbf{t}^*(t) = \tau < t$, and $\mathbf{t}^*(t) = t$ and $\zeta_t = 0$. The proof of sequential rationality in the first two events is the same as for reporting histories. In the last event, sequential rationality follows from the same argument as in the $\xi\left(h_i^{R,t}\right) = \left(t, \zeta_i^t\right)$ case of the proof of Proposition 5.

Construction of $\left(\tilde{\sigma}^{*,k}, \tilde{\mu}^{*,k}\right)$ We first construct a sequence of profiles $\left(\tilde{\sigma}^{*,k}, \tilde{\mu}^{*,k}\right)$ indexed by k that limits to a quasi-SE profile in G^*.

Definition of the Mediator's Strategy $\tilde{\mu}^{*,k}$. As in Proposition 2, in each period t, given player i's canonical report $\tilde{r}_i^{t+1} \in Y_i^t$, the mediator draws a fictitious report $r_{i,t}$ according to $\hat{\sigma}_{i,t}^{R,k}(h_i^{R,t})$, where $h_i^{R,t} = (y_i^t, r_i^t, m_i^t)$ with $y^t = \tilde{r}_i^{t+1}$.[51] Given fictitious reports r^{t+1} and fictitious messages m^t, the mediator draws fictitious message m_t from $\tilde{\phi}_t^k(r^{t+1}, m^t)$, where $\tilde{\phi}^k$ is the behavioral strategy induced by $\tilde{\mu}^k$. He then draws the recommendation $\tilde{m}_{i,t}$ according to $\hat{\sigma}_{i,t}^{A,k}(\tilde{m}_{i,t}|\tilde{r}_i^{t+1}, r_i^{t+1}, m_i^{t+1})$.

Definition of Player i*'s Strategy* $\tilde{\sigma}_i^{*,k}$. We define $\tilde{\sigma}_i^{*,k}$ to specify that player i is honest and obedient, but trembles uniformly over actions with probability $|A_{i,t}|\left(\frac{1}{k}\right)^{3(L+1)T^2}$: for each $a_{i,t} \in A_{i,t}$ and $h_i^{A,t} \in H_i^{A,t}$,

$$\tilde{\sigma}_{i,t}^{*A,k}(a_{i,t}|h_i^{A,t}) = 1_{\{a_{i,t}=m_{i,t}\}}\left(1 - |A_{i,t}|\left(\frac{1}{k}\right)^{3(L+1)T^2}\right) + \left(\frac{1}{k}\right)^{3(L+1)T^2}.$$

Construction of Quasi-SE $(\tilde{\sigma}^*, \tilde{\mu}^*, J, K, \beta)$ Define $\tilde{\sigma}^* = \lim_{k\to\infty} \tilde{\sigma}^{*,k}$ and $\tilde{\mu}^* = \lim_{k\to\infty} \tilde{\mu}^{*,k}$. We will construct J, K, and β such that $(\tilde{\sigma}^*, \tilde{\mu}^*, J, K, \beta)$ is a quasi-SE in G^*. By the first sentence of Lemma 2, this implies that there exists a SE in G^* that implements the same outcome. We will also construct a mediation range Q such that J includes all histories compatible with the mediation range where players have been honest. By the second sentence of Lemma 2, this implies that the SE is canonical.

Definition of Q. For each i and t, if $\tilde{r}_i^{t+1} \in Y_i^t$ we define

$$Q_{i,t}(\tilde{r}_i^{t+1}) = \bigcup_{\substack{\left(r_i^{t+1}, m_i^{t+1}\right) \\ \text{s.t. } r_{i,\tau} \in \operatorname{supp} \hat{\sigma}_{i,t}^{R,k}(y_i^\tau, r_i^\tau, m_i^\tau) \text{ with } y_i^\tau = \tilde{r}_i^{\tau+1} \text{ for each } \tau \leq t \\ m_{i,\tau} \in A_{i,\tau} \setminus D_{i,\tau} \text{ for each } \tau \leq t}} \operatorname{supp} \hat{\sigma}_{i,t}^{A,k}(\cdot|y_i^t, r_i^{t+1}, m_i^{t+1})$$

where $y_i^t = \tilde{r}_i^{t+1}$; and if $\tilde{r}_i^{t+1} \notin Y_i^t$ we define $Q_{i,t}(\tilde{r}_i^{t+1}) = A_{i,t}$.

Definition of (J, K, β). Define $K^{T+} = H_0^{T+1}$ and define $J_i^{A,T+}$ as the set of all histories compatible with the mediation range where players have been honest. Let the other elements of J and K include all truncations of histories in $J^{A,T+}$ and K^{T+}. Since $Q_{i,t}(\tilde{r}_i^{t+1})$ contains all messages ever sent under $\tilde{\mu}^{*,k}$ when the history of communications between the mediator and player i is $\tilde{r}_i^{t+1}$, $(\tilde{\sigma}^*, \tilde{\mu}^*, J, K)$ is valid.

---

[51] If $\tilde{r}_i^{t+1} \in R_i^{*t+1} \setminus Y_i^t$, then the mediator draws $r_{i,t}$ uniformly at random, and similarly for $\tilde{m}_{i,t}$ in what follows.

Next, for each $h_i^{R,t} \in J_i^{R,t}$ and $h^{R,t} \in H^{R,t}[h_i^{R,t}]|_{J,K}$, define

$$\beta_{i,t}\left(h^{R,t}|h_i^{R,t}\right) = \lim_{k\to\infty} \frac{\Pr^{\tilde{\sigma}^{*,k},\tilde{\mu}^{*,k}}\left(h^{R,t}\right)}{\Pr^{\tilde{\sigma}^{*,k},\tilde{\mu}^{*,k}}\left(h_i^{R,t}\right)}.$$

To see that the denominator of this expression is positive (so the quotient is well-defined), note that, since (i) under $\tilde{\mu}^k$ every sequence ζ^t occurs with positive probability and, given $\zeta_t = 1$, for each i the mediator sends each $m_{i,t} \in A_{i,t} \setminus D_{i,t}$ with positive probability at every history, and (ii) players tremble uniformly over actions under $\tilde{\sigma}_i^{*,k}$, it follows that $\Pr^{\tilde{\sigma}^{*,k},\tilde{\mu}^{*,k}}\left(h_i^{R,t}\right) > 0$ for all $h_i^{R,t} \in J_i^{R,t}$. Define $\beta_{i,t}\left(h^{A,t}|h_i^{A,t}\right)$ analogously. These beliefs are consistent by construction.

Sequential Rationality: Since the construction of $\left(\tilde{\sigma}^{*,k}, \tilde{\mu}^{*,k}\right)$ from $\left(\tilde{\sigma}^k, \tilde{\mu}^k\right)$ is the same as the construction of $\left(\tilde{\sigma}^k, \tilde{\phi}^k\right)$ from $\left(\bar{\sigma}^k, \phi\right)$ in the proof of Claim 1 of Proposition 4, sequential rationality of $\left(\tilde{\sigma}^*, \tilde{\mu}^*, J, K, \beta\right)$ follows from sequential rationality of $\left(\tilde{\sigma}, \tilde{\mu}, \tilde{J}, \tilde{K}, \tilde{\beta}\right)$.