Online Appendix to "Coordinating Public Good Provision by Mediated Communication"

Maija Halonen-Akatwijuka and In-Uck Park

May 31, 2020

We provide a formal analysis of the results reported in Section III.C of the main paper. After introducing several preliminary lemmas, we characterize the equilibrium described in Section III.C as an equilibrium that maximizes the social welfare among all equilibria of the communication game provided that the incentive compatibility (IC) is satisfied. Finally, we delineate the parameter values for which the IC is indeed satisfied.

1 Preliminary lemmas

Lemmas A1–A5 are from the proof of Proposition 2 in the main paper.

Lemma A1 *Suppose that* $a_i = (a_i^\ell, a_i^n, a_i^h)$ *is agent* i*'s equilibrium allocation under one pair of posterior beliefs and* $\mathrm{a}_i = (\mathrm{a}_i^\ell, \mathrm{a}_i^n, \mathrm{a}_i^h)$ *is that under another. If* $a_i^n < \mathrm{a}_i^n$*, then* $E(\mathrm{a}_i|\mu_i) - E(a_i|\mu_i) \leq \mathrm{a}_i^n - a_i^n$ *for any* μ_i*, with equality if and only if* a_i^ℓ *and* a_i^h *are interior solutions when* $\mu_i^\ell > 0$ *and* $\mu_i^h > 0$*, respectively.*

Lemma A2 *Let* $a_1 = (a_1^\ell, a_1^n, a_1^h)$ *and* $a_2 = (a_2^\ell, a_2^n, a_2^h)$ *be the equilibrium under posterior* (μ_1, μ_2) *such that* $\mathrm{Eb}(\mu_1) < \mathrm{Eb}(\mu_2)$*. Then,*
(a) $E(a_1|\mu_1) \geq a_1^n + \mathrm{Eb}(\mu_1)$ *with strict inequality if* $\mu_1^\ell > 0$ *and* $a_1^n < |\ell|$*.*
(b) $E(a_2|\mu_2) \leq a_2^n + \mathrm{Eb}(\mu_2)$ *with strict inequality if* $\mu_2^h > 0$ *and* $a_2^n + h > 1$*.*
(c) $E(a_1|\mu_1) - a_1^n = E(a_2|\mu_2) - a_2^n$*.*

Lemma A3 *Let* $\hat{a}_i^t = 1 + t - E(a_{-i}|\mu_{-i})$*, i.e., the unconstrained optimal allocation of agent* i *of* t*-type relative to an allocation vector* $a_{-i} \in [0,1]^3$ *of the other agent with a posterior belief* μ_{-i}*. Then, agent* i*'s utility from* $\hat{a}_i^t$ *is the same regardless of his type and decreases by* y^2 *if his allocation is* y *away from* $\hat{a}_i^t$*.*

Lemma A4 *If* $\mathrm{Eb}(\mu_1) < X < \mathrm{Eb}(\mu_2)$ *for some* $X < h + \ell$*, then the equilibrium value of* a_2^h *is a noninterior solution under the posterior* (μ_1, μ_2)*.*

Lemma A5 *If agent i of a type $t \in \{\ell, h\}$ always gets his unconstrained optimum after sending m_{ik} (even if irrelevant) but weakly prefers sending $m_{ik'}$ even if he sometimes gets less than the unconstrained optimum after $m_{ik'}$, then agent i of n-type strictly prefers sending $m_{ik'}$ to m_{ik}.*

We introduce a few more lemmas below.

Lemma B1 *If* $\mathrm{Eb}(\mu_1) < \mathrm{Eb}(\tilde{\mu}_1)$ *and* $\mathrm{supp}(\mu_1) = \mathrm{supp}(\tilde{\mu}_1) \neq \{h, n, \ell\}$, *then a_1^n is strictly higher in the continuation equilibrium after $(\tilde{\mu}_1, \mu_2)$ than after (μ_1, μ_2) for any μ_2.*

Proof. It is clear when $\mathrm{Eb}(\mu_1) < \mathrm{Eb}(\mu_2) < \mathrm{Eb}(\tilde{\mu}_1)$ from Lemma 2 of the main paper. When $\mathrm{Eb}(\mu_1) < \mathrm{Eb}(\tilde{\mu}_1) \leq \mathrm{Eb}(\mu_2)$ or $\mathrm{Eb}(\mu_2) \leq \mathrm{Eb}(\mu_1) < \mathrm{Eb}(\tilde{\mu}_1)$, if a_1^n is weakly lower after $(\tilde{\mu}_1, \mu_2)$ than after (μ_1, μ_2) then $E(a_1|\tilde{\mu}_1) < E(a_1|\mu_1) \Leftrightarrow 1 - E(a_1|\tilde{\mu}_1) > 1 - E(a_1|\mu_1)$ implying that a_2^n, thus $E(a_2|\mu_2)$, is strictly lower after (μ_1, μ_2) than after $(\tilde{\mu}_1, \mu_2)$, contradicting a_1^n being weakly higher after $(\tilde{\mu}_1, \mu_2)$ than after (μ_1, μ_2). ■

Lemma B2 *If* $\mathrm{Eb}(\mu_1) < \mathrm{Eb}(\mu_2)$ *with $\mu_1^h = 0$ and $\mu_2^\ell = 0$, then $a_1^n = 0$ and $a_2^n = 1$ in the continuation equilibrium after (μ_1, μ_2).*

Proof. If $a_1^n > 0$, then either $E(a_1) < a_1^n$ or $E(a_2) > a_2^n$ so that $a_1^n = 1 - E(a_2) \leq 1 - a_2^n = E(a_1) \leq a_1^n$ with at least one strict inequality, a contradiction. An analogous contradiction obtains when $a_2^n < 0$ as well. ■

Lemma B3 *Suppose a_2^n is higher in the continuation equilibrium after (μ_1, μ_2) than after $(\tilde{\mu}_1, \mu_2)$ where* $\max\{\mathrm{Eb}(\mu_1), \mathrm{Eb}(\tilde{\mu}_1)\} \leq \mathrm{Eb}(\mu_2)$. *Agent 1's utility is the same for n- and h-type and is higher after (μ_1, μ_2) than after $(\tilde{\mu}_1, \mu_2)$, strictly if* $\mathrm{Eb}(\mu_1) < \mathrm{Eb}(\mu_2)$ *and $\mu_2^h > 0$.*

Proof. By Lemmas 1 and 2 of the main paper, both after (μ_1, μ_2) and after $(\tilde{\mu}_1, \mu_2)$, agents 1-n and 1-h obtain unconstrained optimum (hence, identical utility) conditional on a_2 such that $a_2^\ell = a_2^n + \ell$ and $a_2^h = \min\{1, a_2^n + h\}$. Thus, their utility is higher when a_2 has a lower variance. The variance of a_2 is the same at (μ_1, μ_2) and at $(\tilde{\mu}_1, \mu_2)$ if $\mu_2^h = 0$ or $a_2^n + h \leq 1$ at (μ_1, μ_2); but is lower at (μ_1, μ_2) otherwise, which is the case when $\mathrm{Eb}(\mu_1) < \mathrm{Eb}(\mu_2)$ and $\mu_2^h > 0$. ■

At this point, we introduce some terminology to facilitate exposition. Since player i may send multiple messages that generate the same posterior μ_{ik}, we say "any (some) m_{ik}" to mean any (some) message that generate the posterior μ_{ik} when needed. Moreover, since multiple messages may have the same expected bias, we say "any (some) message with $\mathrm{Eb}(\mu_{ik})$." For brevity, we say "agent i-t" to mean agent $i \in \{1, 2\}$ of type $t \in \{h, n, \ell\}$. By "a_i^t after (m_{1j}, m_{2k})" we refer to agent i-t's allocation (to area A) in the continuation equilibrium following the message pair (m_{1j}, m_{2k}).

The following lemmas on comparing messages for agents prove useful.

Lemma B4 *Fix $m_{2\kappa}$ with $\mu_{2\kappa}^h > 0$. In the continuation equilibrium (a_1, a_2) after $(m_{1k}, m_{2\kappa})$ where* $\mathrm{Eb}(\mu_{1k}) < \mathrm{Eb}(\mu_{2\kappa})$, *the utility of agent 1-n is constant so long as*

$a_2^n \le 1-h$ and strictly increases in $a_2^n \ge 1-h$. The utility of agent 1-ℓ is the same as that of agent 1-n if $a_1^n + \ell \ge 0$ and is lower by $(a_1^n+\ell)^2$ if $a_1^n + \ell < 0$.

Proof. The first part is because agent 2's allocation is of a lower variance for higher $a_2^n \ge 1-h$. The second part is clear from Lemma A3. ■

Lemma B5 *Let (a_1, a_2) be the continuation equilibrium after $(m_{ik}, m_{-i\kappa})$ and let $(\tilde{a}_1, \tilde{a}_2)$ be that after $(m_{1k'}, m_{-i\kappa})$. The net benefit of agent i from sending $m_{ik'}$ rather than m_{ik} (conditional on $m_{-i\kappa}$) is lower for ℓ-type (resp. h-type) than for n-type if and only if $\tilde{a}_i^n + \ell \le \min\{0, a_i^n + \ell\}$ (resp. $\tilde{a}_i^n + h \ge \max\{1, a_i^n + h\}$), strictly when the inequality is strict; thus it is no lower if $\tilde{a}_i^\ell$ ($\tilde{a}_i^h$) is interior.*

Proof. Clear from Lemma A3 as a_i^n and $\tilde{a}_i^n$ are interior. ■

Lemma B6 *Suppose a message m_{ik} is optimal for agents i-n and i-ℓ and both of them derive unconstrained optimum after $(m_{ik}, m_{-ik'})$ for all $m_{-ik'} \in M_{-i}$. If another message m_{ij} is optimal for agent i-ℓ, then it is also optimal for i-n and both of them derive unconstrained optimum after $(m_{ij}, m_{-ik'})$ for all $m_{-ik'} \in M_{-i}$.*

Proof. Agents i-n and i-ℓ derive identical equilibrium payoffs from sending m_{ik}, as agent i-ℓ does from sending m_{ij}. If agent i-ℓ did not get unconstrained optimum from m_{ij} sometimes, agent i-n derives a higher overall payoff from m_{ij} (as he always gets the unconstrained optimum) which thus is higher than that from m_{ik}, contradicting optimality of m_{ik}. Thus, i-ℓ always gets unconstrained optimum from m_{ij}, hence the same overall payoff as agent i-n who should, therefore, find m_{ij} optimal. ■

2 Characterization of non-babbling equilibrium

For an arbitrary equilibrium of the communication game, let $M_i = \{m_{i1}, m_{i2}, \cdots, m_{iK_i}\}$ be the set of K_i messages sent by agent i with associated posteriors $\mu_{i1}, \mu_{i2}, \cdots, \mu_{iK_i}$ for $i \in \{1, 2\}$, labelled in such a way that

$$\mathrm{Eb}(\mu_{i1}) \le \mathrm{Eb}(\mu_{i2}) \le \cdots \le \mathrm{Eb}(\mu_{iK_i}) \quad \text{and} \quad \mathrm{Eb}(\mu_{11}) \le \mathrm{Eb}(\mu_{21}).$$

Since $\mathrm{Eb}(\mu_{11}) = \mathrm{Eb}(\mu_0)$ implies the babbling equilibrium as shown in the proof of Proposition 2, below we assume

$$(a0) \qquad \mathrm{Eb}(\mu_{11}) < \mathrm{Eb}(\mu_0) < \mathrm{Eb}(\mu_{1K_1}).$$

We start with the extreme possibility that $\mathrm{Eb}(\mu_{11}) = \ell \Leftrightarrow \mu_{11}^\ell = 1$. From Lemma 2, we have $a_2^n = 1$ and $a_1^n + \ell < 0$ at $(\mu_{11}, \mu_{2\kappa})$ if $\mathrm{Eb}(\mu_{2\kappa}) > \ell$. Consider any message m_{1k} used by agent 1-n so that $\mu_{1k}^n > 0$. Agents 1-ℓ and 1-n get the ideal payoff of 0 from both m_{11} and m_{1k} against any message $m_{2\kappa}$ with $\mathrm{Eb}(\mu_{2\kappa}) = \ell$ (Lemmas 1 and 2 in the main paper).

If there is a message $m_{2\kappa}$ such that $\mathrm{Eb}(\mu_{2\kappa}) > \ell$ and $a_2^n < 1$ at $(\mu_{1k}, \mu_{2\kappa})$, agent 1-$\ell$'s payoff would be lower than that of agent 1-n by a smaller margin at $(m_{1k}, m_{2\kappa})$

than at $(m_{11}, m_{2\kappa})$ by Lemma B4, thus would prefer m_{1k} strictly to m_{11} given that agent 1-n finds m_{1k} optimal, which would contradict $\mu_{11} = \ell$.

Thus, for every message m_{1k} used by agent 1-n, we have $a_2^n = 1$ at $(\mu_{1k}, \mu_{2\kappa})$ if $\text{Eb}(\mu_{2\kappa}) > \ell$. This implies that (i) $\mu_{1k}^h = 0$ for every message m_{1k} used by agent 1-n and (ii) $\mu_{2\kappa}^\ell = 0$ for every message $m_{2\kappa}$ such that $\text{Eb}(\mu_{2\kappa}) > \ell$, so that $\mu_{21} = \ell$. Then, by applying the same reasoning to agent 2, we deduce that (i') $\mu_{2k}^h = 0$ for every message m_{2k} used by agent 2-n and (ii') $\mu_{1\kappa}^\ell = 0$ for every message $m_{1\kappa}$ such that $\text{Eb}(\mu_{1\kappa}) > \ell$.

This means that both agents fully separate among the three types. Then, agent 1-ℓ would obtain his ideal allocation ($1 + \ell$ to A) when agent 2 is of ℓ-type but an allocation of 1 to A otherwise; by pretending to be h-type, instead, he would obtain ideal allocation unless agent 2 is of h-type in which case the allocation is at most 1 to A. Thus, agent 1-ℓ should pretend to be h-type, a contradiction.

The rest of the proof is on the case that $\text{Eb}(\mu_{11}) > \ell$, which we present in two parts depending on whether

[0] there is a message m_{11} such that a_1^h is interior (i.e., $a_1^n + h \leq 1$) after (m_{11}, m_{2k}) for all $m_{2k} \in M_2$.

Note that this condition holds if $\mu_{11}^h > 0$ by Lemmas 2–3.

<u>Part 1: The case that [0] holds.</u>

Fix a message m_{11} for which [0] holds. By Lemmas A4 and A5, agent 1-n strictly prefers sending any m_{1k} with $\text{Eb}(\mu_{1k}) > \text{Eb}(\mu_{21})$ and $\mu_{1k}^h > 0$ to sending m_{11}; and consequently, any m_{1k} with $\text{Eb}(\mu_{1k}) > \text{Eb}(\mu_{21})$ and $\mu_{1k}^n > 0$, thus all m_{1k} with $\text{Eb}(\mu_{1k}) > \text{Eb}(\mu_{21})$. This implies $\mu_{11}^n = 0$ and consequently, $\mu_{11}^\ell, \mu_{11}^h > 0$ since $\ell < \text{Eb}(\mu_{11}) < \text{Eb}(\mu_0)$. Thus, to find m_{11} optimal agent 1-ℓ must benefit less than agent 1-n by sending m_{1k} rather than m_{11} against at least one $m_{2\kappa} \in M_2$. That is,

[1] for every message m_{1k} with $\text{Eb}(\mu_{1k}) > \text{Eb}(\mu_{21})$, there is $m_{2\kappa} \in M_2$ such that agent 1's net benefit of sending m_{1k} rather than m_{11} conditional on $m_{2\kappa}$ is strictly lower for ℓ-type than for n-type, where $\text{Eb}(\mu_{2\kappa}) \geq \text{Eb}(\mu_{1k})$ by Lemma B5.

Fix any m_{1K_1} and $m_{2\kappa}$ as per [1] and consider the continuation equilibrium (a_1, a_2) after $(m_{11}, m_{2\kappa})$, where $a_1^n < a_2^n$ by Lemma 3. Let $(\tilde{a}_1, \tilde{a}_2)$ denote the equilibrium after $(m_{1K_1}, m_{2\kappa})$, where $\tilde{a}_1^n < a_1^n < |\ell|$ by Lemma B5 and [1]. From (6) in the main paper,

$$0 < a_1^n - \tilde{a}_1^n = E(\tilde{a}_2|\mu_{2\kappa}) - E(a_2|\mu_{2\kappa}) \;\leq\; \tilde{a}_2^n - a_2^n = E(a_1|\mu_{11}) - E(\tilde{a}_1|\mu_{1K_1})$$

where the second inequality is from Lemma A1 and consequently,

$$(a1) \qquad E(a_1|\mu_{11}) - a_1^n \;\geq\; E(\tilde{a}_1|\mu_{1K_1}) - \tilde{a}_1^n \;\geq\; \text{Eb}(\mu_{1K_1}) \;>\; \text{Eb}(\mu_0).$$

At this point, we show that $\text{Eb}(\mu_{21}) \leq \text{Eb}(\mu_{1k'})$ for all $m_{1k'} \neq m_{11}$. With a view to reaching a contradiction, suppose there is $m_{1k'} \neq m_{11}$ such that $\text{Eb}(\mu_{11}) \leq \text{Eb}(\mu_{1k'}) <$

$\text{Eb}(\mu_{21})$ and thus, $\mu^n_{1k'} = 0$ as well as $\mu^n_{11} = 0$ by Lemma A5. Then, either $\mu_{11} = \mu_{1k'}$ (in which case m_{11} and $m_{1k'}$ may be identified as they induce the same continuation equilibrium against every m_{2k}) or $\text{Eb}(\mu_{11}) < \text{Eb}(\mu_{1k'})$. In the latter case, a^n_2 is higher in the continuation equilibrium after (m_{11}, m_{2k}) than in that after $(m_{1k'}, m_{2k})$ for every m_{2k} by Lemma B1, hence both agent 1-h and 1-n would strictly prefer sending m_{11} to $m_{1k'}$ by Lemma B3. As this would contradict $\text{Eb}(\mu_{11}) < \text{Eb}(\mu_{1k'})$, we have established $\text{Eb}(\mu_{21}) \leq \text{Eb}(\mu_{1k'})$ for all $m_{1k'} \neq m_{11}$.

Let $(\breve{a}_1, \breve{a}_2)$ denote the continuation equilibrium after (m_{11}, m_{21}). If

[2] agent 2's net benefit of sending $m_{2\kappa}$ rather than m_{21} conditional on m_{11} is weakly larger for h-type than for n-type,

then $a^n_2 \leq \breve{a}^n_2$ must hold by Lemma B5. From (6) and Lemma A1, we deduce $0 \leq \breve{a}^n_2 - a^n_2 = E(a_1|\mu_{11}) - E(\breve{a}_1|\mu_{11}) \leq a^n_1 - \breve{a}^n_1 = E(\breve{a}_2|\mu_{21}) - E(a_2|\mu_{2\kappa})$ and thus,

$$(a2) \qquad E(a_2|\mu_{2\kappa}) - a^n_2 \leq E(\breve{a}_2|\mu_{21}) - \breve{a}^n_2 \leq \text{Eb}(\mu_{21}) \leq \text{Eb}(\mu_0).$$

But, $(a1)$ and $(a2)$ are incompatible by Lemma A2-(c).

We now show that [2] prevails when [0] holds, which establishes that $(a0)$ is unviable if [0] holds. Since agent 2-n gets the same payoff as 2-ℓ from $m_{2\kappa}$ but weakly higher payoff than 2-ℓ from m_{21}, and $\mu^n_{21} = 0$ would imply $\mu^\ell_{21} > 0$, it follows that agent 2-n weakly prefers m_{21} to $m_{2\kappa}$. If

[2a] agent 2-h's net benefit of sending $m_{2\kappa}$ rather than m_{21} is no higher that that of agent 2-n conditional on all $m_{1k'}$ such that $\text{Eb}(\mu_{21}) \leq \text{Eb}(\mu_{1k'}) \leq \text{Eb}(\mu_{1K_1})$,

therefore, negation of [2] would dictate that agent 2-h's overall utility is strictly lower with $m_{2\kappa}$ than with m_{21}. As this would contradict $\mu^h_{2\kappa} > 0$, which is implied by $\text{Eb}(\mu_{2\kappa}) > \text{Eb}(\mu_0)$, this verifies [2] provided that [2a] holds.

Thus, suppose [2a] fails, that is, agent 2's net benefit of sending $m_{2\kappa}$ rather than m_{21} is strictly larger for h-type than for n-type conditional on some m_{1k} with $\text{Eb}(\mu_{21}) \leq \text{Eb}(\mu_{1k})$. By Lemmas 1 and 2 of the main paper, this may happen only if

[3] $\text{Eb}(\mu_{21}) = \text{Eb}(\mu_{1k})$ for some $k \neq 1$ and $\mu^h_{21} = 0$ and $a^n_2 + h > 1$ at the interior solution after (m_{1k}, m_{21}) by a larger margin than that after $(m_{1k}, m_{2\kappa})$.

If $a^n_1 + h \leq 1$ after $(m_{1k}, m_{2k'})$ for all k', then $\mu^n_{1k} = 0 = \mu^n_{11}$ by Lemma A5 and thus, either $\mu_{1k} = \mu_{11}$ in which case m_{1k} and m_{11} may be identified without affecting equilibrium conditions,[1] or $\text{Eb}(\mu_{11}) < \text{Eb}(\mu_{1k})$ in which case agent 1 of both h and n-type would strictly prefer m_{11} to μ_{1k} by Lemma B3 because a^n_2 is higher at $(m_{11}, m_{2k'})$ than at $(m_{1k}, m_{2k'})$, a contradiction. On the other hand, if $a^n_1 + h > 1$ after $(\mu_{1k}, \mu_{2k'})$ for some k', as this is possible only if $\mu^h_{1k} = 0$ and $\text{Eb}(\mu_{1k}) = \text{Eb}(\mu_{2k'}) = \text{Eb}(\mu_{21})$, we may relabel $m_{2k'}$ as m_{21} and repeat the process of

[1]because both messages generate interior a^t_i for all t against any $m_{2k'}$ with the same bias, and generate the unique continuation equilibrium against any other $m_{2k'}$.

verifying [2]. Eventually, either [3] does not hold at some stage, thus verifying [2], or there is a cycle $\mu_{21}, \mu_{1k}, \mu_{2k'}, \mu_{1k''}, \cdots, \mu_{21}$, all of which assign probability 0 to $t_i = h$ and have the same bias, hence they are all identical posteriors.

In the latter case, with a view to reach a contradiction, suppose [2] fails, i.e.,

[2'] $a_2^n + h - 1$ is positive and strictly larger at $(m_{11}, m_{2\kappa})$ than at (m_{11}, m_{21}).

First, suppose $\mu_{21}^n \neq 1$. For [3], letting (a_1, a_2) and $(\hat{a}_1, \hat{a}_2)$ denote the interior solution at (m_{1k}, m_{21}) and the noninterior solution at $(m_{1k}, m_{2\kappa})$, resp., since both $a_1^n + \ell > 0$ and $\hat{a}_1^n + \ell > 0$ we have $\hat{a}_1^n - a_1^n = E(\hat{a}_1|\mu_{1k}) - E(a_1|\mu_{1k}) = a_2^n - \hat{a}_2^n = E(a_2|\mu_{21}) - E(\hat{a}_2|\mu_{2\kappa}) > 0$ where the last equality stems from $\hat{a}_1^n - a_1^n = E(a_2|\mu_{21}) - E(\hat{a}_2|\mu_{2\kappa})$. Solving this equation system we derive

$$(a3) \qquad a_2^n - \hat{a}_2^n \;=\; a_2^n - 1 + \frac{\ell(\mu^\ell - \mu_{2\kappa}^\ell)}{\mu_{2\kappa}^h} \;>\; 0.$$

For [2'], again letting (a_1, a_2) and $(\hat{a}_1, \hat{a}_2)$ denote the solution at (m_{11}, m_{21}) and the noninterior solution at $(m_{11}, m_{2\kappa})$, resp., we have $E(\hat{a}_1|\mu_{11}) - E(a_1|\mu_{11}) = a_2^n - \hat{a}_2^n < 0$ and $E(a_2|\mu_{21}) - E(\hat{a}_2|\mu_{2\kappa}) = \hat{a}_1^n - a_1^n < 0$. From this equation system we derive

$$a_2^n - \hat{a}_2^n \;=\; \frac{(1 - \mu_{11}^\ell)[\mu_{2\kappa}^h(a_2^n - 1) + \ell(\mu^\ell - \mu_{2\kappa}^\ell)]}{\mu_{11}^\ell + \mu_{2\kappa}^h(1 - \mu_{11}^\ell)} \;<\; 0$$

when $\mathrm{Eb}(\mu_{11}) < \mathrm{Eb}(\mu_{21})$, which is incompatible with $(a3)$ because a_2^n is higher at (m_{11}, m_{21}) than at (m_{1k}, m_{21}). If $\mathrm{Eb}(\mu_{11}) = \mathrm{Eb}(\mu_{21})$, we would have $\mu_{11}^\ell > \mu_{21}^\ell = \mu_{1k}^\ell$ since $\mu_{11}^n = \mu_{21}^h = \mu_{1k}^h = 0$ (recall $\mu_{11}^\ell < 1$) and agent 1-ℓ derives identical (unconstrained optimum) utility from m_{11} and m_{1k} against any $m_{2k'}$ with $\mathrm{Eb}(\mu_{2k'}) = \mathrm{Eb}(\mu_{11})$, but a higher utility from m_{1k} than from m_{11} against every $m_{2k'}$ such that $\mathrm{Eb}(\mu_{2k'}) > \mathrm{Eb}(\mu_{11})$ and $a_1^n + \ell < 0$ at $(m_{1k}, m_{2k'})$ by Lemma A5 because a_1^n is lower at $(m_{1k}, m_{2k'})$ than at $(m_{11}, m_{2k'})$. If such a $m_{2k'}$ exists, agent 1-ℓ strictly prefers m_{1k} to m_{11}, a contradiction. Otherwise, agent 1-ℓ derives the same utility as agent 1-n from sending m_{11} and also from sending m_{1k}, contradicting m_{11} being suboptimal for agent 1-n. This completes verifying [2] when $\mu_{21}^n \neq 1$.

Now suppose $\mu_{21}^n = 1$. For any m_{1K} with $\mathrm{Eb}(\mu_{1K}) > 0$, whence $\mu_{1K}^h > 0$, by [1] agent 1-ℓ benefits strictly less than agent 1-n by sending m_{1K} instead of m_{11} against some $m_{2k'}$, termed "moderator (message)", for which a_1^n should be lower than $-\ell$ at $(m_{11}, m_{2k'})$ and decrease at $(m_{1K}, m_{2k'})$. It is routinely verified that the solution value of a_1^n decreases as such only if

$$(a4) \qquad \mu_{1K}^h \;\leq\; \bar{\mu}_{1K}^h = \frac{\ell\mu_{2k'}^\ell(1 - \mu_{11}^h - \mu_{1K}^\ell) + h\mu_{11}^h(\mu_{2k'}^h + \mu_{1K}^\ell - \mu_{2k'}^h\mu_{1K}^\ell)}{h(1 - \mu_{11}^h(1 - \mu_{2k'}^h))},$$

thus every equilibrium posterior μ_{1K} with $\mathrm{Eb}(\mu_{1K}) > 0$ must satisfy $(a4)$.

However, since μ_0 must be in the interior of the convex hull spanned by agent 1's equilibrium posteriors, there exists an equilibrium posterior μ_{1K} with $\mathrm{Eb}(\mu_{1K}) \geq 0$, on the opposite side from $\mu_{1k} = (0,1,0)$ of the hyperplane spanned by μ_0 and μ_{11} in the posterior simplex, which is a condition characterized as

$$\mu_{1K}^h \;\geq\; \underline{\mu}_{1K}^h \;=\; \frac{\mu_0^h(1-\mu_{11}^h) - \mu_0^\ell \mu_{11}^h - (\mu_0^h - \mu_{11}^h)\mu_{1K}^\ell}{1-\mu_0^\ell - \mu_{11}^h} \;=\; \frac{\mu_0^h(1-2\mu_{11}^h - \mu_{1K}^\ell) + \mu_{11}^h \mu_{1K}^\ell}{1-\mu_0^h - \mu_{11}^h}.$$

Consequently,

$$\underline{\mu}_{1K}^h - \bar{\mu}_{1K}^h = \frac{\big[h(\mu_0^h - 2\mu_0^h\mu_{11}^h) - h(1-2\mu_0^h)\mu_{11}^h\mu_{2k'}^h) - \ell(1-\mu_0^h-\mu_{11}^h)\mu_{2k'}^\ell\big](1-\mu_{11}^h - \mu_{1K}^\ell)}{h(1-\mu_0^h-\mu_{11}^h)(1-\mu_{11}^h(1-\mu_{2k'}^h))}.$$

Note that the expression in the bracket is positive if $\mu_{2k'}^\ell \geq \mu_{2k'}^h$ because it increases in $\mu_{2k'}^\ell$ and decreases in μ_{11}^h and obtains $\frac{(h+\ell)\mu_0^h(h-\ell\mu_{2k'}^h)}{h-\ell} > 0$ at $\mu_{2k'}^\ell = \mu_{2k'}^h$ and $\mu_{11}^h = \frac{-\ell}{h-\ell}$ ($\text{Eb}(\mu_{11}) \leq 0$ implies $\mu_{11}^h \leq \frac{-\ell}{h-\ell}$). Since $\text{Eb}(\mu_{11}) < \text{Eb}(\mu_{1K})$ implies $\mu_{11}^h + \mu_{1K}^\ell < 1$ and $\mu_{11}^h < \frac{\mu_0^h}{\mu_0^h + \mu_0^\ell} = 1/2$ implies $\mu_{11}^h + \mu_0^h < 1$, it follows that $\underline{\mu}_{1K}^h > \bar{\mu}_{1K}^h$ if $\mu_{2k'}^\ell \geq \mu_{2k'}^h$. Therefore, for any message m_{1K} with $\text{Eb}(\mu_{1K}) > 0$, the moderator $m_{2k'}$ must have $\mu_{2k'}^\ell < \mu_{2k'}^h$. Moreover, no equilibrium posterior $\mu_{1K} \notin \{\mu_{11}, \mu_{1k}\}$ with $\text{Eb}(\mu_{1K}) = 0$ exists because any such posterior would have $\mu_{1K}^n = 0$ by Lemma A5 and thus would be dominated by m_{11} for agent 1-h by Lemma B1.

Consequently, since μ_0 is the mean of posteriors, there must be a message, say m_{2j}, such that $\text{Eb}(\mu_{2j}) \geq 0$ and $\mu_{2j}^\ell > \mu_{2j}^h$, and for every μ_{1K} with $\text{Eb}(\mu_{1K}) > 0$,

[1'] agent 1's net benefit of sending m_{1K} rather than m_{11} conditional on m_{2j} is no lower for ℓ-type than for n-type.

Suppose that agent 2-n derives a strictly higher utility with m_{21} than m_{2j} against m_{11}, i.e., a_1^n is lower at (m_{11}, m_{21}) than at (m_{11}, m_{2j}). Since $E(a_1|\mu_{11}) = a_1^n + \mu_{11}^h h - (1-\mu_{11}^h)a_1^n$, as μ_{11} is changed to μ_{1K} the negative effect of $(\mu_{1K}^h - \mu_{11}^h)h$ is the same but the positive effect of $(\mu_{1K}^\ell - \mu_{11}^\ell)a_1^n$ is larger when a_1^n is larger, i.e., at (m_{11}, m_{2j}). This means that a_1^n remains lower at (m_{11}, m_{21}) than at (m_{11}, m_{2j}) when μ_{11} is changed to μ_{1K} with $\text{Eb}(\mu_{1K}) > 0$, thus agent 2-n's utility remains higher with m_{21} than m_{2j} against m_{1K}. Since his utility is same at 0 with both m_{21} and m_{2j} against m_{1k}, the message m_{2j} would be suboptimal for agent 2-n. As this is impossible as explained below, it must be the case that he derives no lower utility with m_{2j} than m_{21} against m_{11}. This condition is calculated to be

$$\mu_{2j}^h > \frac{-\ell}{h}\Big(\frac{1-\mu_{11}^h}{\mu_{11}^h}\Big)\mu_{2j}^\ell > \frac{-\ell}{h}\Big(\frac{1-\mu_{11}^h}{\mu_{11}^h}\Big)\mu_{2j}^\ell\Big|_{\mu_{11}^h = -\ell/(h-\ell)} = \mu_{2j}^\ell,$$

contradicting $\mu_{2j}^\ell > \mu_{2j}^h$.

Finally, if m_{2j} is suboptimal for 2-n so that $\text{supp}(\mu_{2j}) = \{h, \ell\}$, then $\text{supp}(\mu_{2\kappa}) = \{h, \ell\}$ is not viable because that would imply $m_{2\kappa}$ being suboptimal for agent 2-h by Lemma B1, nor is $\text{supp}(\mu_{2\kappa}) \ni n$ because that would imply 2-ℓ's utility from m_{2j} being no lower than 2-n's utility from $m_{2\kappa}$ by Lemma A5, contradicting suboptimality of m_{2j}

for 2-n, and nor is $\text{supp}(\mu_{2j}) = \{h\}$ because that would imply $a_1^n = 0$ at $(m_{11}, m_{2\kappa})$, violating [1].

Part 2: The case that [0] fails.

Since $\mu_{11}^h = 0$ and $\text{Eb}(\mu_{11}) = \text{Eb}(\mu_{21})$ by Lemmas 1 and 2 (of the main paper) in this case, we may assume that [0] fails for both agents as they can be relabelled otherwise, so that

[0'] $\mu_{11} = \mu_{21}$ and for each i, $\mu_{i1}^h = 0$ and $a_i^n + h > 1$ at the interior solution after (m_{i1}, m_{-ik}) for some m_{-ik} with $\text{Eb}(\mu_{-ik}) = \text{Eb}(\mu_{i1}) \leq 0$.

This means that μ_{i1} is generated by multiple equilibrium messages for each i. In addition, no other equilibrium posterior $\mu_{ik} \neq \mu_{i1}$ has the same bias as $\text{Eb}(\mu_{i1})$ because $\mu_{ik}^h > 0$ would ensue if it did, confirming [0].

Recall $\ell < \text{Eb}(\mu_{11}) = \text{Eb}(\mu_{21})$. Suppose an equilibrium posterior say for agent 2, $\mu_{2k} \neq \mu_{21}$, has $\mu_{2k}^\ell = 0$. Since $\text{Eb}(\mu_{21}) < \text{Eb}(\mu_{2k})$, agent 2-$h$ derives no higher utility than n-type from sending m_{2k} conditional on every m_{1k}, and h^2 less when $\text{Eb}(\mu_{1k}) = \text{Eb}(\mu_{11})$ by Lemmas A3 and B2; but by sending m_{21}, he derives the same utility as n-type when $\text{Eb}(\mu_{1k}) > \text{Eb}(\mu_{11})$ and less by an amount smaller than h^2 when $\text{Eb}(\mu_{1k}) = \text{Eb}(\mu_{11})$ by Lemma 1 of the main paper. Since n-type derives no lower utility from sending m_{21} than m_{2k} (because $\mu_{21}^n > 0$), this means that h-type strictly prefers m_{21} to m_{2k}, a contradiction unless $\mu_{2k}^n = 1$, leading to the following observation.

[4] $\mu_{ik}^\ell > 0$ unless $\mu_{ik}^n = 1$ for all equilibrium posteriors $\mu_{ik} \neq \mu_{i1}$ for each i.

With a view to reaching a contradiction, suppose $\mu_{i1}^\ell > 0$. If $a_i^n + \ell \geq 0$ at (m_{i1}, m_{-ik}) for all $m_{-ik} \in M_{-i}$ for both $i = 1, 2$, (which implicitly assumes $\mu_{ik}^n \neq 1$ for all μ_{ik}), then both agents i-ℓ and i-n derive unconstrained optimum from m_{i1} against all messages, hence so should both agents from all other messages m_{ik} as well due to Lemma B6 since $\mu_{ik}^\ell > 0$ by [4]. Then, agent i-n and i-ℓ derive higher utility from m_{i1} than from $m_{ik} \neq m_{i1}$ against all $m_{-ik} \neq m_{-i1}$ and the same utility against m_{-i1}, contradicting [4].

Therefore, for some agent, say 1, $a_1^n + \ell < 0$ at (m_{11}, m_{2k}) for some $m_{2k} \neq m_{21}$. If $\mu_{2k}^h = 0$ then $E(a_2|\mu_{2k}) = a_2^n + \text{Eb}(\mu_{2k})$ at (m_{11}, m_{2k}), thus a_2^n is higher and a_1^n is lower at $(m_{11}, m_{2k'})$ for any $m_{2k'}$ with $\text{Eb}(\mu_{2k'}) > \text{Eb}(\mu_{2k})$. Hence, we may assume that $\mu_{2k}^h > 0$. Since $a_1^n + \ell \geq 0$ at any interior solution after (m_{11}, m_{21}), it follows that the net benefit of agent 2 from sending m_{21} rather than m_{2k} is strictly higher for h-type than for n-type against m_{11}; moreover, it is weakly higher for h-type against all messages $m_{1k} \neq m_{11}$. Since n-type weakly prefers m_{21} to m_{2k}, we have reached a contradiction to $\mu_{2k}^h > 0$ as desired.

This leaves us to examine the case that $\mu_{i1}^n = 1$, whence $\text{Eb}(\mu_{i1}) = 0$ and $\mu_{ik}^h > 0$ (as well as $\mu_{ik}^\ell > 0$ by [4]) for all $k \neq 1$. A key step for this case is to establish that

[5] for each $i = 1, 2$, $\{\mu_{i1}, \mu_{iK_i}\}$ is the set of equilibrium posteriors.

Suppose to the contrary that there are more posteriors for, say agent 1, and let μ_{1k} denote the one with the smallest bias, so that $0 = \mathrm{Eb}(\mu_{11}) < \mathrm{Eb}(\mu_{1k}) \leq ... \leq \mathrm{Eb}(\mu_{1K_1})$ where the first strict inequality is implied by negation of [0]. We may also assume that $\mathrm{Eb}(\mu_{2K_2}) \leq \mathrm{Eb}(\mu_{1K_1})$ because if agent 2 had only two posteriors and $\mathrm{Eb}(\mu_{2K_2}) > \mathrm{Eb}(\mu_{1K_1})$ then since agent 1-ℓ derives the same utility of 0 from every m_{1k} with $\mathrm{Eb}(\mu_{1k}) > 0$ against m_{21}, he (thus, agent 1 of all types) must derive the same utility against μ_{2K_2} as well, which means the same continuation equilibrium (a_1, a_2) at every (m_{1k}, m_{2K_2}) with $\mathrm{Eb}(\mu_{1k}) > 0$; but this would imply identical $\mathrm{Eb}(\mu_{1k})$ if $a_1^n + \ell \geq 0$ or identical values of $\mu_{1k}^h h - \mu_{1k}^\ell a_1^n$ otherwise, neither of which is compatible with agent 1-h deriving the same utility from all such m_{1k} against m_{21} (which must be the case given $\mu_{1k}^h > 0$) unless μ_{1k} are all identical (because $a_1^n = 1 + \ell\mu_{1k}^\ell/\mu_{1k}^h$ at (m_{1k}, m_{21})).

Since both agents 1-n and 1-ℓ obtain unconstrained optimum from m_{1K_1} against all $m_{2k'}$, the net benefit of agent 1 from sending m_{1K_1} rather than m_{1k} is lower for h-type than for both n- and ℓ-type against all $m_{2k'}$ such that $\mathrm{Eb}(\mu_{1k}) \leq \mathrm{Eb}(\mu_{2k'}) \leq \mathrm{Eb}(\mu_{1K_1})$. If $\mu_{1k}^n = 0 = \mu_{1K_1}^n$, the net benefit is strictly lower for h-type against $m_{2k'}$ such that $\mathrm{Eb}(\mu_{2k'}) < \mathrm{Eb}(\mu_{1k})$ by Lemma B1, thus both n- and ℓ-type would strictly prefer m_{1K_1} to m_{1k}, contradicting $\mu_{1k}^\ell > 0$. Hence, either $\mu_{1k}^n > 0$ or $\mu_{1K_1}^n > 0$ must hold.

If m_{1K_1} is optimal for agent 1-n (implied by $\mu_{1K_1}^n > 0$), Lemma B6 dictates that both agents 1-n and 1-ℓ derive unconstrained optimum from every message $m_{1k} \neq m_{11}$ and find them optimal. However, m_{1k} cannot be optimal for agent 1-n as it is dominated by m_{11} due to Lemma B7 below, unless there is m_{2j} such that $\mathrm{Eb}(\mu_{11}) < \mathrm{Eb}(\mu_{2j}) < \mathrm{Eb}(\mu_{1k})$ and against which agent 1-n derives a higher utility with m_{1k} than with m_{11}.

Lemma B7 *Consider μ with $\mu^n = 1$ and $0 = \mathrm{Eb}(\mu) < \mathrm{Eb}(\tilde{\mu}) \leq \mathrm{Eb}(\mu_2)$. Let (a_1, a_2) be the noninterior solution at $(\mu_1, \mu_2) = (\mu, \mu_2)$ and let $(\tilde{a}_1, \tilde{a}_2)$ be the solution at $(\mu_1, \mu_2) = (\tilde{\mu}, \mu_2)$. Then, $a_2^n > \tilde{a}_2^n$.*

Proof. Given $\mu_2^h > 0$ we have $a_1^n = 1 - E(a_2|\mu_2)$ so that $a_2^n = 1 - a_1^n = E(a_2|\mu_2) > 1-h$. If $\tilde{a}_2^h$ is an interior solution, i.e, $\tilde{a}_2^n + h \leq 1$, then $a_2^n > \tilde{a}_2^n$. Otherwise, i.e., $\tilde{a}_2^n + h > 1$, then $\tilde{a}_1^n = 1 - E(\tilde{a}_2|\mu_2)$. For $a_2^n \leq \tilde{a}_2^n$ we would need $E(a_1|\mu) = a_1^n \geq E(\tilde{a}_1|\tilde{\mu}) > \tilde{a}_1^n$ where the last inequality stems from $\mathrm{Eb}(\tilde{\mu}) > 0$, so that $E(a_1|\mu) - E(\tilde{a}_1|\tilde{\mu}) < a_1^n - \tilde{a}_1^n$; however, $a_2^n \leq \tilde{a}_2^n$ implies $E(\tilde{a}_2|\mu_2) - E(a_2|\mu_2) \leq \tilde{a}_2^n - a_2^n$ by Lemma A2 so that $a_1^n - \tilde{a}_1^n = E(\tilde{a}_2|\mu_2) - E(a_2|\mu_2) \leq \tilde{a}_2^n - a_2^n = E(a_1|\mu) - E(\tilde{a}_1|\tilde{\mu})$, a contradiction. ■

Thus, assume that m_{1k} is optimal for agent 1-n and there are messages for agent 2, denoted by m_{2j}, such that $0 = \mathrm{Eb}(\mu_{11}) < \mathrm{Eb}(\mu_{2j}) < \mathrm{Eb}(\mu_{1k})$. We may also suppose that $\mathrm{Eb}(\mu_{2K_2}) < \mathrm{Eb}(\mu_{1K_1})$ because if $\mathrm{Eb}(\mu_{2K_2}) = \mathrm{Eb}(\mu_{1K_1})$ we would have encountered a contradiction by relabelling the agents. Moreover, m_{2j} is suboptimal for n-type (thus, $\mu_{2j}^n = 0$) by Lemma B7. If $\mu_{2k'}^n = 0$ for some message $m_{2k'}$ with $\mathrm{Eb}(\mu_{2j}) < \mathrm{Eb}(\mu_{2k'})$, then the net benefit of sending m_{2j} rather than $m_{2k'}$ would be strictly higher for agent 2-h than for 2-ℓ by Lemma B1, contradicting both agents sending both messages. Therefore, we deduce that $\mu_{2k'}^n > 0$ if $\mathrm{Eb}(\mu_{2j}) < \mathrm{Eb}(\mu_{2k'}) \leq$

$\text{Eb}(\mu_{2K_2})$. Consequently, if $\mu^n_{1K_1} > 0$ then agent 2-n would derive a higher utility with m_{21} than with m_{2K_2} against any $m_{1k'} \neq m_{11}$ with $\text{Eb}(\mu_{1k'}) \leq \text{Eb}(\mu_{2K_2})$ because a_1 is interior at $(\mu_{1k'}, \mu_{2K_2})$ by Lemma B6, and against any $m_{1k'}$ with $\text{Eb}(\mu_{1k'}) > \text{Eb}(\mu_{2K_2})$ because a^n_1 is lower (a^n_2 is higher) after $(m_{21}, m_{1k'})$ by Lemma B7, contradicting $\mu^n_{2K_2} > 0$. Thus, $\mu^n_{1K_1} = 0$ must hold.

Then, agent 2 of ℓ-type does equally well as n-type with m_{2K_2} (thus better than h-type) against all $m_{1k''}$ such that $\text{Eb}(\mu_{2j}) < \text{Eb}(\mu_{1k''}) \leq \text{Eb}(\mu_{2K_2})$, whereas h-type does better than ℓ-type with m_{2j} against all such $m_{1k''}$'s. Against m_{11}, both ℓ-type and n-type obtain unconstrained optimum, while h-type doesn't but the margin is smaller with m_{2j} than m_{2K_2}.

Against the message m_{1K_1} (unique because $\mu^n_{1K1} = 0$), either $E(a_2|\mu_{2j}) \leq E(a_2|\mu_{2K_2})$ or $E(a_2|\mu_{2j}) > E(a_2|\mu_{2K_2})$. If the former, agent 2 of h-type does better with m_{2j} than with m_{2K_2} relative to ℓ-type (since a^n_2 is lower with m_{2j}), contradicting $\mu^h_{2K_2} > 0$.

In the latter case, $E(a_2|\mu_{2j}) > E(a_2|\mu_{2K_2})$ against m_{1K_1} requires that $a^n_2 + \ell < 0$ at the noninterior solution following (m_{1K_1}, m_{2j}), which is calculated as

$$\frac{(1-\mu^h_{1K_1})(h\mu^h_{2j} - \ell)}{1-(1-\mu^h_{1K_1})\mu^h_{2j}} + \ell = \frac{(1-\mu^h_{1K_1})(h\mu^h_{2j} - \ell(1+\mu^h_{2j})) + \ell}{1-(1-\mu^h_{1K_1})\mu^h_{2j}} \; < \; 0$$

$$(a5) \qquad \iff \quad \mu^h_{1K_1} \; > \; \underline{\mu}^h_{1K_1} \; = \; \frac{(h-\ell)\mu^h_{2j}}{(h-\ell)\mu^h_{2j} - \ell} \; > \; \left.\frac{(h-\ell)\mu^h_{2j}}{(h-\ell)\mu^h_{2j} - \ell}\right|_{\mu^h_{2j} = \frac{-\ell}{h-\ell}} \; = \frac{1}{2}$$

where the last inequality follows because $\text{Eb}(\mu_{2j}) = \mu^h_{2j}h + (1-\mu^h_{2j})\ell > 0 \Leftrightarrow \mu^h_{2j} > \frac{-\ell}{h-\ell}$ and $\underline{\mu}^h_{1K_1}$ increases in μ_{2j}. Thus, there must exist a message, say m_{1K}, such that $\text{Eb}(\mu_{1k}) \leq \text{Eb}(\mu_{1K}) < \text{Eb}(m_{1K_1})$ and $\mu^h_{1K} < \mu^\ell_{1K}$ because μ_0 is the mean of all posteriors.

By the standard argument, agent 1-h does strictly better with m_{1K} than with m_{1K_1} relative to agent 1-ℓ against all $m_{2k'}$ between them (i.e., m_{1K} and m_{1K_1}); and so does he against m_{21} because it is routinely verified that $a^n_1 + h$ is lower in the noninterior solution after (m_{1K}, m_{21}) than that after (m_{1K_1}, m_{21}) if and only if $\mu^h_{1K_1} > \frac{\mu^h_{1K}}{\mu^h_{1K} + \mu^\ell_{1K}}$ which is the case given $(a5)$ and $\mu^h_{1K} < \mu^\ell_{1K}$. The same also holds against m_{2j} because $a^n_1 + h$ is lower in the noninterior solution after (m_{1K}, m_{2j}) than that after (m_{1K_1}, m_{2j}) as explained below: it is straightforward if $a^n_2 + \ell > 0$ at (m_{1K}, m_{2j}) since $E(a_2|\mu_{2j}) > E(a_2|\mu_{2K_2})$ implies $a^n_2 + \ell < 0$ at (m_{1K_1}, m_{2j}); if $a^n_2 + \ell < 0$ at (m_{1K}, m_{2j}), on the other hand, it is routinely verified to be the case if and only if

$$(a6) \qquad \mu^h_{1K_1} \; > \; \hat{\mu}^h_{1K_1} = \frac{(h-\ell)\mu^h_{1K}\mu^h_{2j} - \ell(1-\mu^\ell_{1K})(1-\mu^h_{2j})}{h\mu^h_{2j} - \ell((\mu^h_{1K} + \mu^\ell_{1k} - 1)\mu^h_{2j} + 1)},$$

which holds because by subtracting the RHS of $(a6)$ from $\underline{\mu}^h_{1K_1}$ we get

$$\frac{(h\mu^h_{2j} - \ell)(h(1-\mu^h_{1K})\mu^h_{2j} + \ell(1-\mu^\ell_{1K} - \mu^h_{2j} + \mu^h_{1K}\mu^h_{2j}))}{(\mu^h_{2j}(h-\ell) - \ell)(\mu^h_{2j}(h+\ell) - (1+(\mu^h_{1K} + \mu^\ell_{1K})\mu^h_{2j})\ell)} \; > \; 0$$

given that $h(1-\mu_{1K}^h)\mu_{2j}^h + \ell(1-\mu_{1K}^\ell - \mu_{2j}^h + \mu_{1K}^h\mu_{2j}^h)$ increases both in μ_{2j}^h and μ_{1K}^ℓ and assumes 0 at $\mu_{2j}^h = \frac{-\ell}{h-\ell}$ and $\mu_{1K}^\ell = \mu_{1k}^h$.

Therefore, there must exist a message, say m_{2J}, such that $\mathrm{Eb}(\mu_{1k}) \le \mathrm{Eb}(\mu_{2J}) < \mathrm{Eb}(\mu_{1K})$ and against which agent 1-h does strictly worse with m_{1K} than with m_{1K_1}. This condition, obtained by an analogous calculation for $(a6)$, is

$$(a7) \qquad \mu_{1K_1}^h \;<\; \frac{(h-\ell)\mu_{1K}^h\mu_{2J}^h - \ell(\mu_{1K}^h + (1-\mu_{1K}^\ell)\mu_{2J}^\ell)}{h\mu_{2J}^h - \ell((\mu_{1K}^h + \mu_{1K}^\ell)(1-\mu_{2J}^\ell) + \mu_{2J}^\ell)}.$$

At the same time, agent 2-ℓ does better with m_{2K_2} than with m_{2J} relative to agent 2-n (and 2-h) against all $m_{1k'}$ such that $\mathrm{Eb}(\mu_{2J}) \le \mathrm{Eb}(\mu_{1k'}) \le \mathrm{Eb}(\mu_{2K_2})$ by the standard argument; against all $m_{1k'}$ with $\mathrm{Eb}(\mu_{1k'}) < \mathrm{Eb}(\mu_{2J})$ both agents 2-n and 2-ℓ obtain unconstrained optimum. Therefore, agent 2-ℓ should do worse, relative to 2-n, with m_{2K_2} than with m_{2J} against m_{1K_1}. This requires that $a_2^n + \ell < 0$ at the noninterior solution after (m_{1K_1}, m_{2J}), which is verified to be the case if and only if $\mu_{1K_1}^h > \frac{h\mu_{2J}^h - \ell(1-\mu_{2J}^\ell)}{h\mu_{2J}^h - \ell(2-\mu_{2J}^\ell)}$, but this is incompatible with $(a7)$ because by subtracting the RHS of $(a7)$ from this lower bound we get

$$\frac{\left[h(1-\mu_{1K}^h)\mu_{2J}^h - \ell(\mu_{1K}^\ell - \mu_{1K}^h(1-\mu_{2J}^\ell) + \mu_{2J}^\ell)\right](h\mu_{2J}^h - \ell)}{(h\mu_{2J}^h - \ell((\mu_{1K}^h + \mu_{1K}^\ell)(1-\mu_{2J}^\ell) + \mu_{2J}^\ell))(h\mu_{2J}^h - \ell(2-\mu_{2J}^\ell))} \;>\; 0$$

where the inequality follows since the expression in the bracket decreases in μ_{1K}^h and assumes $(1-\mu_{1K}^\ell)(h\mu_{2J}^h + \ell\mu_{2J}^\ell) > 0$ when $\mu_{1K}^h = \mu_{1K}^\ell$.

This verifies [5]. Consequently, both players may send two messages one of which is $\mu_{i1} = (0,1,0) =: \mu_n$ and the other is $\mu_{iK_i} = (1/2, 0, 1/2) =: \mu_b$. There are a continuum of continuation equilibria after (μ_b, μ_b) but they all are interior and generate the same payoff for all types. The continuation equilibrium is unique after (μ_b, μ_n) and after (μ_n, μ_b). Jointly controlled lotteries (JCL) are possible via randomization of equilibrium allocations after pairs of messages that generate posterior pair (μ_n, μ_n). Various specifications of JCL correspond to different PBE so long as the incentive compatibility is satisfied for both players to send suitable messages depending on their types. The welfare is the same across all these PBE's because the allocation is the same after each possible type realization. However, the incentive compatibility is satisfied most widely in the equilibrium described in Section III.C of the main paper because, given the symmetry between the two agents, the equal randomization of specialization by the agents in the two areas minimizes the incentive of either agent of biased types to mimic the neutral type.

3 Incentive Compatibility

Lastly, we determine when the IC is satisfied for the equilibrium in Section III.C of the main paper. Consider an ℓ-type agent, say i, sending the equilibrium message b.

If his opponent is of n-type, they play the unique noninterior continuation equilibrium for $\mu_i = \mu_b$ and $\mu_{-i} = \mu_n$ as in Lemma 2 of the main paper; otherwise they play an interior equilibrium for $(\mu_1, \mu_2) = (\mu_b, \mu_b)$ as in Lemma 1. Hence, an ℓ-type agent's expected payoff from sending b is calculated as

$$\frac{1}{3}\left\{-\left[(1+\ell)-\left(1-\frac{h}{2}+\frac{3\ell}{2}\right)\right]^2-\left[(1+\ell)-\left(1+\frac{h}{2}+\frac{\ell}{2}\right)\right]^2\right\}=\frac{-(h-\ell)^2}{6}.$$

Next, suppose an ℓ-type agent deviates by sending n or n'. If his opponent is of n-type, he achieves his ideal allocation $1+\ell$ with probability $1/2$ but ends up with a total allocation of 1 to A with the other probability $1/2$; otherwise, he gets a payoff for the (irrelevant) agent i-ℓ in the unique noninterior continuation equilibrium for $\mu_i = \mu_n$ and $\mu_{-i} = \mu_b$. Thus, his expected payoff from deviating is

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

Therefore, IC holds for an ℓ-type agent if and only if $\ell \leq -(1+\sqrt{5})h/4$.

Analogously, we calculate the expected payoffs of an h-type agent from sending b and sending n or n', respectively, as

$$\frac{-2(h+\ell)^2-(h-\ell)^2}{6} \quad \text{and} \quad \frac{-h^2-4\ell^2}{6}$$

so that IC holds for h-type if and only if $\ell \leq (1-\sqrt{3})h$, which holds whenever the IC holds for ℓ-type because $-(1+\sqrt{5})/4 < 1-\sqrt{3}$. Moreover, the IC always holds for an n-type agent as it is routinely calulated that his expected payoff from sending n or n' is $-2\ell^2/3$ while that from sending b is $-(h-\ell)^2/6 < -2\ell^2/3$.

Therefore, the equilibrium described in Section III.C constitutes the maximim welfare equilibrium of the communication game if and only if $\ell \leq -(1+\sqrt{5})h/4$.