Coordinating Public Good Provision by Mediated Communication

By Maija Halonen-Akatwijuka and In-Uck Park*

We examine a setup where two agents allocate a fixed budget between public goods in two areas. The agents may be biased to one area which is their private information. Without communication the funds are allocated inefficiently resulting in gaps and duplication in public good provision. Direct communication between the agents is ineffective and cannot resolve the coordination failure even when the potential biases are negligible. Coordination can be improved by a mediator who filters the information communicated by the agents. Our results can throw light on how to improve coordination of humanitarian aid by appropriately designed information management system. (JEL D82, D83, H41, H84, L31)

Banerjee (2007) starts his book *Making Aid Work* with an episode from the 2005 earthquake in Pakistan. When international organizations and NGOs rushed in to help, a group of economists got concerned about how the aid would get to the right people. As no one was keeping track of where the aid had been delivered, some villages received many consignments while others had no aid. The economists figured that coordination would be improved by a website to which everyone could report the location and amount of aid sent. Based on this information the humanitarian organizations could decide where the next consignments should go. Disaster management system Risepak was swiftly developed to achieve this goal.[1] However, the humanitarian organizations were largely not willing to share their information and Risepak did not reach a critical mass. Such unwillingness to share information is often reported in the humanitarian sector and has been attributed, e.g., as the cause of the coordination failure of the humanitarian response after the 2010 earthquake in Haiti (UN Inter-Agency Standing Committee 2010; Altay and Labonte 2014).

We approach this allocation problem from the point of sensitivities in informa-

* Halonen-Akatwijuka: University of Bristol, School of Economics, Priory Road Complex, Bristol BS8 1TU, U.K. (email: Maija.Halonen@bristol.ac.uk); Park: Sungkyunkwan University, College of Economics, Seoul, Korea, and University of Bristol, School of Economics, Priory Road Complex, Bristol BS8 1TU, U.K. (email: I.Park@bristol.ac.uk). We thank the editors Johannes Hörner and Leeat Yariv and two anonymous referees for valuable comments. We have also benefited from responses of audiences at the 2016 SIOE meeting in Paris, the 2017 NBER Organizational Economics Meeting, ECARES, Imperial College London and University of Bristol and from discussions with Samuli Leppälä, Heikki Rantakari and Stuart Katwikirize of Plan International. Maija Halonen-Akatwijuka is grateful to Ken Odur Gabelle, Simon Anyanzo and Stephen Mawa Alatawa of Uganda Red Cross Society for discussions and facilitating a visit to Adjumani refugee camp. We gratefully acknowledge financial support from ESRC (Grant ES/H005331/1).

[1]See Amin (2008) for more details about Risepak.

tion sharing. As much as the humanitarian organizations aim to alleviate suffering, there is diversity in primary motivations. We show that in a public goods setup such diversity makes direct information sharing – such as via an open access platform – ineffective. Filtered communication via a mediator – such as an appropriately designed information management system – can, however, improve coordination. While we motivate our study with humanitarian aid, similar information sharing issues arise in various situations where potentially biased agents allocate funds to public goods, e.g., in R&D joint ventures or public agencies.

We examine a setup where two agents allocate a fixed budget of one each between public goods in two areas, A and B. The areas have equal needs and therefore an equal allocation would maximize social welfare. The agents may, however, have biased preferences which is their private information. The agent is aligned with social welfare (neutral type) or biased to area A or B. Without communication the funds are allocated inefficiently resulting in gaps and duplication in public good provision.

Our first result is that direct communication is ineffective and cannot improve the allocation. Since the allocations to each area are public goods, the agent biased to area A has an incentive to represent himself as the type biased to B who would favor area B in his allocation. If the message was perceived as honest, it would steer the other agent to allocate less funds to area B and more funds to area A. It is exactly the agent biased to A who has the greatest incentive to send such a message rendering communication uninformative. In humanitarian context, this result can speak to unsuccessful experiences in direct information sharing, including an open access platform.

This result holds even when the potential biases are negligible in contrast to Crawford and Sobel (1982) where communication can be informative when the sender's bias is small enough. Our setup differs from Crawford and Sobel where the receiver knows the direction of the bias and only the receiver takes an action. In our model the direction of the potential bias is *unknown* and *both* agents take an action in a public good context. These features exacerbate the conflicts of interests and render direct communication ineffective.

Our second result is that communication via a mediator who does not have authority over the agents can improve the allocation. The agents report their type to the mediator as cheap talk (i.e., costless and unverifiable messages *a la* Crawford and Sobel 1982). The mediator can commit to a scheme which determines what information is revealed to the agents. This assumption applies well to an information management system as the mediator. We focus on schemes where the mediator announces a posterior profile (on the agents' types) and "instructs" allocations that form an equilibrium given the announced posteriors. In light of the insight above that revealing the direction of the bias is not incentive compatible, we restrict our attention to schemes that treat the biased types identically and characterize the optimal scheme. We find that it is optimal for the mediator to reveal the types fully only if both agents report neutral. In this case, the mediator randomly assigns an area for each agent to specialize in. It is incentive compatible for the neutral type to follow the mediator's instruction as it results in an equal allocation maximizing each agent's utility and the social welfare.

The mediator filters the rest of the information and only reveals one agent as biased – but not the direction of their bias. The case where both agents are biased is partially pooled equally with each of the two cases where only one is biased. Thus, when the mediator reveals one agent as biased, the other agent may be biased or not.

Since the direction of the bias is never revealed, a biased agent cannot gain anything by representing himself as the opposite type, but he might gain from reporting neutral. Randomization discourages the biased types from untruthfully reporting neutral. An agent biased to area A, when instructed to specialize in area B, may obtain his preferred allocation by diverting some of his budget to A given the other agent will specialize in A. However, he cannot bias the total allocation in his favor when instructed to specialize in area A since the other agent allocates all his budget to B. In addition, partial pooling in other cases generates unequal expected biases for the two agents. This steers their allocations to diverge and restrain certain biases, resulting in a more balanced total allocation.[2] We show that such information management induces the agents to reveal their types truthfully and follow the instructions, provided that the biases to area A and B are not of too different magnitude.

An extensive literature has developed on cheap talk communication since the seminal work of Crawford and Sobel (1982). However, studies on mediated communication have been sparse and largely conducted in very general framework (Forges 1986, 1990) or between an informed party and a decision maker (Goltsman, *et al.* 2009; Ivanov 2010) until recently. The current paper contributes to this growing literature by exploring how mediated communication may benefit multiple, privately informed players who are also action takers in a public good environment.

Two recent papers also study similar issues in different contexts. Goltsman and Pavlov (2014) show that Cournot duopoly firms with private costs may coordinate through a simple mediated mechanism when direct communication does not help. Hörner, Morelli and Squintani (2015) show that mediation can be devised to resolve conflicts between sovereign entities (whose strengths are private information) as effectively as if the mediator had enforcement power. As the players wish to appear as a tough type to their opponents in these contexts, mediation facilitates communication by constraining the aggression of stronger contenders via information filtering. In our context, there is no dominant type the agents wish to dress up as; instead, the problem stems from the agents trying to steer others' contributions in their own favor by misrepresenting their biases. Consequently, filtering information in the current context is devised to mitigate such

[2]To be precise, the allocation is more balanced for all but one type realization and the expected social welfare is higher.

effects so as to foster communication.

Direct communication between agents making voluntary contributions to a public good has been examined by Palfrey and Rosenthal (1991) and Palfrey, Rosenthal and Roy (2017). Their setup differs from ours in that there is only one public good and therefore a freeriding incentive arises. However, the public good is discrete and a threshold of contributions is needed giving the agents an incentive to coordinate. They show that in this setup direct communication can enhance efficiency.

Our work is also related to the literature on organizational design and communication such as Dessein (2002), Alonso, Dessein and Matouschek (2008) and Rantakari (2008). Alonso, Dessein and Matouschek (2008) and Rantakari (2008) compare decentralization and horizontal communication to centralization and vertical communication in a multidivisional organization. They show that when coordination is very important decentralization can be optimal as divisional managers have good incentives to coordinate via horizontal communication. We show that in a public goods context such horizontal communication is ineffective but vertical communication can improve coordination even when the control rights are decentralized. Dessein (2002) examines delegation as an alternative to communication and finds that the principal prefers delegation to communication when the agent's bias is small relative to the principal's uncertainty. The principal furthermore delegates to an intermediary when the agent's bias is moderate. In our setup there is no principal and the intermediary does not have control rights but has a role in mediating communication.

The paper is structured as follows. Section I sets up the allocation game. Section II examines the allocation game when the agents do not communicate. Section III shows that direct communication between the agents cannot improve upon the outcome of no communication. Section IV analyzes the mediated communication outcome and derives the conditions under which it results in welfare improving allocation. Section V discusses applications. Section VI concludes and Appendix contains deferred proofs.

I. Model

There are two agents, 1 and 2, with a budget of 1 to allocate in public goods in two areas, A and B. Agent $i \in \{1, 2\}$ allocates $a_i \in [0, 1]$ to area A leaving $(1 - a_i)$ to area B.[3] The areas have equal needs and the social welfare index is

$$w(a) = -(1 - a)^2$$

where $a = a_1 + a_2$. Social welfare is maximized by allocating half of the total budget of 2 to area A.

[3]For expositional ease, we assume that the agents must allocate all their budgets. This is the case in equilibrium if both agents' utility functions increase in allocations to each area.

The agents are one of three types in $T := \{\ell, n, h\}$ and have a utility function $-(1+t-a)^2$ when of type $t \in T$ where $\ell < n = 0 < h$. That is, an ℓ-type (h-type) is biased against (toward) area A and wants less (more) funds in area A than an n-type who is neutral. Agent's type is their private information. Note that, abusing notation slightly, ℓ, n and h are used both for the types and the degree of biases.

The three types are assumed to be equally likely in the main model,[4] depicted by the prior $\mu_0 = (\mu_0^\ell, \mu_0^n, \mu_0^h) = (1/3, 1/3, 1/3)$. To simplify exposition, we also assume that the potential biases are not too large and have different magnitudes.

ASSUMPTION 1: $0 < |\ell| < h < 1/8$.

Although there is diversity in motivations of the public good providers, we assume that they are not too biased. This is a reasonable assumption for many public good providers, e.g., humanitarian organizations. Different magnitude of bias can result, e.g., from the strength of prior relationship between the agent and the recipient.[5]

We compare the agents' allocation decisions with no communication, direct communication between the agents, and mediated communication. Our interest is finding out when and how communication can improve social welfare.

II. No Communication

The game with no communication is a standard static Bayesian game where the agents simultaneously decide on a_i contingent on their type. Let $a_i = (a_i^\ell, a_i^n, a_i^h)$ denote agent i's allocation strategy where a_i^t is the amount that agent i allocates to area A when its type is $t \in T$. The marginal utility of allocating a_i^t is

$$2 \sum_{s \in T} \frac{1}{3} \left(1 + t - a_i^t - a_{-i}^s\right)$$

where a_{-i}^s is the other agent's allocation. Then his unconstrained optimum from the first order condition is

$$a_i^t = 1 + t - E(a_{-i}) \tag{1}$$

where $E(a_{-i}) = \sum_{s \in T} \frac{1}{3} a_{-i}^s$ is the expected allocation of the other agent.

That is, since the two agents' allocations are perfect substitutes, each type's optimal allocation makes up for the shortfall of the other agent's expected allocation from his ideal allocation, $1 + t$. Note that an n-type agent always allocates such

[4]In Appendix, the main results are proved more broadly than the uniform prior.

[5]Bilateral relationship, e.g., colonial history, trade relationship, common language and geographic proximity, has been shown to increase humanitarian aid (Drury, Olson and Van Belle 2005; Strömberg 2007; Fink and Redaelli 2011).

unconstrained optimal level, (1), because $t = 0$ and $0 \leq E(a_{-i}) \leq 1$. Other types $t \in \{\ell, h\}$ also allocate their unconstrained optimal levels subject to feasibility, i.e., ℓ-type allocates less than n-type by $|\ell|$ and h-type more than n-type by h. In case such allocation is infeasible because it is below 0 for ℓ-type or above 1 for h-type, the constrained optimum would be to allocate 0 or 1, respectively.

We show that every equilibrium is an *interior* equilibrium absent communication, that is, every type of both agents allocates their unconstrained optimal level, (1). This result is recurring in subsequent analysis and prevails more broadly: specifically, if and only if both agents have identical expected biases, defined as

$$\mathrm{Eb}_i \ := \ \ell\mu_i^\ell + h\mu_i^h \quad \text{for agent} \quad i \in \{1, 2\}$$

where μ_i^t is the probability that agent i is of type $t \in T$. (In the current case, $\mathrm{Eb}_1 = \mathrm{Eb}_2 = \ell\mu_0^\ell + h\mu_0^h$.) Therefore, we establish the result for the case that $\mathrm{Eb}_1 = \mathrm{Eb}_2$.

First, consider an arbitrary interior equilibrium (presuming one exists) with allocation strategy $a_i = (a_i^\ell, a_i^n, a_i^h)$ for $i = 1, 2$. Then, $a_i^\ell = a_i^n + \ell$ and $a_i^h = a_i^n + h$ by (1) and, consequently, $E(a_i) = a_i^n + \mathrm{Eb}_i$. Thus, agent i's expected allocation exceeds the allocation of the neutral type by Eb_i.

The neutral type of either agent, say 1, covers the shortfall of the other agent 2's expected allocation, i.e., $a_1^n + E(a_2) = 1$. Since $E(a_2)$ exceeds a_2^n by Eb_2, this implies that $a_1^n + a_2^n + \mathrm{Eb}_2 = 1$. An analogous condition holds also for agent 2, i.e., $a_2^n + a_1^n + \mathrm{Eb}_1 = 1$, in any interior equilibrium. Hence, $\mathrm{Eb}_1 = \mathrm{Eb}_2$ is necessary for an interior equilibrium to exist. In other words, in an interior equilibrium both agents of neutral type cover the shortfall of each other's allocation plus the associated expected bias, which is feasible if and only if the expected biases are the same.

Indeed, if $\mathrm{Eb}_1 = \mathrm{Eb}_2 = \mathrm{Eb}$, so long as

$$(2) \qquad a_1^n + a_2^n \ = \ 1 - \mathrm{Eb} \qquad \text{and}$$

$$(3) \quad a_i^\ell = a_i^n + \ell \geq 0 \ \text{ and } \ a_i^h = a_i^n + h \leq 1 \qquad \Longleftrightarrow \qquad |\ell| \leq a_1^n, a_2^n \leq 1 - h,$$

it constitutes an interior equilibrium for each agent i to allocate $a_i = (a_i^\ell, a_i^n, a_i^h)$. In a symmetric equilibrium, $a_1^n = a_2^n = (1 - \mathrm{Eb})/2$. Additionally, there is a continuum of asymmetric equilibria which can be obtained by increasing agent i's allocation and decreasing agent $-i$'s allocation by the same amount subject to (2) and (3). In all these equilibria, the total allocation is the same contingent on type realization, $a_1^t + a_2^s = a_1^n + t + a_2^n + s = 1 - \mathrm{Eb} + t + s$ for each type pair (t, s).

Table 1 presents the total allocation (to area A) without communication where $\mathrm{Eb} = (h + \ell)/3$. Both agents compensate for the other agent's expected bias in their allocations, resulting in two n-type agents underallocating to area A by Eb. The biased types furthermore add their own bias to the allocation. Thus, the allocation is inefficient, i.e., it diverges from the socially optimal allocation of 1 in every realization. Since the bias of h-type is larger than that of ℓ-type, there is overallocation to area A when at least one agent is of h-type. However, equilibrium payoffs are identical across types and agents, calculated to be $-2(h^2 - h\ell + \ell^2)/9$, because every type t obtains the unconstrained optimum of equalizing the expected allocation to their ideal, $1 + t$.

TABLE 1—TOTAL ALLOCATION TO A WITH NO COMMUNICATION.

$1\backslash 2$	ℓ	n	h
ℓ	$1 - h/3 + 5\ell/3$	$1 - h/3 + 2\ell/3$	$1 + 2(h + \ell)/3$
n	$1 - h/3 + 2\ell/3$	$1 - (h + \ell)/3$	$1 + 2h/3 - \ell/3$
h	$1 + 2(h + \ell)/3$	$1 + 2h/3 - \ell/3$	$1 + 5h/3 - \ell/3$

We have verified above that an interior equilibrium exists if and only if $\text{Eb}_1 = \text{Eb}_2$ and the set of all interior equilibria is fully characterized by (2) and (3).[6]

Next, we show that there is no other equilibrium when $\text{Eb}_1 = \text{Eb}_2$. With a view to reaching a contradiction, suppose (a_1, a_2) is an equilibrium where some a_i^t is a constrained optimum. Assuming $a_1^n < a_2^n$ without loss of generality,[7] either $a_1^\ell = 0 > a_1^n + \ell$ or $a_2^h = 1 < a_2^n + h$ is a constrained optimum (possibly both) and all other values of a_i^t are at their unconstrained optimum, (1). Recall that $E(a_i) = a_i^n + \text{Eb}_i$ if a_i^t is unconstrained optimum for all $t \in T$. Hence, $a_1^\ell = 0 > a_1^n + \ell$ would imply that $E(a_1) > a_1^n + \text{Eb}_1$ and $a_2^h = 1 < a_2^n + h$ would imply that $E(a_2) < a_2^n + \text{Eb}_2$. Since one of the two inequalities should hold, $\text{Eb}_1 = \text{Eb}_2$ implies $E(a_1) - a_1^n > E(a_2) - a_2^n$, or equivalently $a_2^n + E(a_1) > a_1^n + E(a_2)$. This contradicts the equilibrium condition for n-type: $a_i^n + E(a_{-i}) = 1$ for both $i = 1, 2$, as by (1) a_i^n is an unconstrained optimum. Intuitively, if $a_1^\ell = 0$ or $a_2^h = 1$ is the constrained optimum, it would imply that agent 1's expected allocation is biased upward more than agent 2's expected allocation relative to their respective allocation of n-type, incentivising the agents to adjust their allocations closer to each other and moving away from the corner solution.

We summarize the results obtained in the absence of communication below.

PROPOSITION 1: *In the absence of communication, the set of Bayesian Nash equilibria (a_1, a_2) is fully characterized by (2) and (3) where* $\text{Eb} = (h + \ell)/3$. *In all these equilibria, the total allocation to area A for each type realization is given in Table 1. The allocation is inefficient and the equilibrium payoff of each type is* $-2(h^2 - h\ell + \ell^2)/9$.

[6] Note that we have only considered pure allocation strategies. As it is straightforward from the utility function that the unconstrained optimum a_i^t satisfies (1) even if the other agent adopts mixed strategies, it follows that no mixed strategy equilibrium exists.

[7] If $a_1^n = a_2^n$ then $a_i^n + \ell < E(a_1) = E(a_2) < a_i^n + h$ and thus, (1) would imply $2a_1^n + \ell < 1 < 2a_1^n + h$ and consequently, $-\ell < a_i^n < 1 - h$, contradicting some a_i^t being a constrained optimum.

III. Direct Communication

In this section, prior to allocation decisions we allow one round of communication in which agents $i \in \{1,2\}$ send a public cheap talk message m_i simultaneously. Each such message m_i induces a posterior belief $\mu_i = (\mu_i^\ell, \mu_i^n, \mu_i^h)$ on agent i's type by Bayes rule from agent i's message sending strategy. The two agents then make allocation decisions given the posterior profile (μ_1, μ_2), which we call a continuation game. The entire game is referred to as the *direct communication game.*

A. Allocations After Communication

We first examine the allocation decisions in the continuation game after communication with a posterior profile (μ_1, μ_2).[8] These results will be instrumental also in the analysis of mediated communication. Denoting the (continuation) equilibrium allocations as $a_1 = (a_1^\ell, a_1^n, a_1^h)$ and $a_2 = (a_2^\ell, a_2^n, a_2^h)$, agent 1 of type $t \in \{\ell, n, h\}$ solves

$$(4)\quad \max_{a_1^t} \; -\mu_2^\ell(1-a_1^t-a_2^\ell+t)^2 - \mu_2^n(1-a_1^t-a_2^n+t)^2 - \mu_2^h(1-a_1^t-a_2^h+t)^2.$$

The first order condition for the solution $a_1^t \in [0,1]$ is

$$(5)\qquad 2\sum_{s\in T}\mu_2^s\,(1+t-a_1^t-a_2^s\,)\begin{cases}\leq 0 & \text{if } \; a_1^t = 0\\ = 0 & \text{if } \; a_1^t \in (0,1)\\ \geq 0 & \text{if } \; a_1^t = 1.\end{cases}$$

Solving (5) and by symmetry, we deduce that $a_1 = (a_1^\ell, a_1^n, a_1^h)$ and $a_2 = (a_2^\ell, a_2^n, a_2^h)$ constitute an equilibrium if and only if they solve

$$(6)\qquad \begin{cases} a_1^\ell &= \max\{0, 1+\ell-E(a_2)\}\\ a_1^n &= 1-E(a_2)\\ a_1^h &= \min\{1+h-E(a_2), 1\}\end{cases} \text{ and } \begin{cases} a_2^\ell &= \max\{0, 1+\ell-E(a_1)\}\\ a_2^n &= 1-E(a_1)\\ a_2^h &= \min\{1+h-E(a_1), 1\}\end{cases}$$

where $E(a_i)$ is agent i's expected allocation given μ_i. Note that this is the case even when μ_i does not have a full support, in which case a_i^t is said to be "relevant" if $\mu_i^t > 0$ and "irrelevant" otherwise. As before, $a_i^\ell = 0 > 1 + \ell - E(a_{-i})$ and $a_i^h = 1 < 1 + h - E(a_{-i})$ are constrained optimum allocation levels.

An equilibrium is *interior* if all relevant allocations a_i^t are unconstrained optimum, and is *noninterior* otherwise. Below we characterize the set of interior

[8]The dependence of μ_i on messages is suppressed when no confusion arises.

and noninterior equilibria in the continuation game after communication, which depends on whether the two agents' expected biases coincide or not.

Interior equilibria

Consider *interior* equilibria where each relevant a_i^t satisfies the first order condition (5) with equality. By taking expectation of a_1 and a_2 in (6) and rearranging, we get

$$\begin{aligned} E(a_1) + E(a_2) &= 1 + h\mu_1^h + \ell\mu_1^\ell = 1 + h\mu_2^h + \ell\mu_2^\ell \\ \Longrightarrow \quad h\mu_1^h + \ell\mu_1^\ell &= h\mu_2^h + \ell\mu_2^\ell. \end{aligned} \tag{7}$$

Thus, according to (7) a necessary condition for an interior equilibrium is that the two agents have equal expected biases denoted by $\mathrm{Eb}(\mu_i) = h\mu_i^h + \ell\mu_i^\ell$ for $i = 1, 2$.

The analysis and results of Section II extend straightforwardly to this case where $\mathrm{Eb}_i = \mathrm{Eb}(\mu_i)$. We only need to modify Proposition 1 for possible irrelevant allocation values, as stated in the lemma below.

LEMMA 1: *In the continuation game with a posterior profile* (μ_1, μ_2) *such that* $\mathrm{Eb}(\mu_1) = \mathrm{Eb}(\mu_2) = \mathrm{Eb}$*, the set of Bayesian Nash equilibria* (a_1, a_2) *is fully characterized by*

(i) $a_1^n + a_2^n = 1 - \mathrm{Eb}$*, and*

(ii) $a_i^t = a_i^n + t$ *if* $\mu_i^t > 0$ *for* $i \in \{1, 2\}$ *and* $t \in \{\ell, h\}$*.*

In all these equilibria, the total allocation to area A for each type realization is given in Table 2 (so long as $\mu_1^{t_1} \cdot \mu_2^{t_2} > 0$*). The equilibrium payoff of agent i is* $-\sum_{t\in T} \mu_{-i}^t (\mathrm{Eb} - t)^2$ *regardless of type (so long as in the support of* μ_i*).*

TABLE 2—TOTAL ALLOCATION TO A WITH DIRECT COMMUNICATION.

1\2	ℓ	n	h
ℓ	$1 + 2\ell - \mathrm{Eb}$	$1 + \ell - \mathrm{Eb}$	$1 + h + \ell - \mathrm{Eb}$
n	$1 + \ell - \mathrm{Eb}$	$1 - \mathrm{Eb}$	$1 + h - \mathrm{Eb}$
h	$1 + h + \ell - \mathrm{Eb}$	$1 + h - \mathrm{Eb}$	$1 + 2h - \mathrm{Eb}$

Noninterior equilibria

Suppose now that expected biases are not equal, say without loss of generality,

$$\mathrm{Eb}(\mu_1) < \mathrm{Eb}(\mu_2). \tag{8}$$

Then there is a unique noninterior equilibrium as explained below.

Recall that the two agents of n-type just cover the shortfall of the other agent's expected allocation, i.e., $a_i^n = 1 - E(a_{-i})$; or equivalently, they cover the shortfall of each other's allocation, then reduce by $E(a_{-i}) - a_{-i}^n$ to correct for the amount of the other's expected allocation in excess of n-type's allocation, i.e.,

$$a_i^n = 1 - a_{-i}^n - (E(a_{-i}) - a_{-i}^n) \quad \Longleftrightarrow \quad a_1^n + a_2^n = 1 - (E(a_i) - a_i^n) \text{ for } i = 1, 2.$$

Hence, any equilibrium (a_1, a_2) must satisfy $E(a_1) - a_1^n = E(a_2) - a_2^n$, that is, the expected allocation differs from the allocation of the neutral type by the same amount for both agents. If (8) holds, however, this is possible neither in an interior equilibrium where $E(a_i) - a_i^n = \text{Eb}(\mu_i)$, nor in a noninterior equilibrium where $a_1^n \geq a_2^n$ because then either $a_1^h = 1$ or $a_2^\ell = 0$ is noninterior so that $E(a_1) - a_1^n \leq \text{Eb}(\mu_1)$ and $\text{Eb}(\mu_2) \leq E(a_2) - a_2^n$.

Therefore, only noninterior equilibrium with $a_1^n < a_2^n$ is possible, where a_1^ℓ or/and a_2^h is noninterior while all the other allocations are interior solutions to (4). The equilibrium is unique because the optimal level of a_i^n moves in the opposite direction as a_{-i}^n but at a rate no higher than unity, resulting in a unique fixed point (see Appendix). For each agent i, the equilibrium payoff of type t is the same so long as a_i^t is interior solution (which is always the case for $t = n$) because such types obtain unconstrained optimum, but is lower if a_i^t is noninterior because such a type only obtains constrained optimum. The findings are summarized below.

LEMMA 2: *If* $\text{Eb}(\mu_1) < \text{Eb}(\mu_2)$*, there is a unique equilibrium* (a_1, a_2) *and it is noninterior. Moreover,* $a_1^n < a_2^n$ *and* a_1^ℓ *or/and* a_2^h *is noninterior while* a_1^n*,* a_2^n*,* a_1^h *and* a_2^ℓ *are interior solutions (even if irrelevant). For each agent* i*, the equilibrium payoff is lower for a type with a noninterior allocation than for those types with interior allocations who thus have identical payoffs.*

B. Direct Communication Has No Effect

Having examined the allocation choices as above, consider now a *perfect Bayesian equilibrium* (PBE) of the direct communication game. Agent i sends a message m_{ik} from a set $M_i = \{m_{i1}, m_{i2}, \cdots, m_{iK_i}\}$ of K_i messages, inducing a posterior belief μ_{ik} on his type to be held by the other agent. We label messages so that a message labelled higher induces a posterior with a weakly higher expected bias:

$$\text{Eb}(\mu_{i1}) \leq \text{Eb}(\mu_{i2}) \leq \cdots \leq \text{Eb}(\mu_{iK_i}) \quad \text{for} \quad i = 1, 2.$$

If agent i's expected bias does not depend on the message he sends, i.e., $\text{Eb}(\mu_{i1}) = \text{Eb}(\mu_{iK_i})$ for $i = 1, 2$, it must be equal to the prior bias $(h + \ell)/3$. Then, regardless of the messages sent, the continuation equilibrium outcome is the same as that without communication (Proposition 1). This corresponds to the so-called *babbling equilibrium* where agents send arbitrary messages randomly because they will be ignored completely.

The main result to establish is that there is no other equilibrium than a babbling equilibrium in the direct communication game. In a nutshell, the core logic is as follows. If $\mathrm{Eb}(\mu_{11}) < \mathrm{Eb}(\mu_{1K_1})$, agent 1 would appear more likely to be of h-type by sending the maximal message m_{1K_1} than the minimum message m_{11}. This would induce agent 2 to anticipate higher allocation by agent 1 to area A and thus, shift his own allocation away from A to B. Since the allocations are public goods, it is actually ℓ-type (rather than h-type) who would benefit from such shifting by agent 2. Thus, ℓ-type has an incentive to appear as h-type by sending messages that induce higher expected biases. Similarly, an h-type agent would have a greater incentive to send messages with lower expected biases. Then, the equilibrium posteriors of the messages would be self-defeating due to such crossed incentives, rendering communication uninformative. Accordingly, direct communication cannot help the agents to coordinate and the resulting allocation is the same as under no communication.

This result holds even when the potential biases are negligible in contrast to Crawford and Sobel (1982) where informative communication is feasible if the sender's bias is small enough. Our setup differs importantly from Crawford and Sobel where the receiver is the only decision maker and knows the direction of the sender's bias. In our model the direction of the potential bias is unknown and both agents take an action in a public goods context. It is the combination of these factors that results in uninformative communication.

We now provide a more detailed description of why meaningful direct communication is impossible. This proves useful in stating the result precisely and thereby, clarify the relationship between direct and mediated communication. Nevertheless, one may skip this technical part and go to Section IV to see how mediated communication improves welfare.

Returning to the analysis of PBE of the direct communication game, let us examine if a non-babbling equilibrium may be possible. To facilitate illustration, we focus on the case that

$$\mathrm{Eb}(\mu_{11}) \leq \mathrm{Eb}(\mu_{21}) \quad \text{and} \quad \mathrm{Eb}(\mu_{2K_2}) \leq \mathrm{Eb}(\mu_{1K_1}),$$

that is, the range of agent 1's expected biases after communication is weakly wider than that of agent 2. Then, aiming to prove that the expected biases are all equal, suppose they are not. That is, $\mathrm{Eb}(\mu_{11}) < (h+\ell)/3 < \mathrm{Eb}(\mu_{1K_1})$. This implies that $\mu_{1K_1}^h > 0$, i.e., agent 1 of h-type uses the maximal message m_{1K_1} (with a positive probability).

First, consider the case that $\mu_{11}^\ell > 0$, i.e., the minimum message m_{11} is used by agent 1 of ℓ-type, referred to as "agent 1-ℓ" for brevity. After sending m_{11}, agent 1-ℓ encounters a continuation game with a posterior pair (μ_{11}, μ_{2k}) for any $k \in \{1, 2, \cdots, K_2\}$. If agent 2 sends a message that induces expected bias of the same magnitude, $\mathrm{Eb}(\mu_{11}) = \mathrm{Eb}(\mu_{2k})$, the continuation equilibrium is interior and agent 1-ℓ obtains the same continuation payoff as agent 1-n by Lemma 2. If $\mathrm{Eb}(\mu_{11}) < \mathrm{Eb}(\mu_{2k})$, agent 1-$\ell$ obtains the same continuation payoff as agent 1-n if a_1^ℓ is unconstrained optimum but a lower payoff otherwise (Lemma 3).

Therefore, the expected payoff of agent 1-ℓ from sending the message m_{11} is no higher than that of agent 1-n and is strictly lower if a_1^ℓ is noninterior in at least one continuation equilibrium. On the other hand, if agent 1-ℓ sends the maximum message m_{1K_1}, since $\mathrm{Eb}(\mu_{1K_1}) \geq \mathrm{Eb}(\mu_{2k})$ we deduce that agent 1-ℓ always obtains the same payoff as that of agent 1-n because a_1^ℓ is interior in every continuation game with $\mathrm{Eb}(\mu_{1K_1}) \geq \mathrm{Eb}(\mu_{2k})$ by Lemmas 2 and 3.[9] Thus,

(a) The extra payoff from sending m_{1K_1} instead of m_{11} is higher for agent 1-ℓ than for agent 1-n, strictly so if a_1^ℓ is noninterior in at least in one continuation equilibrium after sending m_{11}.

Analogously, by Lemmas 2 and 3, agent 1-h always obtains the unconstrained optimum after sending the message m_{11} because $\mathrm{Eb}(\mu_{11}) \leq \mathrm{Eb}(\mu_{2k})$, but may fail to do so after sending the message m_{1K_1} because $\mathrm{Eb}(\mu_{1K_1}) \geq \mathrm{Eb}(\mu_{2k})$. Hence,

(b) The extra payoff from sending m_{11} instead of m_{1K_1} is higher for agent 1-h than for agent 1-n, strictly so if a_1^h is noninterior in at least in one continuation equilibrium after sending m_{1K_1}.

Finally, if agent 1-n weakly prefers m_{1K_1} to m_{11} in the presumed equilibrium, then agent 1-ℓ should strictly prefer sending m_{1K_1} to m_{11} by (a), provided that a_1^ℓ is noninterior in at least in one continuation equilibrium after sending m_{11}. Alternatively, if agent 1-n weakly prefers m_{11} to m_{1K_1}, agent 1-h should strictly prefer sending m_{11} to m_{1K_1} by (b), provided that a_1^h is noninterior in at least in one continuation equilibrium after sending m_{1K_1}. The former would contradict the supposition above that agent 1-ℓ uses m_{11}, and the latter would contradict the assertion above that agent 1-h uses m_{1K_1}. We show in Appendix that one of these prevails.

Next, consider the case that agent 1-ℓ does not use message m_{11}, i.e. $\mu_{11}^\ell = 0$. Then, $\mu_{11}^n > 0$ and $0 \leq \mathrm{Eb}(\mu_{11})$. In particular, agent 1-n optimally sends the message m_{11} and thus weakly prefers it to m_{1K_1}. It follows as before that agent 1-h would strictly prefer sending m_{11} to m_{1K_1}, contradicting $\mu_{1K_1}^h > 0$, because a_1^h is noninterior in at least one continuation equilibrium after sending m_{1K_1} but not after sending m_{11}. One tricky case for this argument is when $\mu_{11}^n = \mu_{21}^n = 1$ and $a_1^n > 1 - h$ in the continuation equilibrium after the message pair (m_{11}, m_{21}) so that a_1^h is noninterior. In this case, we show in Appendix that agent 2-h would strictly prefer sending m_{21} to m_{2K_2}, contradicting $\mu_{2K_2}^h > 0$.

The arguments sketched above establish that no equilibrium message induces a posterior different from the prior. This is subject to one minor caveat implicitly assumed in resolving the above mentioned tricky case:

($*$) The continuation equilibrium after any message pair is the same so long as the induced posterior profile is the same.

[9] To be fully precise, a_1^ℓ may be noninterior when it is irrelevant (i.e., $\mu_1^\ell = 0$) and $\mathrm{Eb}(\mu_{1K_1}) = \mathrm{Eb}(\mu_{2k})$. Such cases are taken care of in the proof of Proposition 2.

We believe that this is sensible assumption in direct communication game where cheap talk messages have little role beyond the posteriors they induce. However, we discuss what happens if this condition is relaxed in the next subsection.

PROPOSITION 2: *In every PBE of the direct communication game that satisfies* $(*)$*, the total allocation is identical to that in the equilibrium without communication.*

C. Jointly Controlled Lotteries

The condition $(*)$ is innocuous for continuation games with posteriors of unequal expected biases (because the continuation equilibrium is unique) and those with posteriors of equal biases and full supports (because all continuation equilibria generate identical payoffs for all types). When the expected biases are equal and some biased types are irrelevant, however, the irrelevant type's (off-equilibrium) continuation payoff may differ depending on which continuation equilibrium prevails. This opens the possibility of reviving some communication by employing a "jointly controlled lottery" *à la* Forges (1990) to manipulate truthtelling incentives.

It turns out that such manipulation may have an impact only in the continuation game starting with the posterior pair $\mu_1 = \mu_2 = (0, 1, 0)$ when the magnitudes of biases, h and $|\ell|$, are close enough. We outline the efficient PBE in such cases below.

In the efficient PBE, both agents send the same message b if of a biased type, i.e., if $t \in \{h, \ell\}$. This removes the incentive of the biased agent to pretend to be of the opposite type as they simply say 'biased'. When of a neutral type, they randomize between sending messages n and n' with equal probability. The continuation equilibrium is as follows depending on the message pair sent:

- After (b, b), any continuation equilibrium for $\text{Eb}(\mu_1) = \text{Eb}(\mu_2) = (h + \ell)/2$ described in Lemma 1 follows.
- After only one agent reports b, the unique noninterior equilibrium follows (Lemma 2).
- After (n, n) and (n', n'), agent 1 allocates 1 to area A and 2 allocates 0.
- After (n, n') and (n', n), agent 1 allocates 0 to area A and 2 allocates 1.

Any message pair $(m_1, m_2) \in \{n, n'\} \times \{n, n'\}$ induces a posterior pair $\mu_1 = \mu_2 = (0, 1, 0)$ and each agent specializes in one area. The area of specialization is randomized depending on the messages. The role of two equivalent messages, n and n', is to introduce jointly controlled randomization in order to discourage the biased types from pretending to be unbiased.

The description above constitutes a PBE of the direct communication game if sending messages as above is optimal given the stipulated continuation equilibria. It is routinely verified that this is the case if and only if $\ell \leq -\left(1 + \sqrt{5}\right) h/4 \approx -0.81h$, that is, when the biases to area A and B are of a relatively similar magnitude.

In this case, the total allocation is given in Table 3. Compared with a babbling equilibrium (Table 1), separating the neutral type from the biased types increases expected welfare. First, ideal allocation is obtained when both agents are neutral as the agents specialize in different areas. Second, the expected bias is larger when both agents report biased, (b, b), than in the babbling equilibrium, leading to a larger under-allocation (to A) by both agents to correct the larger upward bias of the other agent. This pushes the total allocation closer to (away from) the social optimum when at least one agent is of h-type (both are ℓ-type) relative to the babbling equilibrium. The positive effect dominates because the inefficiency is larger in the babbling equilibrium when both agents are of h-type than of ℓ-type and welfare function is quadratic. Lastly, if one agent is neutral and the other agent is biased, the biased agent's allocation is constrained at 1 when of h-type. Together with the certainty that the opponent is neutral, this reduces the variance of the biased agent's allocation and consequently, improves efficiency.

Table 3—Total allocation to A with jointly controlled lottery.

$1\backslash 2$	ℓ	n	h
ℓ	$1 - h/2 + 3\ell/2$	$1 + \ell$	$1 + h/2 + \ell/2$
n	$1 + \ell$	1	$1 - \ell$
h	$1 + h/2 + \ell/2$	$1 - \ell$	$1 + 3h/2 - \ell/2$

This result relies on the jointly controlled lottery which reduces the incentive for biased types to "lie": by sending n or n', either biased type $t \in \{h, \ell\}$ faces a $1/2$ chance of ending up with $|t|$ away from their ideal allocation.[10]

Such randomization does not appear a particularly natural consequence of direct communication; it rather seems a sensible outcome of guided communication by a mediator. In fact, we show in Section IV that mediated communication achieves even higher welfare by introducing partial pooling.

For a smooth flow, we defer detailed analysis of this Section to Online Appendix, where we also prove that the PBE specified above achieves the maximum efficiency possible via direct communication if $\ell \leq -\left(1 + \sqrt{5}\right) h/4$, but the babbling equilibrium is the unique PBE otherwise.

[10]Without randomization, at least one biased type of either agent can get their ideal allocation for sure by pretending to be unbiased when the other agent is indeed unbiased, rendering separation between biased and unbiased types non-viable under the condition $(*)$. Hence, a variety of different jointly controlled lotteries work in our model, so long as they introduce enough uncertainty in continuation equilibrium for the neutral type to discourage pretension by other types. In other contexts, e.g., Krishna and Morgan (2004), they are used to introduce just the right amount of uncertainty for a threshold type to be indifferent between multiple messages, so that there is a unique jointly controlled lottery that works for a given equilibrium outcome.

IV. Mediated Communication

In this Section we analyze mediated communication between the agents. In the first stage, the agents privately report/send a cheap talk message to the mediator (M). In the second stage, M makes a public announcement interpreted as an "instruction" (which is also cheap talk). M does not have authority over the agents, so they are not obliged to follow her instructions. In the third stage, each agent simultaneously selects an allocation contingent on his type, his report and the instruction received.

Before the first stage, M publicly and credibly commits to an announcement strategy as a function of the agents' reports, which we refer to as a mediation scheme. Given a mediation scheme, we examine the PBE of the game between the two agents as above. Our aim is to study whether and to what extent mediation may improve social welfare relative to the case of direct or no communication examined above.

A. Mediation Schemes

Note that the Revelation Principle applies *à la* Myerson (1982) and thus, we only need to consider mediation schemes that induce a PBE in which the agents report their types truthfully and follow the instructions. Ideally, we would like to characterize optimal mediation schemes that induce a PBE with the maximum social welfare achievable with any mediation scheme. But, finding an optimal mediation scheme is generally known to be difficult, and the current case is no exception. Instead, we characterize below the optimal mediation scheme subject to two conditions that we think are reasonable:

(c1) M's instruction is consistent with a continuation equilibrium (interior or noninterior) characterized in Section III.A.
(c2) M may reveal whether an agent is biased or not, but not the direction of bias.

Any announcement that fails (c1) entails that an agent's posterior on the other's types differ depending on his own type, creating nontrivial complications in figuring out the continuation equilibrium for each such announcement. (c2) reflects our finding that the direction of bias cannot be effectively communicated directly, the underlying strategic reasoning for which also exists inherently in mediated communication.[11] The two conditions obviously restrict the set of mediation schemes we consider. However, we suspect that other schemes are likely to be too complex for practical use.

Hence, we focus our attention to mediation schemes where M announces publicly a posterior profile $\boldsymbol{\mu} = (\mu_1, \mu_2) \in \mathcal{P} := \Delta(T) \times \Delta(T)$ and instructs an

[11] For the simple case that both agents are one of two types, h or ℓ, it can be shown straightforwardly that even mediated communication is ineffective for this reason.

allocation profile $\boldsymbol{a} = (a_1, a_2) \in \mathcal{A} := [0,1]^6$. By virtue of the Revelation Principle, therefore, we define a "mediation scheme" as $\pi : T \times T \to \Delta(\mathcal{P} \times \mathcal{A})$ that specifies, contingently on the report profile $(t_1, t_2) \in T \times T$ received privately from the two agents, the probability with which the mediator publicly announces a posterior profile $\boldsymbol{\mu}$ and an allocation profile $\boldsymbol{a}$ that, by condition (c1), *forms a continuation equilibrium for the posterior profile* $\boldsymbol{\mu}$. Condition (c2) requires that the two biased types are equally likely for each agent according to $\boldsymbol{\mu}$, i.e.,

$$(9) \quad \pi(\boldsymbol{\mu}, \boldsymbol{a}|t_1, t_2) > 0 \text{ for some } (t_1, t_2) \in T \times T \implies \mu_i^\ell = \mu_i^h \text{ for } i = 1, 2.$$

We take it as given that this condition holds for all mediation schemes in what follows.

Given a mediation scheme π, consider an arbitrary announcement $(\boldsymbol{\mu}, \boldsymbol{a})$ made with a positive probability. Conditional on truthful reporting by the agents, each agent i of type t_i forms a posterior belief on the other agent's type by Bayes rule as[12]

$$\mu_{-i}^{t_{-i}}(\boldsymbol{\mu}, \boldsymbol{a}) = \frac{\pi(\boldsymbol{\mu}, \boldsymbol{a}|t_i, t_{-i})\mu_0(t_{-i})}{\sum_{s_{-i}\in T} \pi(\boldsymbol{\mu}, \boldsymbol{a}|t_i, s_{-i})\mu_0(s_{-i})} \quad \forall\, t_{-i} \in T.$$

We say that a scheme π is *consistent* if, for every message $(\boldsymbol{\mu}, \boldsymbol{a})$ in the support of π, the posterior on the other agent $-i$ above is well-defined for every "relevant" type t_i (i.e, $\mu_i^{t_i} > 0$) and coincides with the announced posterior μ_{-i}. Given a consistent scheme π, conditional on truthful reporting, following the instructed allocation is optimal for both players as it constitutes a continuation equilibrium by construction. A consistent scheme is *incentive-compatible (IC)* if truthful-reporting is optimal for both agents conditional on the other agent reporting truthfully. For brevity, we say a scheme π is *CIC* if it is consistent and incentive-compatible. Given a CIC scheme, it is clearly a PBE for both agents to report truthfully and follow the announced allocation. In what follows, we solve for the optimal CIC mediation scheme.

DEFINITION: *A mediation scheme is optimal if it maximizes the expected social welfare among all CIC mediation schemes.*

B. Optimal Mediation Scheme

We show that in any optimal mediation scheme the agents' types are revealed fully only when both agents are neutral. Each agent is then instructed to specialize in a randomly chosen area achieving equal allocation and maximizing welfare, while minimizing the incentive for biased types to appear as neutral. In all other cases, only one agent is revealed as biased, inducing partial pooling on the other agent who may be biased or not. Such filtering of types leads to unequal expected

[12] Note that this posterior is well-defined for t_i if and only if $\pi(\boldsymbol{\mu}, \boldsymbol{a}|t_i, t_{-i}) > 0$ for some t_{-i}, hence may be well-defined for some but not other type(s) of agent i.

biases and noninterior equilibrium reducing the variance of allocations. While direct communication with jointly controlled lotteries can achieve randomized specialization of two neutral agents—albeit quite unnaturally—partial pooling requires a mediator. This is the key reason why mediation can improve welfare.

We now proceed to prove these insights by informally deriving a series of Lemmas. From now on, for brevity we refer to π simply as a "scheme" and denote an announcement $(\boldsymbol{\mu}, \boldsymbol{a})$ of M simply by the posterior pair $\boldsymbol{\mu}$, with the understanding that a corresponding $\boldsymbol{a}$ (which is a continuation equilibrium for $\boldsymbol{\mu}$) is implicit.

Consider a CIC scheme π' and its "dual scheme" π^d obtained from π' by swapping agents 1 and 2, that is, $\pi^d(\mu_2, \mu_1|t, s) = \pi'(\mu_1, \mu_2|s, t)$ for all $(\mu_1, \mu_2) \in \Delta(T) \times \Delta(T)$ and $(s, t) \in T \times T$. Let π be a modified scheme that announces the same as either π' or π^d with equal probability after every report profile (s, t), so that π is a *symmetric* scheme in the sense that the two agents are treated symmetrically, that is,

$$\pi(\mu_1, \mu_2|s, t) \;=\; \pi(\mu_2, \mu_1|t, s) \quad \text{for all} \quad (s, t) \in T \times T. \tag{10}$$

Since π' is consistent by construction, so is π; since both π' and π^d are IC, so is π. Moreover, π' and π^d generate the same welfare as they are identical modulo relabelling of agents and consequently, so does π. Therefore, we may focus on symmetric schemes to characterize an optimal scheme.

LEMMA 3: *There is an optimal mediation scheme that is symmetric.*

Consider a consistent and symmetric mediation scheme π and a posterior pair $\boldsymbol{\mu} = (\mu_1, \mu_2)$ announced (with a positive probability) by π. Given $\mu_i^\ell = \mu_i^h$ by (9), let $\beta_i = \mu_i^\ell + \mu_i^h = 2\mu_i^\ell = 2\mu_i^h$ denote the probability that agent i is of a biased type according to μ_i, hence the probability that he is unbiased is $\mu_i^n = 1 - \beta_i$. Then, μ_i is a convex combination of the biased posterior denoted by $\mu_b := (1/2, 0, 1/2)$ and the neutral posterior denoted by $\mu_n := (0, 1, 0)$ with weights β_i and $1 - \beta_i$:

$$\mu_i = \left(\frac{\beta_i}{2}, 1 - \beta_i, \frac{\beta_i}{2}\right) \;=\; \beta_i \cdot \left(\frac{1}{2}, 0, \frac{1}{2}\right) + (1 - \beta_i) \cdot (0, 1, 0) = \beta_i \mu_b + (1 - \beta_i)\mu_n.$$

We call μ_i a composite posterior if $0 < \beta_i < 1$ (i.e, a proper combination of μ_b and μ_n) and more biased if β_i is higher. Clearly, the more biased is a posterior the higher is its expected bias.

By symmetry, if $\boldsymbol{\mu} = (\mu_1, \mu_2)$ is announced by π then its "dual" (μ_2, μ_1) is also announced with the same probability. If either posterior in $\boldsymbol{\mu}$, say μ_2, is composite, then we can modify the scheme π to another scheme π' by decomposing μ_2 into μ_b and μ_n when paired with μ_1, via revealing whether agent 2 is biased or not: that is, for any (t_1, t_2) such that $\pi(\mu_1, \mu_2|t_1, t_2) > 0$,

- set $\pi'(\mu_1, \mu_n|t_1, t_2) = \pi(\mu_1, \mu_2|t_1, t_2)$ if $t_2 = n$;
- set $\pi'(\mu_1, \mu_b|t_1, t_2) = \pi(\mu_1, \mu_2|t_1, t_2)$ if $t_2 \in \{h, \ell\}$,

and analogously for the dual posterior (μ_2, μ_1). Recall that it is implicit that (μ_1, μ_n) and (μ_1, μ_b) are announced with a corresponding continuation equilibrium allocation. We say that μ_2 is "decomposed (into μ_b and μ_n)" when π is modified to π' as above.

The modified scheme is clearly consistent, but not necessarily IC because reporting incentives may have changed. In Lemma 4 below, we examine the welfare effects of a few key modifications conditional on truthful reporting. As a result, we will identify a relatively simple consistent scheme with maximum welfare conditional on truthful reporting. Then, we show that it is indeed an optimal scheme (i.e., satisfies IC) unless h and $|\ell|$ are too far apart.

LEMMA 4: *Consider a posterior pair (μ_1, μ_2) announced by a consistent scheme π, where μ_1 is composite and weakly less biased than μ_2. Conditional on truthful reporting, the expected welfare*

(a) *strictly increases when μ_2 is decomposed;*

(b) *strictly decreases when $\mu_2 = \mu_b$ and μ_1 is decomposed; and*

(c) *strictly increases when $\mu_2 = \mu_b$ and the posterior pair $(\mu_1, \mu_2) = (\mu_1, \mu_b)$ is merged with another posterior pair $(\widetilde{\mu}_1, \mu_b)$ announced by π.*

The first modification in Lemma 4 starts with a composite posterior pair (μ_1, μ_2). Suppose for the sake of illustration that this is a "trivial" scheme where M simply announces the prior, $\mu_1 = \mu_2 = \mu_0$, and the allocations of a babbling equilibrium. Such scheme is obviously CIC. Then modify the scheme so that it reveals whether agent 2 is neutral or biased. According to Lemma 4(a), this modification leads to higher expected welfare. The welfare improvement stems from constrained allocation of biased types in noninterior equilibria. Specifically, when μ_2 is decomposed and (μ_1, μ_n) is announced instead of (μ_1, μ_2), the continuation equilibrium is noninterior where agent 2 of n-type, "agent 2-n" for short, obtains a higher payoff (which coincides with social welfare) because he optimizes against agent 1's allocation which now has a lower variance (since agent 1-h's allocation is constrained). When (μ_1, μ_b) is announced instead of (μ_1, μ_2), agent 2-h's allocation is constrained so that the total allocation is closer to social optimum on average. Hence, the expected social welfare is improved.[13]

Next, consider the case that agent 2 is revealed to be biased. According to Lemma 4(b), modifying the scheme to reveal further whether agent 1 is biased or not would lower welfare. When (μ_b, μ_b) is announced instead of (μ_1, μ_b), the continuation equilibrium is interior which hurts welfare due to an enlarged disparity in allocation between biased types of agent 2. When (μ_n, μ_b) is announced, welfare improves because a_2^h gets more constrained due to lower $\mathrm{Eb}(\mu_1)$. But, the former effect dominates because the (quadratic) welfare index drops more quickly as the total allocation moves away from the optimum than the index rises as it moves toward the optimum.

[13]We show in Appendix that the same result obtains by decomposing the more biased posterior if $\mu_1 \neq \mu_2$ (because a posterior pair with μ_b is announced with a higher probability where the welfare improvement is more pronounced).

Hence, decomposing μ_1 does not help in a posterior pair with $\mu_2 = \mu_b$. If there is another posterior pair announced by a consistent scheme with $\mu_2 = \mu_b$, merging the two posterior pairs (in the obvious manner described in Appendix) increases expected welfare according to Lemma 4(c), because consolidating agent 1's posteriors reduces uncertainty. Thus, only one posterior pair may be announced with $\mu_2 = \mu_b$.

Lastly, we consider the case where one of the agents was revealed to be neutral, which arises when the more biased posterior is decomposed *à la* Lemma 4(a). Lemma 5 below shows that revealing also whether the other agent is biased does not change welfare but relaxes ICs if combined with randomized specialization when both are unbiased. For consistency, we continue to treat agent 1 as the less biased agent.

LEMMA 5: *Suppose a posterior pair (μ_n, μ_2) is announced by a consistent scheme where μ_2 is composite. Decomposing μ_2 does not affect welfare conditional on truthful reporting, but reduces the incentive to misreport for biased types $t \in \{h, \ell\}$ if (μ_n, μ_n) is announced with instructions $(a_1^n, a_2^n) = (1, 0)$ and $(0, 1)$ with equal probability.*

Consider a posterior pair (μ_n, μ_2) announced where μ_2 is composite. In the continuation equilibrium, since agent 1's allocation is certain at a_1^n, agents 2-n and 2-ℓ optimally cover the shortfall from their ideal, i.e., $a_2^n = 1 - a_1^n$ and $a_2^\ell = 1 + \ell - a_1^n$ from (6). Then, for $a_1^n = 1 - E(a_2)$ to hold we must have $a_2^h - a_2^n = |\ell|$. Since $a_2^h = 1$ is constrained, it follows that $a_2^n = 1 + \ell$, $a_2^\ell = 1 + 2\ell$ and $a_1^n = -\ell$. Note that this holds even if $\mu_2 = \mu_b$; and that $a_1^n + a_2^n = 1$ which will also be the case when the posterior pair (μ_n, μ_n) is announced. Therefore, the total allocation does not change when μ_2 is decomposed, hence neither does social welfare. However, decomposing μ_2 can reduce the incentive of biased types to report neutral by introducing random specialization when both report neutral: as explained earlier, when both agents report neutral, a misreporting type would half of the time face the worst case that the other agent allocates nothing in his preferred area.

In the light of Lemmas 4 and 5, from any symmetric scheme one can keep modifying to increase the expected welfare or reduce misreporting incentives until no further such modification is possible. At the end of this process, the more biased posterior in any pair cannot be composite (Lemma 4(a) and Lemma 5) and only one pair (and its dual) survives with μ_b as the more biased posterior by Lemma 4(c). Hence, the final scheme, denoted by π^*, announces three posterior pairs (μ_n, μ_n), $(\mu_{0.5}, \mu_b)$, $(\mu_b, \mu_{0.5})$ where $\mu_{0.5} = 0.5(\mu_b + \mu_n) = (1/4, 1/2, 1/4)$, as specified below:

$$(11) \qquad \begin{cases} 1 = \pi^*(\mu_n, \mu_n|n, n) = \pi^*(\mu_b, \mu_{0.5}|b, n) = \pi^*(\mu_{0.5}, \mu_b|n, b) \\ 1/2 = \pi^*(\mu_b, \mu_{0.5}|b, b) = \pi^*(\mu_{0.5}, \mu_b|b, b). \end{cases}$$

Here, we use b to denote agents' reports $t_i \in \{h, \ell\}$ as an argument of π^*.

Thus, π^* reveals both agents to be neutral when they are (with random allocation instructions); but does not reveal when both are biased by partially pooling with the case where only one agent is biased. Partial pooling avoids the welfare loss stemming from high variances of interior equilibria when both agents are revealed to be biased. Such partial pooling requires coordination by a mediator, hence is infeasible under direct communication even with jointly controlled lottery. The total allocation, presented in Table 4, is therefore more efficient than under direct communication.

Table 4—Total allocation to A with mediation scheme π^*.

$1\backslash 2$	ℓ	n	h
ℓ	$1-h/4+7\ell/4$	$1-h/4+3\ell/4$	$1+(h+\ell)/4$ or $1+3(h+\ell)/4$
n	$1-h/4+3\ell/4$	1	$1+h/4-3\ell/4$
h	$1+3(h+\ell)/4$ or $1+(h+\ell)/4$	$1+h/4-3\ell/4$	$1+5h/4-3\ell/4$

The scheme π^* is clearly consistent, and maximizes welfare subject to truthful reporting because it is obtained by modifying any consistent scheme without reducing welfare. Hence, π^* would be the optimal scheme if it is IC. The IC conditions are strongest when the posterior (μ_n, μ_n) is announced with allocation pairs $(a_1, a_2) = (\underline{a}, \bar{a})$ and $(a_1, a_2) = (\bar{a}, \underline{a})$ with equal probabilities where $\underline{a} = (\underline{a}^\ell, \underline{a}^n, \underline{a}^h) = (0, 0, h)$ and $\bar{a} = (\bar{a}^\ell, \bar{a}^n, \bar{a}^h) = (1+\ell, 1, 1)$, which we postulate for π^*.[14] Then, so long as h and $|\ell|$ are not too far apart, π^* is IC and thus is optimal as stated in the following Proposition.

PROPOSITION 3: *The scheme π^* generates the maximum welfare among all consistent schemes conditional on truthful reporting. It is the optimal scheme if and only if $\ell \leq (1 - 2\sqrt{14})h/11 \approx -0.59h$.*

If agent 1 reports biased, he induces $(\mu_b, \mu_{0.5})$ with probability 2/3 or $(\mu_{0.5}, \mu_b)$ with probability 1/3, while if he reports neutral, he induces (μ_n, μ_n) with probability 1/3 or $(\mu_{0.5}, \mu_b)$ with probability 2/3. Hence, the biased type's incentive to misreport increases with the extra payoff he obtains in the continuation equilibrium after (μ_n, μ_n) or $(\mu_{0.5}, \mu_b)$ as compared to after $(\mu_b, \mu_{0.5})$. This is higher for ℓ-type than h-type because the randomization after (μ_n, μ_n) is less damaging for ℓ-type; and is overwhelming for ℓ-type when $h - |\ell|$ is large enough because then the damage is sufficiently small. Since n-type has no incentive to misreport as biased, it follows that π^* is IC if $|\ell|$ is not too small relative to h as stated in

[14] The probabilities do not matter with which allocation pairs $(\underline{a}, \bar{a})$ and $(\bar{a}, \underline{a})$ are announced along with the posterior (μ_n, μ_n). This is because, whatever the probabilities, each agent expect $(\underline{a}, \bar{a})$ and $(\bar{a}, \underline{a})$ with equal probability due to symmetric treatment of the two agents. However, it is critical that the instructed allocations are complete specialization of each agent in one area, because any less specialization would increase the incentive for biased agents to report neutral.

Proposition 3. The IC holds for a substantially larger range of $|\ell|$ than under direct communication with jointly controlled lottery. This is because the mediated partial pooling enhances the payoff of biased types from truthful reporting.

V. Applications

A. Coordination of Humanitarian Aid

Our results can throw light on coordination of humanitarian aid. In an emergency humanitarian organizations (HOs) allocate aid to affected areas. Coordination failure is a recurring theme in evaluations of aid delivery. There can be duplication of aid in some areas and gaps in others. Some evaluations attribute coordination failure to HOs' reluctance to share information. Our results suggest that such environments have an inherent incentive problem which results in ineffectiveness of direct communication even when the HOs are fairly aligned.

Our result can also speak to the experience of disaster management system Risepak. Risepak was developed to improve coordination in the 2005 earthquake in Pakistan. The idea was that each HO could display at a village level the amount and type of aid delivered to improve coordination of subsequent deliveries. Despite its potential, most HOs were not willing to share their information and Risepak failed to achieve critical mass. "*Although a great deal of thought was dedicated to the potential applications of Risepak once the system reached a critical mass of data, less attention seems to have been paid to creating immediate incentives for organizations to provide the information in the first place.*" (Amin 2008, p. 260). Such open access website is in effect equivalent to direct communication. Our result on ineffectiveness of direct communication provides an explanation for Risepak's experience and, by identifying the inherent incentive problem, provides a starting point for the solution.

Our result on mediated communication suggests that coordination can be improved by appropriately designed information management system. In this respect our model is too simple for practical purposes. However, we conjecture that the ability to raise welfare by employing a mechanism that filters information is robust and the reasonable restrictions we imposed on the mechanism will be helpful for further research.

There are some information management tools currently used within the Cluster Approach. Cluster Approach was introduced by the Humanitarian Reform of 2005 to improve coordination of humanitarian aid.[15] Cluster Approach divides the response to various clusters, e.g., shelter, nutrition and health, and assigns a leader organization to coordinate each cluster. The cluster leads do not have authority over the partners but they can induce coordination, e.g., by information sharing. Information management tools available include 'Who does What

[15] https://www.humanitarianresponse.info/en/about-clusters/what-is-the-cluster-approach

Where' (3W) and Humanitarian Dashboard. 3W reports the number of organizations operating in each cluster and in each district. However, the districts are large[16] and there is no information about the budgets. Humanitarian Dashboard reports the percentage of aid requirements met in each cluster in a given emergency but typically has no geographical information.[17]

The current information management tools clearly filter the information, although not necessarily optimally. Let us focus on regional allocation and 3W.[18] According to our results, an HO has an incentive to downplay its operations in its priority region and exaggerate them in its low-priority region.[19] However, since only the operational presence – but not its scale – is reported in 3W, untruthful report would be easily detected and could result in loss of reputation. The crude reporting of 3W is therefore robust to the type of gaming we have analyzed. 3W improves coordination by very rough identification of gaps and overlaps and is particularly helpful when the HOs enter the emergency sequentially or expand their operations to new regions.

Our result on mediated communication raises the question whether coordination of more fine-tuned budget allocations after entry could be improved. Budget allocations and priorities are soft information and it would be possible for the HOs to exaggerate or downplay them. Therefore HOs' reports need to be appropriately filtered to give incentives for informative communication. Further research is required to find practical insights for a more complex setup than we have analyzed in this paper.

B. *Other Applications*

Our analysis has wider applicability than coordination of humanitarian aid. There are various situations where potentially biased agents allocate funds between public goods. In a research joint venture, the partners allocate funds between different projects and may have biased preferences depending, e.g., on the importance of the potential innovation for the parent company. Public agencies (research grant councils, for example) allocate funds between different regions and projects and may have objectives that are not fully aligned. Development aid agencies allocate aid between social sectors (health and education) and other

[16]The districts in 3W are counties while Risepak was working at a village level.

[17]Cluster Approach also offers additional benefits to the humanitarian organizations, such as security and consolidated appeals for funding, which are important to ensure participation.

[18]Regional allocation (rather than allocation between different clusters as in Humanitarian Dashboard) is the relevant dimension for humanitarian organizations which typically specialize in one cluster.

[19]Nonneutrality of humanitarian aid is well established at the country level. In addition to needs, news coverage and bilateral relationship (e.g. colonial history, trade relationship, common language and geographic proximity) increase humanitarian aid (Drury, Olson and Van Belle 2005; Eisensee and Strömberg 2007; Strömberg 2007; Fink and Redaelli 2011). At regional level, spatial inertia favors regions where the humanitarian organizations have prior operations (Jayne et al. 2002). Some governments target relief aid to regions with stronger political support (Jayne et al. 2001; Plümper and Neumayer 2009; Francken, Minten and Swinnen 2012) or to more informed electorates (Besley and Burgess 2001, 2002). Furthermore, it is generally believed that NGOs locate to media hotspots as visibility and demonstrable activity are important for securing funding (Cooley and Ron 2002).

public expenditure programs (e.g. transportation) as well as regionally and may have different priorities.[20] Our main result is that even when the agents are fairly aligned, but the direction of their potential bias is unknown, coordination by direct communication fails and results in inefficient allocation of funds.

Coordination can, however, be improved by an appropriately designed information management system. The agents report their planned allocations (or alternatively, their priorities) and the system filters information so that no information is revealed about the direction of any reported priorities.[21] Appropriately designed information management system gives the agents the incentives to report truthfully and improves the allocation of funds.

VI. Conclusion

We studied two potentially biased agents providing public goods in two areas, A and B, with equal needs. When the agents do not communicate, the allocation of funding is inefficient. Our main result is that direct communication is ineffective even when the agents are fairly aligned if the direction of their potential bias is not known. An agent biased to A would represent himself as biased to B with the aim of influencing the other agent to allocate more funds to area A. We furthermore show that a mediator (or an information management system) who filters the information communicated by the agents can improve coordination even when the mediator does not have any control rights. We have analyzed allocation in a stylized setup but we conjecture that our key insights are not sensitive to the details of our model.

Our main application is in humanitarian aid. Our result on ineffective direct communication provides a novel explanation to unsuccessful coordination attempts, e.g., via an open access website. Our results on mediated communication points to the direction of a well-designed information management system to coordinate humanitarian aid. Our model is also applicable to various other situations where several agents with potentially diverse motivations allocate funds (or time and attention) to public goods, for example, in R&D joint ventures and public agencies.

An interesting direction for future research is to explore management as coordination. For instance, a manager may act as a coordinator across two divisions allocating funds between public goods for the company. The manager then has a role even when the decision rights are decentralized and the relevant information is dispersed in the company. Our model could be developed further to be more suitable for such management settings.

[20] See Halonen-Akatwijuka (2007) on coordination failure in development aid caused by asymmetric information about budget sizes.

[21] In direct revelation mechanism the agents report their types. In equivalent indirect mechanism the agents could report their planned allocations and a balanced allocation can be interpreted as neutral type and allocation in favor of A (respectively, B) as the type biased to A (respectively, B).

REFERENCES

Alonso, Ricardo, Wouter Dessein, and Niko Matouschek. 2008. "When Does Coordination Require Centralization?" *American Economic Review* 98 (1): 145-179.

Altay, Nezih, and Melissa Labonte. 2014. "Challenges in Humanitarian Information Management and Exchange: Evidence from Haiti." *Disasters* 38 (S1): S50-S72.

Amin, Samia. 2008. "Data Management Systems after the Earthquake in Pakistan: The Lessons of Risepak." in Amin, Samia, and Markus Goldstein (eds) *Data Against Natural Disasters: Establishing Effective Systems for Relief, Recovery, and Reconstruction.* The World Bank.

Banerjee, Abhijit. 2007. *Making Aid Work.* Cambridge: MIT Press.

Besley, Timothy, and Robin Burgess. 2001. "Political Agency, Government Responsiveness and the Role of the Media." *European Economic Review* 45 (4–6): 629–640.

Besley, Timothy, and Robin Burgess. 2002. "The Political Economy of Government Responsiveness: Theory and Evidence from India." *Quarterly Journal of Economics* 117 (4): 1415-1451.

Cooley, Alexander, and James Ron. 2002. "The NGO Scramble: Organizational Insecurity and the Political Economy of Transnational Action." *International Security* 27 (1): 5-39.

Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50 (6): 1431-1451.

Dessein, Wouter. 2002. "Authority and Communication in Organizations." *Review of Economic Studies* 69 (4): 811-838.

Drury, A. Cooper, Richard Stuart Olson, and Douglas A. Van Belle. 2005. "The Politics of Humanitarian Aid: U.S. Foreign Disaster Assistance, 1964-1995." *Journal of Politics* 67 (2): 454-473.

Eisensee, Thomas, and David Strömberg. 2007. "News Droughts, News Floods, and U.S. Disaster Relief." *Quarterly Journal of Economics* 122 (2): 693-728.

Fink, Gunt, and Silvia Redaelli. 2011. "Determinants of International Emergency Aid – Humanitarian Need Only?" *World Development* 39 (5): 741-757.

Francken, Nathalie, Bart Minten, and Johan F.M. Swinnen. 2012. "The Political Economy of Relief Aid Allocation: Evidence from Madagascar." *World Development* 40 (3): 486–500.

Forges, Francoise. 1986. "An Approach to Communication Equilibria." *Econometrica* 54 (6): 1375–1385.

Forges, Francoise. 1990. "Universal Mechanisms." *Econometrica* 58 (6): 1341–1364.

Goltsman, Maria, Johannes Hörner, Gregory Pavlov, and Francesco Squintani. 2009. "Mediation, Arbitration and Negotiation." *Journal of*

Economic Theory 144 (4): 1397-1420.

Goltsman, Maria, and Gregory Pavlov. 2014. "Communication in Cournot Oligopoly." *Journal of Economic Theory* 153 : 152-176.

Halonen-Akatwijuka, Maija. 2007. "Coordination Failure in Foreign Aid." *The B.E. Journal of Economic Analysis & Policy* 7 (1).

Hörner, Johannes, Massimo Morelli, and Francesco Squintani. 2015. "Mediation and Peace." *Review of Economic Studies* 82 (4): 1483-1501.

Ivanov, Maxim. 2010. "Communication via a Strategic Mediator." *Journal of Economic Theory* 145 (2), 869–884.

Jayne, Thomas S., John Strauss, Takashi Yamano, and Daniel Molla. 2001. "Giving to the Poor? Targeting of Food Aid in Rural Ethiopia." *World Development* 29 (5): 887–910.

Jayne, Thomas S., John Strauss, Takashi Yanamo, and Daniel Molla. 2002. "Targeting of Food Aid in Rural Ethiopia: Chronic Need or Inertia?" *Journal of Development Economics* 68 (2): 247-288.

Krishna, Vijay, and John Morgan. 2004. "The Art of Conversation: Eliciting Information from Experts through Multi-stage Communication." *Journal of Economic Theory* 117 (2): 147-179.

Myerson, Roger. 1982. "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems." *Journal of Mathematical Economics* 10 (1): 67-81.

Palfrey, Thomas, and Howard Rosenthal. 1991. "Testing for Effects of Cheap Talk in a Public Goods Game with Private Information." *Games and Economic Behavior* 3 (2): 183-220.

Palfrey, Thomas, Howard Rosenthal, and Nilanjan Roy. 2017. "How Cheap Talk Enhances Efficiency in Threshold Public Goods Games." *Games and Economic Behavior* 101 (1):234-259.

Plümper, Thomas, and Eric Neumayer. 2009. "Famine Mortality, Rational Political Inactivity, and International Food Aid." *World Development* 37 (1): 50–61.

Rantakari, Heikki. 2008. "Governing Adaptation." *Review of Economic Studies* 75 (4): 1257–1285.

Strömberg, David. 2007. "Natural Disasters, Economic Development, and Humanitarian Aid." *Journal of Economic Perspectives* 21 (3): 199-222.

UN Inter-Agency Standing Committee. 2010. "Response to the Humanitarian Crisis in Haiti Following the 12 January 2010 Earthquake: Achievements, Challenges and Lessons to be Learned." Available at http://www.alnap.org/resource/5941.aspx (accessed January 5, 2017).

"What is Cluster Approach." https://www.humanitarianresponse.info/en/about-clusters/what-is-the-cluster-approach (accessed January 5, 2017).

Mathematical Appendix

A1. Appendix A

In Appendix A, we prove Lemma 2 and Proposition 2 in Section III.

Proof of Lemma 2.

We have show in the main text that any equilibrium is noninterior where $a_1^n < a_2^n$ and either $a_1^\ell = 0$ or/and $a_2^h = 1$ is noninterior. It remains to show the uniqueness of equilibrium. By optimality,

$$a_1^n = 1 - \sum \mu_2^t \cdot a_2^t(a_1^n) \tag{A1}$$

where $a_2^t(a_1^n)$ is the optimal allocation a_2^t of (6) when $a_1 = (\max\{0, a_1^n + \ell\}, a_1^n, a_1^n + h)$. Given (8), the RHS of ($A1$) is nonnegative when $a_1^n = 0$, increases continuously in a_1^n at a rate strictly less than unity while a_1^n is low enough so that $a_1^n \leq |\ell|$ or $a_2^n(a_1^n) > 1 - h$ (i.e., while either a_1^ℓ or $a_2^h(a_1^n)$ is noninterior), but at a rate equal to unity for higher a_1^n such that $a_1^n < a_2^n(a_1^n)$. Moreover, it assumes $a_1^n + \mathrm{Eb}(\mu_1) - \mathrm{Eb}(\mu_2) < a_1^n$ when $a_1^n = |\ell|$ or $a_2^h(a_1^n) = 1$, whichever happens later. Therefore, there is a unique a_1^n such that $0 \leq a_1^n < a_2^n(a_1^n)$ and satisfies ($A1$), hence a unique equilibrium. Q.E.D.

Proof of Proposition 2.

We prove Proposition 2 for general prior distributions denoted by $\mu_0 = (\mu_0^\ell, \mu_0^n, \mu_0^h)$ with the property that $0 < \mathrm{Eb}(\mu_0) < h + \ell$.[22] We start with some preliminary results on allocation games under general posterior beliefs (that ensue after communication).

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.*

Proof. By Lemmas 1 and 2, a_i satisfies $a_i^\ell = \max\{0, a_i^n + \ell\} \leq a_i^n \leq a_i^h = \min\{1, a_i^n + h\}$ and similarly for a_i. The conclusions then follow straightforwardly from $E(\mathrm{a}_i|\mu_i) - E(a_i|\mu_i) = \mu_i^\ell(\mathrm{a}_i^\ell - a_i^\ell) + \mu_i^n(\mathrm{a}_i^n - a_i^n) + \mu_i^h(\mathrm{a}_i^h - a_i^h)$. ■

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) $E(a_1|\mu_1) - a_1^n = E(a_2|\mu_2) - a_2^n$.

Proof. Straightforward from (6) and Lemma 2. ■

[22]This property helps us address complications stemming from the fact that an agent's equilibrium allocation need not move in the opposite direction when the other agent's expected bias changes.

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 $\hat{a}_i^t$.*

Proof. The expected utility of agent i of t-type from $\hat{a}_i^t$ is

$$-\mu_{-i}^\ell(E(a_{-i}|\mu_{-i}) - a_{-i}^\ell)^2 - \mu_{-i}^n(E(a_{-i}|\mu_{-i}) - a_{-i}^n)^2 - \mu_{-i}^h(E(a_{-i}|\mu_{-i}) - a_{-i}^h)^2$$

independently of agent i's type t. Subtracting from this his utility when his allocation changes by y from $\hat{a}_i^t$, i.e.,

$$-\mu_{-i}^\ell(E(a_{-i}|\mu_{-i})-a_{-i}^\ell-y)^2-\mu_{-i}^n(E(a_{-i}|\mu_{-i})-a_{-i}^n-y)^2-\mu_{-i}^h(E(a_{-i}|\mu_{-i})-a_{-i}^h-y)^2,$$

we get

$$-2y\big[\mu_{-i}^\ell(E(a_{-i}|\mu_{-i})-a_{-i}^\ell)+\mu_{-i}^n(E(a_{-i}|\mu_{-i})-a_{-i}^n)+\mu_{-i}^h(E(a_{-i}|\mu_{-i})-a_{-i}^h)\big]+y^2 \;=\; y^2. \blacksquare$$

Lemma A4 *If* $\mathrm{Eb}(\mu_1) < \mathrm{Eb}(\mu_2)$ *and* $\mathrm{Eb}(\mu_1) < h + \ell$, *then a_2^h is a noninterior solution at the continuation equilibrium after (μ_1, μ_2).*

Proof. Suppose to the contrary that a_2^h is an interior solution. Then, by Lemma 5, a_1^ℓ must be noninterior, i.e., $a_1^n < |\ell|$. In such an equilibrium, we would have

$$a_2^n = 1 - E(a_1) = 1 - (1-\mu_1^\ell)a_1^n - \mu_1^h h \quad\text{and}\quad a_1^n = 1 - E(a_2) = 1 - a_2^n - \mu_2^h h - \mu_2^\ell \ell.$$

Solving these simultaneous equations, we get

$$a_2^n = \frac{h(\mu_2^h(1-\mu_1^\ell) - \mu_1^h) + \ell\mu_2^\ell(1-\mu_1^\ell) + \mu_1^\ell}{\mu_1^\ell}; \quad a_1^n = \frac{h(\mu_1^h - \mu_2^h) - \ell\mu_2^\ell}{\mu_1^\ell}$$

Thus,

$$\begin{aligned} a_2^n + h - 1 &= \frac{h(\mu_2^h(1-\mu_1^\ell) - \mu_1^h + \mu_1^\ell) + \ell\mu_2^\ell(1-\mu_1^\ell)}{\mu_1^\ell} \\ &= \frac{\mathrm{Eb}(\mu_2)(1-\mu_1^\ell) + h\mu_1^\ell - h\mu_1^h}{\mu_1^\ell} \\ &> \frac{\mathrm{Eb}(\mu_1)(1-\mu_1^\ell) + h\mu_1^\ell - h\mu_1^h}{\mu_1^\ell} \;=\; h + \ell - \mathrm{Eb}(\mu_1) \;>\; 0, \end{aligned}$$

contradicting the supposition that a_2^h is interior. This completes the proof. ■

Returning to the task of proving Proposition 2, recall that $M_i = \{m_{i1}, m_{i2}, \cdots, m_{iK_i}\}$ is 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

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

In the case that $\mathrm{Eb}(\mu_{11}) = \mathrm{Eb}(\mu_{21})$, label agents so that $\mu^h_{11} > 0$ if possible; else, i.e., if $\mu^h_{11} = \mu^h_{21} = 0$ then label agents so that a^h_1 is interior in the continuation equilibrium after (μ_{11}, μ_{21}), which is possible due to $(*)$.

If $\mathrm{Eb}(\mu_{11}) = \mathrm{Eb}(\mu_0)$, then $\mathrm{Eb}(\mu_{11}) = \mathrm{Eb}(\mu_{1K_1})$ and $\mathrm{Eb}(\mu_{11}) = \mathrm{Eb}(\mu_{21}) = \mathrm{Eb}(\mu_0) = \mathrm{Eb}(\mu_{2K_2})$, thus the equilibrium outcome is equivalent to the case of no communication by Lemma 1. Therefore, below we assume

$$(A2) \qquad \mathrm{Eb}(\mu_{11}) < \mathrm{Eb}(\mu_0) < \mathrm{Eb}(\mu_{1K_1}).$$

To facilitate exposition, we state a lemma on comparing messages for different types.

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}.*

Proof. Recall that an n-type agent always gets his unconstrained optimum. For each message of the other agent, therefore, the net gain of agent i from sending $m_{ik'}$ rather than m_{ik} for t-type is the same as that for n-type if t-type gets unconstrained optimum after $m_{ik'}$, but is smaller otherwise. Thus, if the t-type weakly prefers sending $m_{ik'}$ to m_{ik} then an n-type must strictly prefer sending $m_{ik'}$. ■

Note that the equilibrium level of a^h_1 is an interior solution (i.e., the unconstrained optimum) under the posterior (μ_{11}, μ_{2k}) for any $k \in \{1, \cdots, K_2\}$ by Lemmas 1 and 2, but is noninterior under (μ_{1K_1}, μ_{21}) by Lemma A4. As $\mu^h_{1K_1} > 0$ from $\mathrm{Eb}(\mu_{1K_1}) > \mathrm{Eb}(\mu_0) > 0$, agent 1 of h-type weakly prefers sending m_{1K_1} to m_{11}. Hence, by Lemma A5, agent 1 of n-type should strictly prefer sending m_{1K_1} to m_{11}, implying $\mu^n_{11} = 0$.

In addition, agent 1 of ℓ-type obtains his unconstrained optimum after sending m_{1K_1} if agent 2 sent m_{2k} such that $\mathrm{Eb}(\mu_{2k}) < \mathrm{Eb}(\mu_{1K_1})$ by Lemma 2. Thus, if there is no message $m_{2\kappa} \in M_2$ such that $\mathrm{Eb}(\mu_{2\kappa}) \geq \mathrm{Eb}(\mu_{1K_1})$ and agent 1's net benefit of sending m_{1K_1} rather than m_{11} conditional on $\mu_2 = \mu_{2\kappa}$ is strictly lower for ℓ-type than for n-type, then agent 1's unconditional net benefit of sending m_{1K_1} rather than m_{11} is no lower for ℓ-type than for n-type by Lemma A3. As agent 1 of n-type strictly prefers sending m_{1K_1} to m_{11} as argued above, this would imply that so does ℓ-type, and thus that $\mu^\ell_{11} = \mu^n_{11} = 0$, contradicting $\mathrm{Eb}(\mu_{11}) < \mathrm{Eb}(\mu_0)$. Therefore,

[1] for some message $m_{2\kappa}$ such that $\mathrm{Eb}(\mu_{2\kappa}) \geq \mathrm{Eb}(\mu_{1K_1})$, agent 1's net benefit of sending m_{1K_1} rather than m_{11} conditional on $\mu_2 = \mu_{2\kappa}$ is strictly lower

for ℓ-type than for n-type.

Fix such a message $m_{2\kappa}$ and consider the continuation equilibrium allocation $a_1 = (a_1^\ell, a_1^n, a_1^h)$ and $a_2 = (a_2^\ell, a_2^n, a_2^h)$ under $(\mu_{11}, \mu_{2\kappa})$. Then, $a_1^n < a_2^n$ by Lemma 2.

Let $(\tilde{a}_1, \tilde{a}_2)$ denote the equilibrium under $(\mu_{1K_1}, \mu_{2\kappa})$. By [1] above, $\tilde{a}_1^n < a_1^n$ and $\tilde{a}_1^\ell = 0$ must hold. Hence, from (6),

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

where the second inequality is from Lemma A1. Thus,

$$(A3) \qquad E(a_1|\mu_{11}) - a_1^n \ \geq \ E(\tilde{a}_1|\mu_{1K_1}) - \tilde{a}_1^n \ \geq \ E(a_1|\mu_{1K_1}) - a_1^n \ > \ \text{Eb}(\mu_0)$$

where the second inequality is from Lemma A1.

At this point, we show that $\text{Eb}(\mu_{21}) \leq \text{Eb}(\mu_{12})$. With a view to reaching a contradiction, suppose otherwise. Then, $\mu_{12}^n = 0$ by the same reasoning that led to $\mu_{11}^n = 0$ (cf. Lemma ma A5). Thus, we may assume $\text{Eb}(\mu_{11}) < \text{Eb}(\mu_{12}) < \text{Eb}(\mu_{21})$ because messages m_{11} and m_{12} may be identified if $\text{Eb}(\mu_{11}) = \text{Eb}(\mu_{12})$ by $(*)$. From the formula (6) it can then be verified that for each m_{2k}, a_2^n is higher in the continuation equilibrium after (m_{11}, m_{2k}) than in that after (m_{12}, m_{2k}). This would mean that agent 2's allocation has lower variance after (m_{11}, m_{2k}) than after (m_{12}, m_{2k}), hence that agent 1 of h-type would strictly prefer sending m_{11} to m_{12}. As this would contradict $\text{Eb}(\mu_{11}) < \text{Eb}(\mu_{12})$, we deduce that $\text{Eb}(\mu_{21}) \leq \text{Eb}(\mu_{12})$ must hold.

Given this, we now show that either

[2] agent 2's net benefit of sending $m_{2\kappa}$ rather than m_{21} conditional on $\mu_1 = \mu_{11}$ is weakly larger for h-type than for n-type, or

[2'] a_2^h is noninterior in the continuation equilibrium after (m_{12}, m_{21}), which is possible only if $\text{Eb}(\mu_{21}) = \text{Eb}(\mu_{12})$.

With a view to reaching a contradiction, suppose otherwise. Then, agent 2's unconditional net benefit of sending $m_{2\kappa}$ rather than m_{21} is strictly larger for n-type than for h-type, and the net benefit is no lower for ℓ-type than n-type. Since $\mu_{2\kappa}^h > 0$, implying that agent 2 of h-type weakly prefers sending $m_{2\kappa}$ to m_{21}, it would follow that agent 2 of both ℓ-type and n-type strictly prefer $m_{2\kappa}$ to m_{21}, contradicting $\text{Eb}(\mu_{21}) < \text{Eb}(\mu_0)$.

The remaining part of the proof differs slightly depending on whether [2] holds or not. If [2] holds, let $(\breve{a}_1, \breve{a}_2)$ denote the continuation equilibrium after (μ_{11}, μ_{21}). By Lemma A3 and [2] above, $a_2^n \leq \breve{a}_2^n$ must hold. From (6) and Lemma A1, we deduce $0 \leq \breve{a}_2^n - a_2^n = E(a_1|\mu_{11}) - E(\breve{a}_1|\mu_{11}) \leq a_1^n - \breve{a}_1^n = E(\breve{a}_2|\mu_{21}) - E(a_2|\mu_{2\kappa})$ and thus,

$$(A4) \qquad E(a_2|\mu_{2\kappa}) - a_2^n \leq E(\breve{a}_2|\mu_{21}) - \breve{a}_2^n < \text{Eb}(\mu_{21}) \leq \text{Eb}(\mu_0)$$

where the second inequality follows from Lemma A2-(b). But, $(A3)$ and $(A4)$ cannot hold at the same time by Lemma A2-(c). This proves that $(A2)$ is not viable.

Finally, if [2] does not hold, then [2'] should hold. Then, by an analogous argument, [1] should hold with m_{11} replaced by m_{12} and consequently, so should $(A3)$ with μ_{11} replaced by μ_{12} and (a_1, a_2) representing the continuation equilibrium under $(\mu_{12}, \mu_{2\kappa})$. In addition, by the same reasoning underlying [2], it follows that

[2"] agent 2's net benefit of sending $m_{2\kappa}$ rather than m_{21} conditional on $\mu_1 = \mu_{12}$ is weakly larger for h-type than for n-type.

This further leads to $(A4)$ which is incompatible with $(A3)$, thus establishing that $(A2)$ is not viable. The proof of Proposition 2 is complete. Q.E.D.

A2. Appendix B

In Appendix B, we provide deferred proofs in Section IV.

Proof of Lemma 4.

(*a*) Let $\mu_i = \beta_i \mu_b + (1 - \beta_i)\mu_n$ for $i = 1, 2$, where $0 < \beta_1 \le \beta_2 < 1$. The continuation equilibrium for the posterior (μ_1, μ_2) is straightforwardly calculated from Section III as

$$(A5) \qquad a_1^n = \frac{(h+\ell)\beta_1(2-\beta_2)}{2\beta_2} - \ell \quad \text{and} \quad a_2^n = 1 + \ell - \frac{(h+\ell)\beta_1}{\beta_2}$$

with $a_i^t = a_i^n + t$ for $t \in \{h, \ell\}$ and $i = 1, 2$, except $a_2^h = 1$. Note that this includes the case $\beta_1 = \beta_2$ when the continuation equilibrium is interior. From this, the expected welfare conditional on (μ_1, μ_2) having been announced is routinely calculated as

$$W_0 = \frac{\beta_1^2(\beta_2 - 2)(h+\ell)^2 - 2\beta_1\beta_2(h^2 - 2h\ell - \ell^2) - 4\beta_2^2\ell^2}{4\beta_2}$$

Now suppose μ_2 is decomposed into $\mu_n = (0, 1, 0)$ and $\mu_b = (1/2, 0, 1/2)$. The continuation equilibrium after (μ_1, μ_n) is clearly $a_2^n = -\ell$ and $a_1^t = 1 + \ell + t$ for $t \in T$, and that after (μ_1, μ_b) is $(A5)$ when $\beta_2 = 1$, again with $a_i^t = a_i^n + t$ for $t \in \{h, \ell\}$ and $i = 1, 2$, except $a_2^h = 1$. Thus, the expected welfare conditional on either (μ_1, μ_n) or (μ_1, μ_b) is announced is calculated as

$$W_S = \frac{-\beta_1^2\beta_2(h+\ell)^2 - \beta_1(4\ell^2 + \beta_2(2h^2 - 4h\ell - 6\ell^2)) - 4\beta_2\ell^2}{4}.$$

Therefore,

$$W_S - W_0 = \frac{\beta_1(1-\beta_2)(h+\ell)\big[2\beta_2(h-3\ell)+\beta_1(2+\beta_2)(h+\ell)\big]}{4\beta_2} \;>\; 0$$

where the inequality ensues because, given $\beta_1 \leq \beta_2$, the expression inside the bracket decreases in ℓ reaching the lowest value of $h(2\beta_2 + \beta_1(2+\beta_2)) > 0$ at $\ell = 0$.

(b) The continuation equilibrium for the posterior $(\mu_1, \mu_2) = (\mu_1, \mu_b)$ is ($A5$) when $\beta_2 = 1$, with $a_i^t = a_i^n + t$ for $t \in \{h, \ell\}$ and $i = 1, 2$, except $a_2^h = 1$. From this, the expected welfare conditional on (μ_1, μ_b) is announced is calculated as

$$W_1(\beta_1) \;=\; \frac{2h\ell\beta_1(2-\beta_1) - \ell^2(4-2\beta_1+\beta_1^2) - h^2\beta_1(2+\beta_1)}{4}.$$

After decomposition, the continuation equilibrium after (μ_n, μ_b) is clearly $a_1^n = -\ell$ and $a_2^t = 1 + \ell + t$ for $t \in T$, and that after (μ_b, μ_b) is as in Lemma 1 of the main text when $\mu_i^h = \mu_i^\ell = 1/2$. Thus, the expected welfare conditional on either (μ_1, μ_n) or (μ_1, μ_b) is announced is calculated as $-(1-\beta_1)\ell^2 - \beta_1(3h^2 - 2h\ell + 3\ell^2)/4$. Subtracting this from $W_1(\nu_1)$, we get $(h+\ell)^2\beta_1(1-\beta_1)/4 > 0$.

(c) Let $p = \sum_t \pi(\mu_1, \mu_b|t)$ and $\widetilde{p} = \sum_t \pi(\widetilde{\mu}_1, \mu_b|t)$ denote the ex-ante probability that (μ_1, μ_b) and $(\widetilde{\mu}_1, \mu_b)$ announced by π. When the two are merged, it forms a posterior pair $(\hat{\mu}_1, \mu_b)$ where $\hat{\mu}_1 = (p\mu_1 + \widetilde{p}\widetilde{\mu}_1) / (p + \widetilde{p})$. Hence, subtracting the initial total expected welfare from that after merging, we get

$$(p+\widetilde{p})W_1\Big(\frac{p\beta_1 + \widetilde{p}\widetilde{\beta}_1}{p+\widetilde{p}}\Big) - pW_1(\beta_1) - \widetilde{p}W_1(\widetilde{\beta}_1) \;>\; 0$$

where the inequality ensues because $W_1(\cdot)$ in strictly concave as verified by its second derivative being $-(h+\ell)^2/2$. ■

Proof of Lemma 5.

The continuation equilibrium after $(\mu_1, \mu_2) = (\mu_n, \mu_2)$ is

$$a_1^n = -\ell, \quad a_2^h = 1, \quad a_2^n = 1 + \ell, \quad a_2^\ell = 1 + 2\ell,$$

which is also the continuation equilibrium after (μ_n, μ_b), while that after (μ_n, μ_n) is clearly any (a_1^n, a_2^n) such that $a_1^n + a_2^n = 1$. Thus, it is evident that the total allocation to area A is the same for each possible type profile under (μ_1, μ_2) and when μ_2 is separated, establishing the equivalence of the expected welfare.

To check IC for each agent, we need to consider the dual posterior as well because both agents are equally likely to play the role of agent 1 and agent 2 after reporting. Conditional on (μ_1, μ_2) is announced, each agent of all types derive the same expected payoff $u_1 = -\beta_2\ell^2$ as agent 1, but as agent 2 types ℓ and n derive a payoff of 0 while type h derives $-(h+\ell)^2$. By reporting $r = h$ or $r = \ell$, therefore, an agent derives the payoff of agent 2 with 1/2 of the probability $\left[\nu_0(1-\nu_0)\pi(\mu_1, \mu_2|n, b)\right]/\left[\beta_0(1-\beta_0)+\beta_0^2\right] = (1-\beta_0)\pi(\mu_1, \mu_2|n, b)$ where $\beta_0 = 1-\mu_0^n = 2/3$. By reporting $r = n$, however, he derives agent 1's payoff with 1/2 of the probability $(1-\beta_0)\pi(\mu_1, \mu_2|n, n) + \beta_0\pi(\mu_1, \mu_2|n, b)$ and agent 2's payoff with 1/2 of the probability $(1-\beta_0)\pi(\mu_1, \mu_2|n, n)$. Hence, the net benefit of reporting n rather than $r \neq n$ is 1/2 of

$$-\left[(1-\beta_0)\pi(\mu_1, \mu_2|n, n)+\beta_0\pi(\mu_1, \mu_2|n, b)\right]\beta_2\ell^2+(1-\beta_0)\left[\pi(\mu_1, \mu_2|n, n)-\pi(\mu_1, \mu_2|n, b)\right]u_2^t$$

where $u_2^\ell = u_2^n = 0$ and $u_2^h = -(h+\ell)^2$.

Upon decomposition, conditional on (μ_1, μ_b), each agent of all types derive the same expected payoff $-\ell^2$ as agent 1, but as agent 2 types ℓ and n derive a payoff of 0 while type h derives $-(h+\ell)^2$. Conditional on $(\mu_1, \mu_n) = (\mu_n, \mu_n)$, each agent's payoff is $-t^2/2$ depending on type $t \in \{\ell, n, h\}$.

By reporting $r = h$ or $r = \ell$, therefore, an agent derives 0 if $t \neq h$ and $-(h+\ell)^2$ if $t = h$ with 1/2 of the probability $(1-\beta_0)\pi(\mu_1, \mu_2|n, b)$. By reporting $r = n$, however, he derives $-t^2/2$ of the probability $(1-\beta_0)\pi(\mu_1, \mu_2|n, n)$ and $-\ell^2$ with probability 1/2 of $\beta_0\pi(\mu_1, \mu_2|n, b)$ Hence, the net benefit of reporting n rather than $r \neq n$ is 1/2 of

$$-(1-\beta_0)\pi(\mu_1, \mu_2|n, n)t^2 - \beta_0\pi(\mu_1, \mu_2|n, b)\ell^2 - (1-\beta_0)\pi(\mu_1, \mu_2|n, b)u_2^t.$$

Subtracting this net benefit of reporting n from that before separation, we get $(1-\beta_2)(\pi_{n,n}+2\pi_{n,b})\ell^2/6 > 0$, $\big(2\pi_{n,b}-\beta_2(\pi_{n,n}+2\pi_{n,b})\big)\ell^2/6 > 0$, $\big((1-\beta_2)(\pi_{n,n}+2\pi_{n,b}) - 2\pi_{n,n})\ell^2 - 2\pi_{n,n}h\ell\big)/6 > 0$ for types $t = \ell$, n and h, respectively, where $\pi_{n,n} = \pi(\mu_1, \mu_2|n, n)$ and $\pi_{n,b} = \pi(\mu_1, \mu_2|n, b)$. The signs are verified as above because $1-\beta_2 = \pi_{n,n}/\left(\pi_{n,n}+\pi_{n,b}\right)$. Consequently, the incentive to misreport is lower for $t = h, \ell$, but higher for $t = n$, after decomposition. ■

Proof of Proposition 3.

From any symmetric and consistent scheme, one can (weakly) increase the welfare conditional on truthful reporting, by decomposing according to Lemmas 4(a) and 5 and merge according to Lemma 4(c) until no such modification is possible. Let π^* be the resulting scheme which is clearly consistent. Any posterior profile $\boldsymbol{\mu} = (\mu_1, \mu_2)$ in the support of π^* must satisfy (i) at most one posterior (i.e., not both μ_1 and μ_2) is composite, (ii) there is a unique posterior, say μ', such that if $\mu_i = \mu_b$ then $\mu_{-i} = \mu'$, and (iii) if $\mu_i = \mu_n$ then $\mu_{-i} \in \{\mu_n, \mu_b\}$.

Note that μ' in (ii) is a composite posterior because when one agent is biased the other agent may be both biased and unbiased. Then, (iii) implies that if $\mu_i = \mu_n$ then $\mu_{-i} = \mu_n$. In conjunction with symmetry, it further follows that if $\mu_i = \mu_b$ then agent $-i$ is equally likely to be biased and not, i.e., $\mu_{-i} = 0.5\mu_b + 0.5\mu_n$. This verifies (11).

Therefore, π^* defined in (11) maximizes the welfare among all consistent schemes conditional on truthful reporting, because any other consistent (and symmetric) scheme can be modified to π^* without reducing welfare.

It remains to show when π^* is optimal, which amounts to delineating when π is IC. When both agents are neutral, the probabilities do not matter with which allocation pairs $(\underline{a}, \bar{a})$ and $(\bar{a}, \underline{a})$ are announced (along with the posterior (μ_n, μ_n)) where $\underline{a} = (0, 0, h)$ and $\bar{a} = (1 + \ell, 1, 1)$. This is because, whatever the probabilities, each agent expect $(\underline{a}, \bar{a})$ and $(\bar{a}, \underline{a})$ with equal probability due to symmetric treatment of the two agents. However, it is optimal that the instructed allocations are complete specialization of each agent in one area, because any less specialization would increase the incentive for biased agents to report neutral.

Consider agent i of type ℓ. By reporting $r = \ell$ (or h), he obtains an expected payoff u_2^ℓ after $(\mu_{0.5}, \mu_b)$ as agent 2 with probability 2/3, and an expected payoff of u_1^ℓ after $(\mu_{0.5}, \mu_b)$ as agent 1 with probability 1/3, hence an overall expected payoff of $(u_1^\ell + 2u_2^\ell)/3$. By untruthfully reporting $r = n$, on the other hand, he obtains an expected payoff of $-\ell^2/2$ after (μ_n, μ_n) with probability 1/3, and either u_1^ℓ after $(\mu_{0.5}, \mu_b)$ or u_2^ℓ after $(\mu_b, \mu_{0.5})$ with the remaining probability, hence an overall expected payoff of $(-\ell^2/2 + 2u_1^\ell)/3$. Hence, the net gain from untruthful reporting for an ℓ-type agent is

$$\frac{-\ell^2/2 + u_1^\ell - 2u_2^\ell}{3} = \frac{5h^2 + 2h\ell - 11\ell^2}{48}.$$

In an analogous manner we calculate the net gain from untruthful reporting for an h-type agent as

$$\frac{-h^2/2 + u_1^h - 2u_2^h}{3} = \frac{5h^2 + 18h\ell + 5\ell^2}{48}.$$

and that for an n-type as

$$\frac{u_1^n - 2u_2^n}{3} = \frac{5h^2 + 2h\ell - 3\ell^2}{48}.$$

Thus, it is routinely verified that n-type has no incentive to report untruthfully always, h-type if $\ell \leq -(9 - 2\sqrt{14})h/5 \approx -0.3h$, and ℓ-type if $\ell \leq (1 - 2\sqrt{14})h/11$, completing the proof. Q.E.D.

Remark on extending Proposition 3: Note that Lemmas 4 and 5 concern consistent posterior pairs that may be announced subject to (c1) and (c2), hence continue to hold for prior beliefs μ_0 even if $\mu_0^n \neq 1/3$ so long as $\mu_0^h = \mu_0^\ell$. In addition, Lemmas 4 and 5 continue to hold for prior beliefs such that μ_0^h and μ_0^ℓ are close enough, due to continuity of continuation equilibrium in the posterior profile.[23] Therefore, Proposition 3 extends to an open set of prior beliefs that contains all μ_0 such that $\mu_0^h = \mu_0^\ell \in (0, 1/2)$, with an appropriately modified range of ℓ for which π^* is an optimal scheme.

[23]For Lemma 5, it is straightforward to verify that the continuation equilibrium after (μ_n, μ_2) coincides with that after (μ_n, μ_b) even if $\mu_0^h \neq \mu_0^\ell$.