Supplementary Material
Why Echo Chambers are Useful

Ole Jann — CERGE-EI

Christoph Schottmüller — University of Cologne and TILEC

January 27, 2026

Contents

1. Bipolar polarization – step-by-step derivation of the results (including proof theorem 2)

We denote by n_0 (n_b) the number of people with bias 0 (b) and without loss of generality we let $n_0 \geq n_b$.

1.1. Segregation and full information as equilibrium

First, we ask the question when segregation, i.e. all players with bias $b_i = 0$ choosing room 1 and all players with bias $b_i = b$ choosing room 2, is an equilibrium. Clearly, every player is truthtelling in the most informative messaging equilibrium in this case and player i's expected payoff is

$$U_i^{fs} = -\alpha n_{-i} b^2 - 1/4\left[n + \alpha(n-1)n\right] + (1/4 - p(1-p))\left[(1-\alpha)n_i + \alpha n_i^2 + \alpha n_{-i}^2\right] \quad (1)$$

where n_i is the number of players with the same bias as player i and n_{-i} is the number of players with the other bias.

Proposition 3. *A segregation equilibrium exists if and only if one of the two following condition is met: (i) $b > n_0(p-1/2)$, (ii) $p-1/2 < b \leq n_0(p-1/2)$ and $n_0 \leq (1+\alpha)(n_b-1)$.*

Proof of proposition 3: Consider the incentives of player i to unilaterally deviate in his room choice in a segregation situation. There are three relevant cases that we will consider in turn: (i) after the deviation everyone including i still sends truthful messages, (ii) after the deviation i will babble but the players in the room he is switching to still send truthful messages and (iii) i's room switch leads to babbling by all players in the room he is switching to.

First, truthful messages after switch. This occurs if and only if $b \leq p - 1/2$. If i's bias group is the (weakly) smaller group, then i will benefit from a room switch in this case, see equation (1). Hence, segregation does not exist as equilibrium if b is less than $p - 1/2$.

Second, i babbles after the switch but everyone else remains truthtelling. This happens if and only if $p - 1/2 < b \leq n_{-i}(p - 1/2)$. In this case, i's deviation payoff is

$$U_i^d = -\alpha n_{-i} b^2 - 1/4\left[n + \alpha(n-1)n\right] + (1/4 - p(1-p))\left[1 + n_{-i} + \alpha n_{-i}^2 + \alpha(n_i - 1)^2\right]$$

which is higher than U_i^{fs} if and only if $n_{-i} > (1+\alpha)(n_i - 1)$. Clearly, the members of the smaller group are the ones for who this constraint is more stringent. That is, a segregation equilibrium exists given $p - 1/2 < b \leq n_{-i}(p - 1/2)$ if and only if n_0 is at most $1+\alpha(n_b-1)$. Intuitively, members of the smaller group have a lot of information to gain form a switch if the larger group is very large. Hence, a segregation equilibrium only exists if the larger group is not too large.

Third, the switch leads to complete babbling in the room switched to. This will occur if and only if $b > n_{-i}(p-1/2)$. In this case, it is obvious that the deviation is unprofitable and a segregation equilibrium exists. □

Next we check for which parameter values a full information equilibrium, i.e. all players choosing the same room and reporting their signal truthfully, exists. Then, truthful reporting is an equilibrium of the messaging game if and only if $b(1-(n_b-1)/(n-1)) \leq p-1/2$ which is equivalent to $bn_0/(n-1) \leq p-1/2$. It is obvious that unilateral deviations in room choice from a fully integrated room are not profitable whenever truthtelling is by all players is an equilibrium in this room. We state this below as formal result and illustrate the results on full segregation and full information in figure 1.[1]

Proposition 4. *A full information equilibrium exists if and only if* $bn_0/(n-1) \leq p-1/2$.

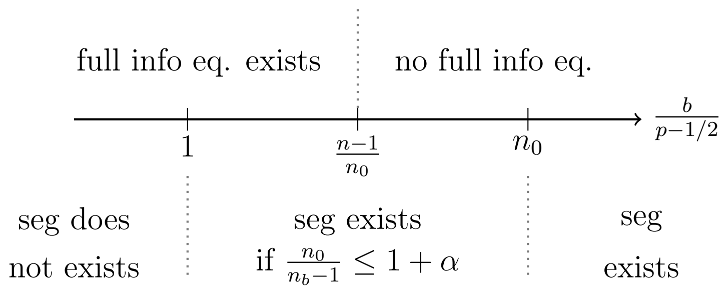

Figure 1: Segregation and full information as room choice equilibria

1.2. Intermediate results on welfare optimal room allocation

We will now consider how a planner would assign players to rooms in order to maximize the sum of players' payoffs. The planner's only tool is room assignment knowing the players' biases, i.e. the planner does not observe signals and cannot influence the messages sent or actions taken by the players.

From proposition 1, it is clear that the planner's objective is equivalent to maximizing $\sum_i \zeta_i$, i.e. the total number of pieces of information by all players. We proceed in a number of lemmas. To avoid case distinction, we use the convention that a player who is alone in a room will send a truthful message.

Lemma 2. *There is a welfare optimal room assignment without a room in which all players babble.*

[1] Note that $(n-1)/n_0 \leq n_0$ as $n_0 \geq n/2$ and $n \geq 2$.

Proof of lemma 2: In such a room there would be at least one player of each bias type. Splitting the room into two according to bias type would lead to weakly more transmitted pieces of information. □

Lemma 3. *Assume there are two rooms R_1 and R_2 such that in equilibrium players with bias $x \in \{0, b\}$ send truthful messages in both rooms. Then players with bias x will also send truthful messages in a merged room $R_1 \cup R_2$.*

Proof of lemma 3: As players send truthful messages by assumption in room R_i for $i \in \{1, 2\}$,

$$\frac{b}{p - 1/2} n_{R_i,-x} \leq n_{R_i} - 1$$

has to hold where $n_{R_i,-x}$ is the number of players in room R_i who do not have bias x. Summing this inequality over the two rooms $i = 1, 2$ yields

$$\frac{b}{p - 1/2} \left(n_{R_1,-x} + n_{R_2,-x}\right) \leq n_{R_1} + n_{R_2} - 2$$

which is sufficient for

$$\frac{b}{p - 1/2} \left(n_{R_1,-x} + n_{R_2,-x}\right) \leq n_{R_1} + n_{R_2} - 1.$$

The latter inequality implies that truthful messages in room $R_1 \cup R_2$ are optimal for players with bias x. □

Corollary 3. *The welfare maximizing room assignment will not include rooms R_1 and R_2 such that either*

- *both R_1 and R_2 are both populated exclusively by players with bias $x \in \{0, b\}$, or*
- *both R_1 and R_2 are populated by players with both biases and all players send truthful messages, or*
- *R_1 and R_2 are populated by players with both biases and only the players with bias $x \in \{0, b\}$ send truthful messages, or*
- *R_1 is exclusively populated by players with bias $x \in \{0, b\}$ and R_2 is populated by players with both biases but only players with bias x send truthful messages in R_2.*

Proof of corollary 3: In each of these cases, merging the two rooms maintains truthtelling incentives for those that originally sent truthful messages by lemma 3. As the truthful messages are received by more players, merging clearly increases the planner's objective. □

Lemma 4. *The welfare maximizing room assignment will not include rooms R_1 and R_2 such that both rooms contain players of each bias and all players in R_1 send truthful messages while only players of bias $x \in \{0, b\}$ send truthful messages in R_2.*

Proof of lemma 4: First, we consider the case where more players are in room R_1, i.e. $n_{R_1,x} + n_{R_1,-x} \geq n_{R_2,x} + n_{R_2,-x}$. Note that the number of pieces of information generated in those two rooms is $n_{R1}^2 + n_{R_2,x}(n_{R_2,x} + n_{R_2,-x})$. Note that $n_{R2,x} > n_{R_2,-x}$ as otherwise x would not be truthtelling in R_2 while $-x$ is not. Consider now an alternative room assignment that differs from the original one in the way that $n_{R2,-x}$ players of each bias are moved from R_2 to R_1. Denote everythign after the change using $\tilde{\cdot}$. That is, $\tilde{R}_2$ will contain $n_{R_2,x} - n_{R_2,-x}$ players of bias x and none of bias $-x$ while $\tilde{R}_1$ will contain $n_{R_1,x} + n_{R_2,-x}$ players of bias x and $n_{R_1,-x} + n_{R_2,-x}$ of type $-x$. The crucial result is that all players in room $\tilde{R}_1$ find truthtelling optimal: As truthtelling was optimal for players of both biases by assumption in R_1,

$$\frac{b}{p - 1/2} \leq \frac{n_{R_1} - 1}{\max\{n_{R_1,x}, n_{R_1,-x}\}}.$$

Now note that

$$\frac{n_{R_1} - 1}{\max\{n_{R_1,x}, n_{R_1,-x}\}} \leq \frac{n_{R_1} - 1 + 2n_{R_2,-x}}{\max\{n_{R_1,x} + n_{R_2,-x}, n_{R_1,-x} + n_{R2,-x}\}}$$

as the latter fraction is increasing in $n_{R_2,-x}$ and therefore truthtelling is still optimal in $\tilde{R}_1$. Clearly, truthtelling is also optimal in $\tilde{R}_2$ as only players with bias x are left there. Consequently, the total number of pieces of information in $\tilde{R}_1$ and $\tilde{R}_2$ is $(n_{R_1} + 2n_{R_2,-x})^2 + (n_{R_2,x} - n_{R_2,-x})^2$ which, by $n_{R_1} \geq n_{R_2} > n_{R_2,x}$, is strictly greater than the number of pieces of information generated by rooms R_1 and R_2. Hence, the planner prefers the room assignment $\tilde{R}_1$, $\tilde{R}_2$ over the room assignment R_1 and R_2 (keeping room assignment for players not in those rooms fixed).

Second, consider the case where more players are in room R_2, i.e. $n_{R_1,x} + n_{R_1,-x} < n_{R_2,x} + n_{R_2,-x}$. In this case, we argue that merging the two rooms to $R_1 \cup R_2$ will yield more information than keeping them separate. By lemma 3, players with bias x will still be truthtelling in room $R_1 \cup R_2$. Assume that players with bias $-x$ will not tell the truth in $R_1 \cup R_2$ (otherwise merging the two rooms is clearly optimal). Note that this implies $n_{R_1,x} + n_{R_2,x} > n_{R_1,-x}, n_{R_2,-x}$ as players with bias x tell the truth in $R_1 \cup R_2$ and players with bias $-x$ do not. The number of pieces of information generated in $R_1 \cup R_2$ is then $(n_{R_1,x} + n_{R_2,x})(n_{R_1,x} + n_{R_2,x} + n_{R_1,-x} + n_{R_2,-x})$. This is greater than the number of pieces of information generated in R_1 and R_2 separately as

$$n_{R_1,x}^2 + 2n_{R_1,x}n_{R_2,x} + n_{R_1,x}n_{R_1,-x} + n_{R_1,x}n_{R_2,-x} + n_{R_2,x}^2 + n_{R_2,x}n_{R_1,-x} + n_{R_2,x}n_{R_2,-x}$$
$$> n_{R_1,x}^2 + 2n_{R_1,x}n_{R_1,-x} + n_{R_1,-x}^2 + n_{R_2,x}^2 + n_{R_2,x}n_{R_2,-x}$$

$$\Leftrightarrow 2n_{R_1,x}n_{R_2,x} - n_{R_1,x}n_{R_1,-x} + n_{R_1,x}n_{R_2,-x} + n_{R_2,x}n_{R_1,-x} - n_{R_1,-x}^2 > 0$$

$$\Leftrightarrow -n_{R_1,-x}n_{R_1} + n_{R_1,x}n_{R_2} + n_{R_2,x}n_{R_1} > 0$$

which holds true by $n_{R_1} < n_{R_2}$ and $n_{R_1,x} + n_{R_2,x} > n_{R_1,-x}, n_{R_2,-x}$. Consequently, the planner would prefer $R_1 \cup R_2$ to R_1 and R_2 separately. □

Lemma 5. *In a welfare maximizing room allocation, the following cannot occur: There are three rooms R_0 populated only of players with bias 0, R_b populated only with players of bias b and R_m populated with players of both biases.*

Proof of lemma 5: If R_m induces babbling, it is clearly better to assign the bias 0 (b) players in there to R_0 (R_b) instead. If only players with bias $x \in \{0, b\}$ send truthful messages in R_m, then it is clearly better to merge this room with R_x which maintains truthtelling incentives for players with bias x, see lemma 3.

Hence, we only have to consider the case where players with both biases send truthful messages in R_m. For concreteness assume there are more players in R_0 than in R_b denoted by $n_{R_0} \geq n_{R_b}$ (the reverse case is analyzed analogously). We consider two cases in turn. First, consider $n_{R_0} \geq n_{R_m} + n_{Rb}$. We claim that merging R_m and R_0 will then lead to more information than keeping these rooms separate. By lemma 3, player with bias 0 will still be truthtelling in $R_0 \cup R_m$ and the number of pieces of information generated by rooms R_b and $R_0 \cup R_m$ is at least $n_{R_b}^2 + (n_{R_0}+1)(n_{R_m}+n_{R_0})$ which is strictly higher than the number of pieces of information generated by R_0, R_b and R_m, i.e. $n_{R_b}^2 + n_{R_0}^2 + n_{R_m}^2$, as $n_{R_0} + 1 > n_m$.

Second, consider $n_{R_0} < n_{R_m} + n_{R_b}$. The following change in the room allocation creates more information: Move n_{R_b} players from R_b and n_{R_b} players from R_0 to R_m. This leaves no one in R_b, $n_{R_m} + 2n_{R_b}$ in R_m and $n_{R_0} - n_{R_b}$ players in R_0. As in the proof of lemma 4, the move maintains truthtelling incentives for players in R_m. The number of pieces of information generated after the move is $(n_{R_m} + 2n_{R_b})^2 + (n_{R_0} - n_{R_b})^2 = n_{R_m}^2 + 4n_{R_b}n_{R_m} + 5n_{R_b}^2 + n_{R_0}^2 - 2n_{R_0}n_{R_b}$ which is higher than the number of pieces of information generated by R_0, R_b and R_m without the move, i.e. $n_{R_b}^2 + n_{R_0}^2 + n_{R_m}^2$, by $n_{R_0} < n_{R_b} + n_{R_m}$. □

Corollary 4. *The welfare optimal room assignment consists of at most two rooms and at most one room in which players of both biases are present.*

Proof of corollary 4: This follows from the combination of corollary 3 and lemmas 2, 4 and 5. □

1.3. Welfare optimal room allocation and equilibrium

Corollary 4 (and the preceding lemmas) leave the following possibilities for welfare optimal room assignment:

- segregation: each bias group has its own exclusive room.
- full integration: one room for all players
- mix: one room exclusively with players of bias x and one room with both bias types in which either
  - only players of type $-x$ send truthful messages
  - all players send truthful messages.

The idea behind the mix situation is the following: If there are more players with bias $-x$ than players with bias x and assigning all players to one room would lead to babbling, then it can be optimal not to separate completely but to assign some players of the minority x to the majority room $-x$. The x players will babble there but they receive a lot of information (truthful messages of all the $-x$ players). If such a situation is optimal, then clearly it must be the case that assigning one more minority player to the mixed room would lead to babbling. This is the first mixed assignment possibility.

The second mixed assignment possibility refers to a situation where the minority bias group would babble if all players were in one room while the majority would send truthful messages. In this case, taking some players of the majority to a separate room can restore truthtelling by both groups in the mixed room and might be optimal.

As should be clear from the discussion above, the mixed scenario is only welfare optimal if the group sizes differ. With equal group sizes either segregation or full integration is optimal. Which of the two is optimal depends on whether the bias difference b is sufficiently small to obtain truthtelling in a fully integrated room. While it is straightforward to prove this directly, we will here prove the more general theorem 2 and come back to the special case of equal group sizes afterwards.

Proof of theorem 2 on page 12.

By corollary 4, we can focus on at most two rooms in the welfare optimal room assignment.

If $b/(p-1/2) \leq (n_0+n_b-1)/n_0$, then full information, i.e. all players in one room and every player sends a truthful message, is feasible and a single room is obviously welfare optimal.

If $(n_0 + n_b - 1)/n_0 < b/(p - 1/2) \leq (n_0 + n_b - 1)/n_b$ (which is only possible if $n_0 > n_b$), then players with bias b would babble in a single room. It is important to notice that segregation cannot be optimal in this situation: segregation leads to $n_b^2 + n_0^2$

pieces of information while in one single room there would still be $n_0 * (n_0 + n_b)$ pieces of information which is greater than $n_b^2 + n_0^2$ by $n_0 > n_b$.[2] Consequently, there are two options for the welfare maximal room assignment. Either all players are in one single room and babbling by players with bias b is tolerated or some players with bias 0 are assigned to a separate room in order to balance the mixed room and restore truthtelling incentives for the bias b players. The maximal number of bias 0 players that can remain in the mixed room without inducing babbling by bias b players is $n_{m0} = \lfloor (n_b - 1)/(b/(p - 1/2) - 1) \rfloor$. Furthermore, truthtelling by the majority group clearly also requires that players with bias 0 are (weakly) in the majority. Hence, consider for now $n_{m0} \geq, n_b$. The number of pieces of information in the scenario with n_{m0} bias 0 players and all bias b players in one room and $n_0 - n_{m0}$ bias 0 players in a separate room is $(n_b + n_{m0})^2 + (n_0 - n_{m0})^2$. Whether this is higher or lower than the number of pieces in a fully integrated room, i.e. $n_0(n_0 + n_b) + n_b$, depends on the parameters. In particular, if $b/(p - 1/2)$ is close to the lower boundary (and n_b is not too small), two rooms will be optimal as n_{m0} will be relatively high. If $b/(p - 1/2)$ is close to the upper boundary, however, n_{m0} will be low and one room with babbling by the minority will be optimal. Finally, if $n_{m0} < n_b$, then babbling by one group occurs in any mixed room and therefore it is optimal to have one integrated room in which the minority players babble.

Finally, we consider $b/(p-1/2) > (n_0+n_b-1)/n_b$. In this case, even the bias 0 players will babble if all players are in one room. This implies that putting some bias zero players in an own separate room will no longer help: All the remaining bias zero players would have even higher incentives to babble and it would be more informative to fully separate the two bias groups. Consequently, only two options remain: Either the just mentioned segregation or enough bias b players are assigned to an own separate room to restore truthtelling incentives for the bias 0 players in the mixed room. The maximum number of bias b players that can remain in the mixed room without destroying truthtelling by bias 0 players is $n_{mb} = \lfloor (n_0 - 1)/(b/(p-1/2) - 1) \rfloor$. This yields $n_0(n_0 + n_{mb}) + n_{mb} + (n_b - n_{mb})^2$ pieces of information which can, depending on the parameters, be higher or lower than the $n_0^2 + n_b^2$ pieces of information created by segregation.

It remains to check when the welfare optimal room assignment constitutes an equilibrium of the game. If full information is feasible, i.e. if $b/(p-1/2) \leq (n_0+n_b-1)/n_0$, then it is clearly an equilibrium. For $(n_0+n_b-1)/n_0 < b/(p-1/2) \leq (n_0+n_b-1)/n_b$, the welfare optimal room assignment is definitely an equilibrium if it is optimal to keep all players in the same room (and have the smaller group babbling): The point is that this can only be optimal if isolating one player of the larger group would not lead to truthtelling in a room with all other players. (If this was the case, then isolating one player of the larger group would be optimal.) But this implies that any deviation in room choice from the situation

[2]Similarly, it is not optimal to put only some bias b players into their own room: As $n_b \leq n_0$, they would have less information there than they would have if they were babbling in one single room.

with one big mixed room will lead to less information for the deviating player and less (or the same) information for the other players. Following proposition 1, such a deviation will therefore reduce the deviating player's expected payoff. Hence, the welfare maximizing room assignment is an equilibrium if $(n_0+n_b-1)/n_0 < b/(p-1/2) \leq (n_0+n_b-1)/n_b$ and $(n_b+n_{m0})^2+(n_0-n_{m0})^2 \leq n_0(n_0+n_b)$. For the case $(n_b+n_{m0})^2+(n_0-n_{m0})^2 > n_0(n_0+n_b)$ where two rooms are optimal – one mixed room with truthtelling by all players and one with only players of the larger group – the relevant question is whether the $n_0 - n_{m0}$ players in the room only for bias 0 players would want to deviate to the mixed room. Note that this deviation will destroy truthtelling by the bias b players. However, a unilaterally deviating player will obtain information from n_{m0} other players which is more information than the $n_0 - n_{m0} - 1$ truthful messages he obtains when not deviating.[3] It is therefore unsurprising that the welfare optimal room assignment will only be an equilibrium if α is sufficiently high. Following 1, the deviation will not be profitable if and only if it decreases $\zeta_i + \alpha \sum_{j \neq i} \zeta_j$. This is the case if and only if

$$\begin{aligned} n_{m0} + 1 + \alpha \left((n_0 - n_{m0} - 1)^2 + (n_{m0} + 1)(n_b + n_{m0}) \right) \\ \leq n_0 - n_{m0} + \alpha \left((n_0 - n_{m0} - 1)(n_0 - n_{m0}) + (n_{m0} + n_b)^2 \right). \end{aligned}$$

$$\Leftrightarrow \alpha \geq \frac{2n_{m0} - n_0 + 1}{n_0 - 2n_{m0} - 1 + n_b^2 + n_b(n_{m0} - 1)}. \tag{2}$$

Hence, if $(n_0+n_b-1)/n_0 < b/(p-1/2) \leq (n_0+n_b-1)/n_b$ and $(n_b+n_{m0})^2+(n_0-n_{m0})^2 > n_0(n_0+n_b)$ as well as $n_{m0} \geq n_b$, the welfare optimal room assignment is an equilibrium if and only if (2) holds.

For $b/(p-1/2) > (n_0+n_b-1)/n_b$, the welfare optimal room assignment can be either a mixed room combined with a room exclusively for the smaller group or segregation. In the first case it is straightforward to see that this is an equilibrium. Recall that moving one more player from the separate room to the mixed room would lead to babbling by all players in the mixed room. Clearly, such a deviation is not profitable. In the second case, we already showed before that total separation is an equilibrium if either $b/(p-1/2) \geq n_0$ or $\alpha \geq (1 + n_0 - n_b)/(n_b - 1)$. □

Note that for $b/(p - 1/2) > n_0$ we get $n_{bm} = 0$ and in this case segregation is always welfare optimal. Furthermore, an increase in $b/(p - 1/2)$ will decrease both n_{m0} and n_{mb}. This implies that we can simply step through the different cases of theorem 2 as $b/(p - 1/2)$ increases. However, depending on parameter values, in particular n_0 and n_b,

[3]To see this, first note that $n_{m0} \geq n_b$ as no truthtelling mixed room exists if bias b players were not truthtelling in a perfectly balanced, i.e. same number of players from each bias group, room. Then note that $(n_b+n_{m0})^2+(n_0-n_{m0})^2 > n_0(n_0+n_b)$ is more demanding for higher n_0 (keeping other parameters fixed). It is straightforward to calculate that the inequality cannot be satisfied (given $n_b \leq n_{m0}$) for $n_0 \geq 5n_{m0}/3$. Hence, $n_{m0} \geq 4n_0/5$ whenever this room assignment is welfare maximizing which implies $n_{m0} > n_0 - n_{m0} - 1$.

some of the cases may be skipped. However, the first and last case will never be skipped as they always occur for sufficiently low, respectively high, values of $b/(p-1/2)$.

If the welfare optimal room allocation is not an equilibrium, then the welfare optimal equilibrium features too little segregation: In case (2b), the equilibrium has all players in a single room while the welfare optimal allocation has two rooms. In case (3b), the welfare optimal room allocation is full segregation which is not an equilibrium.

To illustrate, consider the special case of equal group sizes, i.e. $n_0 = n_b = n/2$. This rules out case (2) of theorem 2. Furthermore, the condition defining case (3a) is not met as it can be rewritten as $n_{mb}+1-n/2 \geq 0$ but given the condition $b/(p-1/2) > (n-1)/n_b$ of case (3) $n_{mb} < n/2-1$ and therefore case (3a) is not possible if $n_b = n_0 = n/2$. Hence, only case (1) and (3b) remain with equal group sizes and therefore the welfare optimal room allocation is either full segregation or full integration in this case. Full integration is welfare optimal if $b/(p-1/2) \leq (n-1)/(n/2)$ and is also an equilibrium in this case. If $b/(p-1/2) > (n-1)/(n/2)$, segregation is welfare optimal but only an equilibrium if $\alpha \geq 1/(n/2-1)$. Otherwise, there is too little segregation in equilibrium.

2. Two specific types of bias distributions

To find out how polarized biases need to be so that segregation is optimal and an equilibrium, we can consider two stylized cases. First, we will consider biases that are evenly distributed on an interval of the real line. We can think of this case as having "zero polarization", whereas clustering of biases around certain values exhibits positive polarization. Second, we consider biases that are tightly clustered around a central value – we could think of this as "negative polarization".

2.1. Uniformly distributed biases

We will show the following result:

Result 3. *Let $b_i = (i-1)*k/(n-1)$ for $i = 1, \ldots, n$. Then the welfare optimal room allocation assigns either all players to one single room or all but one extreme player to the same room. Assigning all players to the same room is also an equilibrium.*

Intuitively, we can start by considering the fully integrated room, in which some people whose biases are close to the overall average tell the truth, and the rest babble and learn from the truth-tellers. Since biases are evenly distributed by assumption, there is little welfare to be gained by moving the bias average around by allocating people to another room. (This can only work because of integer effects – i.e. because changes in the average bias have discrete effects on who tells the truth – which is precisely what gives us the exceptions in the second half of the proposition.) Any room that includes only part of the players will have a shorter truth-telling interval, which (again, absent integer effects) means fewer truth-tellers. But if we cannot increase the number of truth-tellers by segregating into smaller rooms, then the fully integrated room must be welfare-optimal and also an equilibrium: Every player receives the highest possible number of truthful messages while the number of players having their own signal in addition to this number of messages is also maximal. The remainder of this section shows this result formally.

Proposition 5. *Let $b_i = (i-1) * k/(n-1)$ for $i = 1, \dots, n$. Then one single room with all players is both welfare optimal and an equilibrium if either*

$$\underline{b} - \left[k/2 - (p - 1/2)\frac{n-1}{n}\right] \leq \frac{k}{2(n-1)} \tag{3}$$

or

$$k(1 - 1/n)/2 - (p - 1/2)\frac{n-2}{n-1} > \underline{b} - \frac{k}{n-1}. \tag{4}$$

If neither of these two conditions holds, isolating player n in one room and all other players in one room is welfare optimal. This is only an equilibrium if

$$\alpha \geq \frac{\lfloor \frac{n-1}{n}(2p-1)/(k/(n-1)) \rfloor - 1}{n-2}. \tag{5}$$

Proof of proposition 5: Theorem 1 states that in the most informative equilibrium of the messaging subgame players in room R will tell the truth if and only if $b_i \in \left[\overline{b} - \frac{n_R-1}{n_R}(p - \frac{1}{2}), \overline{b} + \frac{n_R-1}{n_R}(p - \frac{1}{2})\right]$. If this interval covers $[0, k]$, then one room leads to truthtelling by all players and one single room is clearly optimal. In the remainder of this proof, we therefore assume that this is not the case. The length of the interval $\left[\overline{b} - \frac{n_R-1}{n_R}(p - \frac{1}{2}), \overline{b} + \frac{n_R-1}{n_R}(p - \frac{1}{2})\right]$ is $\frac{n_R-1}{n_R}(2p-1)$. The number of players telling the truth in any room is consequently bounded from above by $\lfloor \frac{n_R-1}{n_R}(2p-1)/(k/(n-1)) \rfloor + 1$ as the players' biases are equally spaced with distance $k/(n-1)$ between two consecutive players' biases. This bound may not be attained by any feasible room due to the discrete nature of the problem. More specifically, if we take the fully integrated room, then the number of truthtelling players will be either $\lfloor \frac{n-1}{n}(2p-1)/(k/(n-1)) \rfloor + 1$ or $\lfloor \frac{n-1}{n}(2p-1)/(k/(n-1)) \rfloor$.

Let t^* be the maximal number of truthtelling players in any possible room. From the above, it is clear that $t^* \in \left\{ \lfloor \frac{n-1}{n}(2p-1)/(k/(n-1)) \rfloor + 1, \lfloor \frac{n-1}{n}(2p-1)/(k/(n-1)) \rfloor \right\}$. Suppose t^* is the number of truthtelling players if all players are in the same room. Then the number of pieces of information generated in this room is t^*n+n-t^*. We will show that in this case no other room configuration generates more pieces of information: The total number of pieces of information in r rooms is: $\sum_R t_R n_R + n_R - t_R = \sum_R t_R(n_R - 1) + n_R \leq \sum_R t^*(n_R - 1) + n_R = t^*(n-r) + n \leq t^*n + n - t^*$. By proposition 1, one big room with all players is then welfare optimal if this leads to t^* truthtelling players.

Next consider the situation where one integrated room with all players leads not to t^* but only to $t^* - 1$ truthtelling players. Suppose that there is some room R^* with

$n-1$ players in which t^* players are truthtelling. We show that in this case the room configuration $(R^*, \{1, \ldots, n\} \setminus R^*)$ is welfare optimal. This will lead to $t^*(n-1)+n-1-t^*+1 = t^*(n-2)+n$ pieces of information. The big integrated room leads to only $(t^*-1)n+n-t^*+1 = t^*(n-1) < t^*(n-1)-t^*+n$ pieces of information and is therefore welfare inferior. Any other room configuration with r rooms leads to $\sum_R t_R n_R + n_R - t_R = \sum_R t_R(n_R-1)+n_R \leq \sum_R t^*(n_R-1)+n_R = t^*(n-r)+n \leq t^*(n-2)+n$ pieces of information which is also (weakly) less than $(R^*, \{1, \ldots, n\} \setminus R^*)$. Hence, in this case $(R^*, \{1, \ldots, n\} \setminus R^*)$ is welfare optimal.

Finally, we show that the conditions in the proposition lead to either of the two just described cases. Note that in the fully integrated room $\bar{b} = k/2$. Hence, condition (3) states that the distance from the lowest player's bias who tells the truth to the lower boundary of the truthtelling interval is less than 1/2 the distance between two consecutive players' biases. By symmetry of the truthtelling interval around $\bar{b}$ and the equal spacing of biases, this is also true for the distance of the highest bias player telling the truth and the upper boundary of the truthtelling interval. First, let (3) hold strictly. Then it is clear that shifting the truthtelling interrval (by changing $\bar{b}$) cannot lead to more players being truthtelling. Furthermore, the length of the truthtelling interval is strictly decreasing in the number of players in the room. Hence, in no other room can there be more truthtelling players than in the fully integrated room. This holds also if (3) holds with equality as the length of the truthtelling inequality is strictly decreasing in the number of players in the room. Consequently, t^* is achieved by the fully integrated room and the argument two paragraphs above shows that then the fully integrated room is welfare optimal.

Now consider the case where (3) does not hold. Start from the fully integrated room. If (3) does not hold, shifting the truthtelling interval by $k/(2(n-1))$ down (by – for now magically – reducing $\bar{b}$ by this amount), will imply that this interval contains 1 more player than in the fully integrated room. Furthermore, the distance of this lowest truthtelling player after the shift to the lower boundary of the truthtelling interval will be less than $k/(2(n-1))$ by the assumption that (3) did not hold. Now note that removing player n from the fully integrated room will reduce $\bar{b}$ by exactly $k/(2(n-1))$ (from $k/2$ to $(k - k/(n-1))/2$). But note that removing this player also implies that $n_R = n-1$ and therefore the length of the truthtelling interval is reduced. Condition (4) states that due to the shrinking of the interval when moving from n to $n-1$ players the one player whose truthtelling was gained by shifting the interval down is lost again. Furthermore, the "shrinking" occurs at the upper as well as the lower boundary to the same extent. This implies that also at the upper boundary one truthtelling player is lost due to the shrinking (while the shifting did not lose anyone as (3) was violated by assumption). Consequently, the room without player n will have one less truthtelling player than the fully integrated room if (3) is violated and (4) holds. In this case, no room with $n-1$ (or less) players can have more truthtelling players than the fully integrated room and therefore t^* is attained in the fully integrated room. Consequently, the fully integrated room is by the results above welfare optimal.

If neither (3) nor (4) holds, then the "shifting" argument above implies that the room allocation $(\{1,\dots,n-1\},\{n\})$ leads to one more truthtelling player in $R^* = \{1,\dots,n-1\}$ than in the fully integrated room. Consequently, t^* is attained in R^* and $(\{1,\dots,n-1\},\{n\})$ is welfare optimal by the results above.[4]

In terms of equilibrium, it is immediate that no player wants to deviate from the fully integrated room by isolating himself as self-isolation leads to less information for himself and no more information for other players. The same argument applies for players in room R^* in case (3) and (4) are violated. However, the isolated player might have an incentive to join R^*: This would reduce the amount of information as only t^*-1 instead of t^* players would be truthtelling in the resulting fully integrated room reducing the number of pieces information of all other players in this room from $t^*(n-1)+n-1-t^*$ to $(t^*-1)(n-1)+n-t^*$. However, the deviating player would gain more information for himself, i.e. the number of pieces of information he observes is t^* instead of 1. From 6, it follows that the deviation is profitable if and only if $\alpha < (t^*-1)/(n-2)$. Note that $t^* = \lfloor \frac{n-1}{n}(2p-1)/(k/(n-1)) \rfloor$ in the here analyzed case where one integrated room is not optimal. This gives the condition in the proposition. □

2.2. Symmetrically single peaked bias distribution

We now move to symmetrically, single peaked distribution of biases: Assume that biases are on an equally spaced grid $0, d, 2d, \dots, Kd$ for some $d > 0$ and $K \in \mathbb{N}$. The number of players with bias $b_i = kd$ is increasing up to $Kd/2$ and decreasing thereafter. Furthermore, we assume that the number of players with bias kd equals the number of players with bias $(K-k)d$ for $k = 0, 1, \dots, \lfloor K/2 \rfloor$.

To state our proposition we need the following notation: Let $\underline{k}$ be the lowest k such that $kd \geq Kd/2 - (p-1/2)(n-1)/n$ and let $\bar{k}$ be the highest k such that $kd \leq Kd/2 + (p-1/2)(n-1)/n$. Note that due to the discreteness of the grid and following theorem 1, the truthtelling interval in a fully integrated room will cover all players with $b_i \in [\underline{k}d, \bar{k}d]$.

Proposition 6. *With a symmetric, single peaked distribution of biases, one room containing all players is welfare optimal and also an equilibrium if*

$$\bar{k}d - \underline{k}d + d > (2p-1)\frac{n-2}{n-1}. \tag{6}$$

Proof of proposition 6: Theorem 1 states that in the most informative equilibrium of the messaging subgame players in room R will tell the truth if and only if $b_i \in \left[\overline{b} - \frac{n_R-1}{n_R}(p-\frac{1}{2}),\, \overline{b} + \frac{n_R-1}{n_R}(p-\frac{1}{2})\right]$. If this interval covers $[0, Kd]$, then one room

[4]It should be noted that similar arguments as above, with an upward instead of a downward shift, lead to the optimality of $(\{1\},\{2,\dots,n\})$ which will also attain t^* if (3) and (4) are violated.

leads to truthtelling by all players and one single room is clearly optimal. In the remainder of this proof, we therefore assume that this is not the case. Note that – holding $\bar{b}$ fixed – the length of the interval is increasing in n_R. If we turn to the case of one fully integrated room, then the truthtelling interval is $\left[Kd/2 - \frac{n-1}{n}(p - \frac{1}{2}),\, Kd/2 + \frac{n-1}{n}(p - \frac{1}{2})\right]$ as $\bar{b} = Kd/2$. We will first show the result under a condition slightly stronger than (6), namely under the condition

$$\bar{k}d - \underline{k}d + d > (2p-1)\frac{n-1}{n}. \tag{7}$$

Condition (7) states that the length of the truthtelling interval is less than $\bar{k} - \underline{k} + d$. (Note that the length of the truthtelling interval is weakly larger than $\bar{k}d - \underline{k}d$ due to the discrete grid on which biases are distributed.) This implies that the truthtelling interval would not cover more grid points if it was moved up or down while keeping its length constant. As the truthtelling interval is shorter for any other room (because of $n_R < n$) and the distribution of biases is single-peaked, this implies that there is no room in which more players are truthtelling than in the fully integrated room.

The same conclusion follows if (6) holds instead of (7): (6) states that the length of the truthtelling interval in any room different from the fully integrated room (which therefore contains at most $n-1$ players) is less than $\bar{k}d - \underline{k}d + d$ which again implies that the truthtelling interval of such a room cannot cover more grid points than the fully integrated room and by single peakedness it can therefore also not contain more truthtelling players.

Let t^* be the maximal number of truthtelling players in any possible room. From the above, t^* is attained by the fully integrated room if (6) holds. In this case, the number of pieces of information generated in the fully integrated room is $t^*n + n - t^*$. We will show that no other room configuration generates more pieces of information: The total number of pieces of information in r rooms is: $\sum_R t_R n_R + n_R - t_R = \sum_R t_R(n_R - 1) + n_R \leq \sum_R t^*(n_R - 1) + n_R = t^*(n-r) + n \leq t^*n + n - t^*$. By proposition 1, one big room with all players is therefore welfare optimal if (6) holds.

In case a single fully integrated room is welfare maximal it is also an equilibrium: Unilateral self-isolation would lead to less information for the deviating player and also – by welfare optimality of the fully integrated room – to less information over all. The deviation is therefore unprofitable. □

3. Example: Too much segregation in equilibrium

In the case of two bias groups, we have shown that the welfare-optimal equilibrium room allocation is either the overall welfare-optimum, or has too little segregation compared to it. If there are three or more bias groups, this is not generally true anymore – now it is possible that the welfare-optimal equilibrium involves *too much* segregation compared to the welfare-optimum. This can occur when a player wants to deviate from the welfare-optimum to another room where he can learn more and thereby destroys the truth-telling incentives of people in the room that he is leaving. The following paragraphs provide an example for such a situation.

Consider a bias configuration in which 13 people have bias $b_1 = -1000$, 10 people have bias $b_2 = 0$ and 2 people have bias $b_3 = 500$. We can easily see that there exists no possible room with members of exactly two bias groups in which anyone tells the truth. Even in rooms that involve all three bias groups, no one with biases b_1 and b_3 will ever tell the truth. The only way to get anyone with bias 0 to tell the truth in a mixed room is to create a room with one person with bias b_1, two people with bias b_3 and an arbitrary number of people with bias 0. This leads us to the welfare-optimal room allocation: Room 1 consists of 12 people with bias b_1 and generates 144 pieces of information, and room 2 contains everybody else and generates 133 pieces of information, for a total $\sum \zeta_i = 277$. For low α, this allocation is not an equilibrium: The person with bias b_1 in room 2 can change to room 1 and have 13 pieces of information instead of 11.

Now consider the room allocation where bias groups are fully segregated: This generates $169+100+4 = 273$ pieces of information and is also an equilibrium: No one can learn anything by switching to another room. Hence, this is the welfare-optimal equilibrium, while the first allocation we described is welfare-optimal – which means that there is too much segregation in the welfare-optimal equilibrium. (Note that this example is generic in the sense that we could find an open ball of bias configurations around this particular bias configuration in which our conclusions remain valid.)

4. Non-credible threats could increase information

In the paper we select the most informative cheap talk equilibrium in the messaging stage of the game. If we allow players to commit to play less informative equilibria in certain rooms, then this threat can be used to affect room choice in a way that increases the amount of information transmission. We do not consider such threats in the paper as they seem non-credible to us. As soon as players find themselves in a given room allocation, every player (even those players in other rooms!) benefit strictly from playing the most informative messaging equilibrium.

This section, however, shows how one can easily generate examples in which such threats help to increase the amount of transmitted information. We stick to the case of two bias groups, i.e. n_0 players have bias 0 and n_b players have bias b. To simplify further, let $n_0 = n_b + 1$. We will focus on cases in which the welfare optimal room allocation is full segregation while the equilibrium room allocation in case the most informative equilibrium is played in every room is such that room R_1 hosts all players with bias 0 (truthtelling) as well as 2 players of bias b (babbling) while all other players of bias b are in room R_2.[5] Using the conditions from the paper, this scenario will emerge if the following conditions are met:

1. truthtelling by bias 0 in R_1: $\frac{2}{n_0+2}b - \frac{n_0+1}{n_0+2}\left(p - \frac{1}{2}\right) \leq 0$,
2. babbling by bias 0 if a third bias b player joined R_1: $\frac{3}{n_0+3}b - \frac{n_0+2}{n_0+3}\left(p - \frac{1}{2}\right) > 0$,
3. bias b players prefer R_1 (i.e. full segregation is not an equilibrium): $\frac{b}{p-1/2} < n_0$ and $\alpha < (1 + n_0 - n_b)/(n_b - 1) = 2/(n_b - 1)$.

The equilibrium with 2 deviating (and babbling) bias b players generates $(n_b - 2)^2 + 2 * (n_0 + 1) + n_0^2 = (n_b - 2)^2 + 2 * (n_b + 2) + (n_b + 1)^2 = 2n_b^2 + 9$ pieces of information.

We will now focus on the following messaging equilibrium: There is one player with bias 0 who is babbling whenever all players with bias 0 are in the same room. All other players (and all players in all room allocations in which not all players with bias 0 are in the same room) will play the most informative messaging equilibrium. With this play in the messaging phase, full segregation, i.e. all players of bias 0 in room R_1 and all players of bias b in R_2, is an equilibrium if

$$\alpha \geq (1 + n_0 - 1 - n_b)/(n_b - 1) = 1/(n_b - 1).$$

Note that such full information (with the one designated bias 0 player babbling) creates $n_b^2 + (n_0 - 1)^2 + n_0 = 2n_b^2 + n_b + 1$ pieces of information which exceeds $2n_b^2 + 9$ if $n_b > 8$. That is for parameters satisfying $n_b > 8$, $n_0 = n_b + 1$, $\alpha \in [1/(n_b - 1), 2/(n_b - 1))$ as well as the three conditions above, the threat of babbling in a certain room configuration helps to increase information transmission through its effect on room choice. Example parameters satisfying all conditions are $n_b = 9$, $n_0 = 10$, $p = 5/6$, $b = 3/2$, $\alpha = 1/5$.

5. Uncertainty

So far, we have assumed that all biases b_i are common knowledge. This may not always be the case, especially in environments where communication is somewhat anonymous, such as on the internet. In such cases, it seems reasonable to assume that both the state of the world and the types of all players are subject to uncertainty.

5.1. Main results and intuition

Let all biases b_i be randomly and independently distributed on $\mathbb{R}$ according to distribution F_i. Each player observes his own bias b_i, but only knows the distributions of the biases of other players. Let $b_i^e = \int_{-\infty}^{\infty} b_i \, dF_i$ be the expected value of b_i. This can be thought of

[5] It is straightforward to show that we need at least 2 deviating players in this equilibrium to generate more information through the threat.

as a generalization of the paper's main model, in which all biases were always identical to their expected value. When we talk about "introducing" or "adding" uncertainty in this context, we think of starting with the model in which all biases are known with certainty, and replacing each bias with a bias distribution that has the same expected value. Throughout this section, we will be comparing across distributions that have the same expected value. The following paragraphs intuitively analyze the model with uncertainty; the corresponding formal statements and analysis are in section 5.2 below.

To find the messaging equilibria within a room, we need to consider i's problem of choosing a message m_i after observing b_i and σ_i, but only knowing F_j for all $j \in R_i$. We can show that this problem is very similar to knowing all biases with certainty. In particular, recall that i's willingness to tell the truth depended only on the distance between b_i and the average of all other b_j's in the model with certainty. This insight applies analogously to a model in which all biases are unknown: Now i cares only about the difference between b_i and the average of all b_j^e, i.e. the expected values of other people's bias.

A difference in describing equilibria with uncertainty arises since i may want to tell the truth for some values of b_i and not for others, and the other players are unsure about b_i when interpreting m_i. Their belief about how likely i is to tell the truth hence depends on how b_i is distributed. For each possible probability with which i tells the truth, there exists an interval around $\frac{\sum_{j\in R_i, j\neq i} b_j^e}{n_{R_i}-1}$ such that i wants to tell the truth if the realized b_i lies within this interval. Since the distribution of b_i is common knowledge, that gives us the following equilibrium condition: The beliefs of all other players about i's probability of truth-telling need to give rise to a truth-telling interval for i around the average of all b_j such that i wants to tell the truth with exactly the probability with which the other players believe that he tells the truth.

This translates into a slightly generalized version of theorem 1 which, for any distribution of b_i, gives us the highest probability with which i can tell the truth in any equilibrium. Intuitively, the more concentrated F_i is around $\frac{\sum_{j\in R_i, j\neq i} b_j^e}{n_{R_i}-1}$, the higher the probability with which i can tell the truth in equilibrium. Interestingly, only the probability mass of F_i that is sufficiently close to $\frac{\sum_{j\in R_i, j\neq i} b_j^e}{n_{R_i}-1}$ matters; whether or not b_i^e itself is close to the average or not is not directly relevant for whether i is able to tell the truth in equilibrium.

In particular, this means that we can choose any set of expected biases, regardless of how close they are to each other, and construct bias distributions such that none of the players ever wants to tell the truth to anyone in any room allocation. This means that for any bias configuration, uncertainty has the potential to completely destroy all chances of creating a room in which information is exchanged.

Proposition 7. *Take a set of n players with biases $\{b_1, b_2, \ldots, b_n\}$ such that there exists a room allocation in which some (or all) players tell the truth. Then there exists a set of probability distributions $\{F_1, F_2, \ldots, F_n\}$ of biases with expected values $\{b_1, b_2, \ldots, b_n\}$*

such that in any room allocation of the n players, no player will tell the truth in any equilibrium. (Proof on page 23.)

This is, of course, a very stark result. Uncertainty need not always destroy communication. It can, in fact, make communication possible where it was previously impossible, by moving probability mass of b_i's distribution closer to the average of other biases. This effect, however, is more limited and can never lead to full truth-telling if there is no full truth-telling in a model with certain biases and identical expected values.

Proposition 8. *If b_i is such that there exists no equilibrium in room R_i where i tells the truth, there exists a distribution F_i with expected value $b_i^e = b_i$ such that there exists an equilibrium in R_i where i tells the truth with positive probability. However, there exists no F_i such that i tells the truth with probability 1 in any equilibrium. (Proof on page 23.)*

While uncertainty can make some truth-telling possible where it was not possible with certainty, large amounts of uncertainty will always destroy any truth-telling and make all messages arbitrarily uninformative unless they preserve sufficient probability mass in the neighborhood of $\frac{\sum_{j\in R_i, j\neq i} b_j^e}{n_{R_i}-1}$. Because of the large space of possible distributions and possible orderings on uncertainty, we show this result in two ways. First, we consider any continuous bias distribution and show that by "stretching" it, any equilibrium will become arbitrarily uninformative. Then we consider discrete bias distributions with bounded support, and show that any way of increasing the variance of such a distribution will likewise eventually erode all informative equilibria. In the following propositions, μ_{ji}^l is j's belief about θ_i, given that i has sent the signal m^l; the other expressions involving μ are defined analogously.

Proposition 9. *Let F be a continuous distribution function that is continuous at its expected value b_i^e and symmetric around b_i^e. Let $F^\kappa(x) = F(b_i^e + \kappa(x - b_i^e))$, i.e. $b_i = b_i^e$ almost surely for $\lim_{\kappa\to\infty} F^\kappa$. For any F and $\varepsilon > 0$, there exists a $\bar{\kappa} > 0$ such that $\mu_{ji}^h - \mu_{ji}^l < \varepsilon$ if $F_i = F^\kappa$ and $\kappa \leq \bar{\kappa}$. (Proof on page 24.)*

Proposition 10. *Fix the expected bias b_i^e of all players in a given room and a bounded support for all bias distributions F_i. Assume that there is at least one element in the support that is smaller than $\frac{\sum_{j\in R_i, j\neq i} b_j^e}{n_{R_i}-1} - (2p-1)$ and at least one element that is larger than $\frac{\sum_{j\in R_i, j\neq i} b_j^e}{n_{R_i}-1} + (2p-1)$. Then for each $\varepsilon > 0$ there exists some $\overline{\sigma_{F_i}}$ such that for all such F_i with $\mathrm{Var}\,(b_i) \geq \overline{\sigma_{F_i}}^2$, $\mu_{ji}^h - \mu_{ji}^l \leq \varepsilon$. (Proof on page 24.)*

Figure 2 illustrates propositions 8 to 10. The bias configuration is identical to the one in figure 2 on page 10, except that there is some mean-preserving uncertainty about the biases of players 2 and 5, whose biases are now distributed according to a bell-shaped distribution function. Under certainty, player 2 was telling the truth, but is now only telling the truth if his realized b_2 falls within the interval (proposition 9). Player 5 was

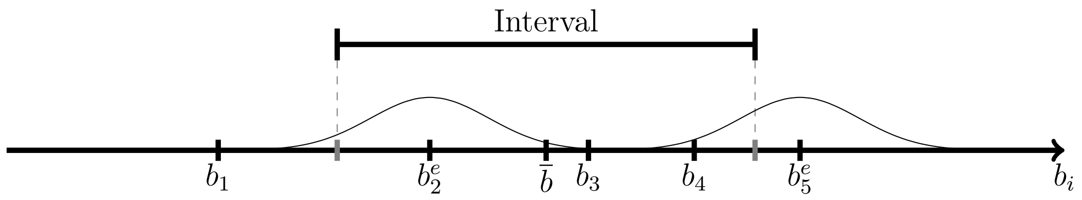

Figure 2: An illustration of propositions 8 to 10. (The biases are identical to the one in figure 2 of the paper except that b_2 and b_5 are now uncertain.)

babbling, but will now sometimes send an informative message if his realized bias is close enough to $\bar{b}$ (proposition 8).[6]

These results already contain statements about room choice with uncertainty: If truth-telling is greatly reduced or becomes impossible, there is not much to be gained from being in one room. Of course, truth-telling between people with identical bias distributions is not necessarily easier – note that proposition 7 contained no assumption that people differ in how their biases are distributed. So are the effects of uncertainty simply to make communication hard in general? Not necessarily. Consider a model where full integration is welfare-optimal and an equilibrium if biases are known. We can show that for any such model, uncertainty can cause segregation between groups to become Pareto-superior to integration, and such segregation may also be an equilibrium of the room choice game.

Proposition 11. *Let the number of players be weakly larger than 4 and let $b_i^e \in \{0, b\}$, with $b \in \left(0, \frac{n-1}{n}(2p-1)\right]$. Let the two bias groups be of equal size, i.e. $n_0 = n_b = n/2$. Then in the room-choice game:*

- *If $b_i = b_i^e$ with certainty, the fully integrated room is welfare-optimal and an equilibrium.*
- *If biases are uncertain, we can find distributions F_i that keep all b_i^e constant such that full segregation between the two bias groups is welfare-optimal. For $\alpha \geq \frac{2}{n-2}$, this is also an equilibrium.*

(Proof on page 24.)

To illustrate this result, let us consider the example of finding an optimal tax rate (i.e. the optimal amount of redistribution) when information about the state of the economy is dispersed. Assume that the world consists of liberals and conservatives. Liberals generally prefer higher taxes than conservatives, but everybody is aware that the optimal tax level depends on how bad taxes are for economic growth. If the exact political preference of each person is known, an informative exchange is possible even across party lines as long

[6]This graphic is meant as an illustration and ignores the fact that, while the interval's length remains constant, its precise location may shift depending on the exact beliefs of the receiving players in equilibrium.

as preferences are not too different. But now assume that instead, each member of each political group is either a moderate or an extremist. It is only observable whether anyone is liberal or conservative, not whether they are extremists or moderates. Both have equal probability, so that in expectation each person is still an "average" liberal or conservative.

Consider the problem of a liberal who is unsure whether he is listening to a moderate conservative or a conservative extremist. He knows that a conservative extremist would always tell a liberal that taxes are bad for the economy, regardless of what his information is. Any statement about the damages of taxes has hence become less informative, while being more likely to be made, than if the liberal was talking to an average conservative. The same is true for a conservative listening to a liberal. Yet while discussion across party lines has become less informative, this is not true for discussion within parties: The possible biases within groups are still close enough so that both moderates and extremists want to truthfully reveal their knowledge to other members of their party. It is hence better for liberals to only talk to other liberals and for conservatives to only talk to conservatives, than for any cross-party discussion to take place – not because of strong inherent differences in preferences, but because of uncertainty about who one's interlocutor is.

5.2. Detailed analysis and proofs for the model with uncertainty

5.2.1. Preliminary analysis

Similarly to the derivation of expression (15), we can write

$$
\begin{aligned}
U_i(m^h) &= \mathbb{E}\left[const - \alpha \sum_{j \in R_i, j \neq i} \left(b_j - b_i + \mu_{ji}^h + \sum_{k \neq i} \mu_{jk} - \theta_i - \sum_{k \neq i} \theta_k \right)^2 \middle| \sigma_i \right] \\
U_i(m^l) &= \mathbb{E}\left[const - \alpha \sum_{j \in R_i, j \neq i} \left(b_j - b_i + \mu_{ji}^l + \sum_{k \neq i} \mu_{jk} - \theta_i - \sum_{k \neq i} \theta_k \right)^2 \middle| \sigma_i \right].
\end{aligned}
$$

Note that we are interested in the difference of the two expressions. Hence, while all b_js are now unknown, this uncertainty only matters where b_j is multiplied by μ_{ji}^h and μ_{ji}^l, respectively. We can hence write

$$
\begin{aligned}
\Delta U_i(\sigma_i) &= (U_i(m^h) - U_i(m^l))/\alpha \\
&= 2(\mu_{ji}^h - \mu_{ji}^l)(n_{R_i} - 1) \left[-\frac{\mu_{ji}^h + \mu_{ji}^l}{2} - \frac{\sum_{j \in R_i, j \neq i} b_j^e}{n_{R_i} - 1} + b_i + \mathbb{E}\left[\theta_i \middle| \sigma_i \right] \right], \quad (8)
\end{aligned}
$$

which is identical to (15) except that we have substituted b_j^e for b_j. i's problem remains virtually unchanged, except that he now considers the expected value of biases of other people within the room.

Now consider i's messaging strategy. In the following, let

$$\begin{aligned}
\lambda^h &= \Pr(m_i = m^h | \sigma_i = \sigma^h) \text{ and} \\
\lambda^l &= \Pr(m_i = m^l | \sigma_i = \sigma^l)
\end{aligned}$$

i.e. λ^h and λ^l are the marginal probabilities with which i truthfully reveals his signal, averaging over all possible bias types. For example, if b_i has two possible values with equal probability and i only reveals σ^h truthfully for one of them, then $\lambda^h = \frac{1}{2}$. The resulting beliefs of player j are

$$\begin{aligned}
\mu_{ji}^h &= \frac{p\lambda^h + (1-p)(1-\lambda^l)}{1+\lambda^h - \lambda^l} \\
\mu_{ji}^l &= \frac{p(1-\lambda^h) + (1-p)\lambda^l}{1-\lambda^h+\lambda^l}.
\end{aligned}$$

We can also write the following two terms, which both appear in equation (8):

$$\begin{aligned}
\mu_{ji}^h - \mu_{ji}^l &= \frac{2p\lambda^h + 2p\lambda^l - 2p - \lambda^h - \lambda^l + 1}{(\lambda^h - \lambda^l + 1)(\lambda^l - \lambda^h + 1)} \\
&= (2p-1)\frac{(\lambda^h + \lambda^l - 1)}{(\lambda^h - \lambda^l + 1)(\lambda^l - \lambda^h + 1)} \qquad (9) \\
\mu_{ji}^h + \mu_{ji}^l &= \frac{2p\lambda^h - 2p\left(\lambda^h\right)^2 - 2p\lambda^l + 2p\left(\lambda^l\right)^2 - 2\left(\lambda^l\right)^2 - \lambda^h + \lambda^l + 2\lambda^h\lambda^l + 1}{(\lambda^h - \lambda^l + 1)(\lambda^l - \lambda^h + 1)} \\
&= \frac{4p\left(\lambda^l\right)^2 - 2\left(\lambda^l\right)^2 - 4p\lambda^h\lambda^l + 2\lambda^h\lambda^l + 2p\lambda^h - \lambda^h - 2p\lambda^l + \lambda^l - 2p + 1}{(\lambda^h - \lambda^l + 1)(\lambda^l - \lambda^h + 1)} + 2p \\
&= (2p-1)\frac{2\left(\lambda^l\right)^2 - 2\lambda^h\lambda^l + \lambda^h - \lambda^l - 1}{(\lambda^h - \lambda^l + 1)(\lambda^l - \lambda^h + 1)} + 2p \\
&= (2p-1)\left(\frac{\left(\lambda^l\right)^2 - \lambda^h\lambda^l - \lambda^l}{(\lambda^h - \lambda^l + 1)(\lambda^l - \lambda^h + 1)} + \frac{\left(\lambda^l\right)^2 - \lambda^h\lambda^l + \lambda^h - 1}{(\lambda^h - \lambda^l + 1)(\lambda^l - \lambda^h + 1)}\right) + 2p \\
&= (2p-1)\left(\frac{\lambda^l}{\lambda^h - \lambda^l - 1} + \frac{\lambda^l - 1}{\lambda^h - \lambda^l + 1}\right) + 2p. \qquad (10)
\end{aligned}$$

From (9), we can see that the condition $\mu_{ji}^h \geq \mu_{ji}^l$ translates to $\lambda^h + \lambda^l \geq 1$. We can distinguish two cases:

- $\lambda^h + \lambda^l = 1$. Then $\mu_{ji}^h - \mu_{ji}^l = 0$ and i's messages are completely uninformative.
- $\lambda^h + \lambda^l > 1$. We will focus on this case, in which messages by i have some informative content.

We can intuitively see that if i's messages are believed to contain some information about σ_i, i should never want to misrepresent σ^h if b_i is high compared to the average bias of other players (and vice versa if b_i is low). In fact, we can show the following result:

Lemma 6. *Assume that $\lambda^h + \lambda^l > 1$. Then i always strictly prefers to truthfully reveal (i) σ^h if $b_i \geq \mathbb{E}\left[\frac{\sum_{j\in R_i, j\neq i} b_j}{n_{R_i}-1}\right]$ and (ii) σ^l if $b_i \leq \mathbb{E}\left[\frac{\sum_{j\in R_i, j\neq i} b_j}{n_{R_i}-1}\right]$.*

Proof. Consider case (i) and assume that the opposite was true, i.e. $\Delta U_i(\sigma^h) \leq 0$ for some $b_i \geq \mathbb{E}\left[\frac{\sum_{j\in R_i, j\neq i} b_j}{n_{R_i}-1}\right]$. Then, since $(\mu_{ji}^h - \mu_{ji}^l) > 0$ by assumption and $b_i \geq \mathbb{E}\left[\frac{\sum_{j\in R_i, j\neq i} b_j}{n_{R_i}-1}\right]$, it must be that $\frac{\mu_{ji}^h + \mu_{ji}^l}{2} - \mathbb{E}\left[\theta_i | \sigma_i\right] > 0$ or $\frac{\mu_{ji}^h + \mu_{ji}^l}{2} - p > 0$, which means $\left(\frac{\lambda^l}{\lambda^h - \lambda^l - 1} + \frac{\lambda^l - 1}{\lambda^h - \lambda^l + 1}\right) > 0$. But we know that $\lambda^h - \lambda^l - 1 < 0$ and $\lambda^h - \lambda^l + 1 > 0$ from $\lambda^h + \lambda^l > 1$, which implies that $\left(\frac{\lambda^l}{\lambda^h - \lambda^l - 1} + \frac{\lambda^l - 1}{\lambda^h - \lambda^l + 1}\right) < 0$. We can analogously prove (ii). □

Now we can consider which conditions need to be in place for an equilibrium to exist in which i tells the truth with probabilities λ^h and λ^l. To be clear: We are still considering pure equilibria, since i has a strict preference for lying or telling the truth for any b_i except for non-generic boundary cases. However, given F_i (the distribution of b_i), we can determine how often i's messages will be truthful once we have established for which b_i i wants to tell the truth and for which he wants to lie. We can think of λ^h and λ^l as the marginal probabilities of truth-telling by i.

Lemma 7. *There exists an equilibrium in which i truthfully reveals σ^h with marginal probability λ^h and truthfully reveals σ^l with marginal probability λ^l if and only if*

$$1 - F_i\left(\frac{\sum_{j\in R_i, j\neq i} b_j^e}{n_{R_i} - 1} + \left(p - \frac{1}{2}\right)\cdot\left(\frac{\lambda^l}{\lambda^h - \lambda^l - 1} + \frac{\lambda^l - 1}{\lambda^h - \lambda^l + 1}\right)\right) \leq \lambda^h$$

and

$$F_i\left(\frac{\sum_{j\in R_i, j\neq i} b_j^e}{n_{R_i} - 1} + \left(p - \frac{1}{2}\right)\left(\frac{\lambda^h - 1}{\lambda^h - \lambda^l - 1} + \frac{\lambda^h}{\lambda^h - \lambda^l + 1}\right)\right) \geq \lambda^l.$$

Both inequalities hold with equality if F_i is continuous at the argument.

Proof. From equation 8 we get that $\Delta U_i(\sigma_i) \geq 0 \Leftrightarrow$

$$b_i - \frac{\sum_{j\in R_i, j\neq i} b_j^e}{n_{R_i} - 1} \geq \frac{\mu_{ji}^h + \mu_{ji}^l}{2} - \mathbb{E}\left[\theta_i | \sigma_i\right].$$

Recall that $\mathbb{E}\left[\theta_i | \sigma_i = \sigma^h\right] = p$ and $\mathbb{E}\left[\theta_i | \sigma_i = \sigma^l\right] = 1 - p$. We can make use of the expression for $\mu_{ji}^h + \mu_{ji}^l$ that we have derived in (10) to get $\Delta U_i(\sigma^h) \geq 0 \Leftrightarrow$

$$b_i - \frac{\sum_{j\in R_i, j\neq i} b_j^e}{n_{R_i} - 1} \geq \left(p - \frac{1}{2}\right)\cdot\left(\frac{\lambda^l}{\lambda^h - \lambda^l - 1} + \frac{\lambda^l - 1}{\lambda^h - \lambda^l + 1}\right)$$

and $\Delta U_i(\sigma^l) \leq 0 \Leftrightarrow$

$$b_i - \frac{\sum_{j\in R_i, j\neq i} b_j^e}{n_{R_i} - 1} \leq \left(p - \frac{1}{2}\right)\left(\frac{\lambda^h - 1}{\lambda^h - \lambda^l - 1} + \frac{\lambda^h}{\lambda^h - \lambda^l + 1}\right).$$

In an equilibrium, the beliefs of the receivers of m_i must be correct on average. In this case, this means that it must be sufficiently likely for b_i to fulfill either of the two inequalities, which gives us the conditions from the proposition. If F_i is continuous at the argument, correct beliefs require that the inequalities hold with equality. If it is not, there could potentially be mixed equilibria in which for the borderline type, i mixes between different messages and beliefs are correct on average. $\square$

Note that that $\left(\frac{\lambda^h-1}{\lambda^h-\lambda^l-1}+\frac{\lambda^h}{\lambda^h-\lambda^l+1}\right)-\left(\frac{\lambda^l}{\lambda^h-\lambda^l-1}+\frac{\lambda^l-1}{\lambda^h-\lambda^l+1}\right)=2$. Lemma 7 consequently describes conditions on the distribution function F at two points that are $2p-1$ apart. In particular if F_i is continuous at these two points the conditions state that probability mass in the interval between these two points has to equal $\lambda^l+\lambda^h-1$. More importantly, the conditions can be used to show that player i babbles in a given room if F_i does not have enough probability mass around the average bias of the other players in the room. To be precise, if F_i has no probability mass in $\frac{\sum_{j\in R_i, j\neq i} b_j^e}{n_{R_i}-1} \pm (2p-1)$, then the conditions of lemma 6 imply $\lambda^l+\lambda^h=1$ and therefore uninformative messages.[7]

5.2.2. Proofs

Proof of proposition 7 on page 17.

Without loss of generality, let b_1 and b_n be the smallest and largest biases respectively. We can represent each bias as the expected value of a distribution that only places density on the values $b_1-(2p-1)$ and $b_n+(2p-1)$. For this set of distributions $\{F_1, F_2, \ldots, F_n\}$, the conditions of lemma 7 imply $\lambda^h+\lambda^l=1$, and hence there exists no equilibrium in which any of the players tells the truth. $\square$

Proof of proposition 8 on page 18.

We can construct a distribution F_i that has positive density on $\frac{\sum_{j\in R_i, j\neq i} b_j^e}{n_{R_i}-1}$, which means that the conditions of lemma 7 imply that there exists an equilibrium in which a message by i is informative.

To achieve full truth-telling (i.e. $\lambda^h=\lambda^l=1$), lemma 7 implies we would have to be able to construct an F_i that only has density inside the interval $\frac{\sum_{j\in R_i, j\neq i} b_j^e}{n_{R_i}-1} \pm \left(p-\frac{1}{2}\right)$. However, this would contradict our starting assumption that if b_i is b_i^e for sure, there exists no equilibrium in which i tells the truth. $\square$

[7] To be precise, both points at which F_i is evaluated in lemma 6 lie in the interior of the interval $\left[\frac{\sum_{j\in R_i, j\neq i} b_j^e}{n_{R_i}-1}-(2p-1), \frac{\sum_{j\in R_i, j\neq i} b_j^e}{n_{R_i}-1}+(2p-1)\right.$ and therefore F_i will be continuous at both points and equal to the same value if there is no probability mass in this interval. As the conditions in lemma 6 then hold with equality, they imply $\lambda^h+\lambda^l=1$ which in turn implies $\mu_{ji}^h-\mu_{ji}^l=0$.

Proof of proposition 9 on page 18.

By the symmetry of F, all F^κ have the same expected value. We can find a $\bar{\kappa}$ small enough so that F^κ has less than $\varepsilon' > 0$ probability mass within $\frac{\sum_{j \in R_i, j \neq i} b_j^e}{n_{R_i}-1} \pm (2p-1)$ for any $\kappa \leq \bar{\kappa}$. Then it follows from lemma 7 that there exists no equilibrium for which $\lambda^l + \lambda^h > 1 + \varepsilon'$. The result follows now from the continuity of (10) and the fact that $\mu_{ji}^h - \mu_{ji}^l = 0$ if $\lambda^h + \lambda^l = 1$. □

Proof of proposition 10 on page 18.

Let the lower (upper) bound of the support be $\underline{b}_i$ ($\bar{b}_i$). Note that by assumption $\underline{b}_i \leq \frac{\sum_{j \in R_i, j \neq i} b_j^e}{n_{R_i}-1} - (2p-1)$ and $\bar{b}_i \geq \frac{\sum_{j \in R_i, j \neq i} b_j^e}{n_{R_i}-1} + (2p-1)$ which implies by lemma 6 that player i sends uninformative messages in equilibrium. Now fix $\underline{b}^\varepsilon = \frac{\sum_{j \in R_i, j \neq i} b_j^e}{n_{R_i}-1} - (2p-1)$ and $\bar{b}^\varepsilon = \frac{\sum_{j \in R_i, j \neq i} b_j^e}{n_{R_i}-1} + (2p-1)$. This implies that $\mu_{ji}^h - \mu_{ji}^l \leq \varepsilon$ whenever the probability that $b_i \geq \bar{b}^\varepsilon$ plus the probability that $b_i < \underline{b}^\varepsilon$ is more than $1 - \varepsilon'$ for some $\varepsilon' > 0$ (by lemma 6 and the continuity of $\mu_{ji}^{\cdot}$ in λ^h and λ^l). Let $\overline{\sigma_{F_i}}^2$ be defined by

$$\overline{\sigma_{F_i}}^2 = (1-\varepsilon')\left(\frac{\bar{b}_i - b_i^e}{\bar{b}_i - \underline{b}_i}(\underline{b}_i - b_i^e)^2 + \frac{b_i^e - \underline{b}_i}{\bar{b}_i - \underline{b}_i}(\bar{b}_i - b_i^e)^2\right) + \varepsilon'\left(\frac{\bar{b}^\varepsilon - b_i^e}{\bar{b}^\varepsilon - \underline{b}^\varepsilon}(\underline{b}^\varepsilon - b_i^e)^2 + \frac{b_i^e - \underline{b}^\varepsilon}{\bar{b}^\varepsilon - \underline{b}^\varepsilon}(\bar{b}^\varepsilon - b_i^e)^2\right).$$

Any distribution with variance above $\overline{\sigma_{F_i}}$ has to have more than ε' probability mass above $\bar{b}^\varepsilon$ or below $\underline{b}^\varepsilon$ as $\overline{\sigma_{F_i}}$ is the variance of the distribution maximizing variance under the constraint that only $1 - \varepsilon'$ probability mass is outside the interval $[\underline{b}^\varepsilon, \bar{b}^\varepsilon]$. Consequently, any distribution with variance above $\overline{\sigma_{F_i}}$ will lead to $\mu_{ji}^h - \mu_{ji}^l \leq \varepsilon$. □

Proof of proposition 11 on page 19.

Fix 0 and a $b > 0$. Consider the distributions putting probability 1/2 on $-(p - 1/2)$ and 1/2 on $p - 1/2$ instead of 0 for sure and 1/2 on $b - (p - 1/2)$ and 1/2 on $b + (p - 1/2)$. Under segregation everyone is (just!) truthtelling. In any room including at least 1 player with another bias than the own one, a bias 0 (b) player will however lie if his bias is the lower (higher) element of the support:

Take for example a player with bias $b+p-1/2$ that got a low signal. Then $\Delta U(\sigma^l) > 0$ can be written as $b + p - 1/2 - \frac{\sum_{j \in R_i, j \neq i} b_j^e}{n_{R_i}-1} > (\mu_{ji}^h + \mu_{ji}^l)/2 - (1-p)$. The right hand side of this inequality is bounded from above by $p - 1/2$ because $\mu_{ji}^h \leq p$ and $\mu_{ji}^l = 1 - p$ by lemma 6 according to which $\lambda^h = 1$. As $b - \frac{\sum_{j \in R_i, j \neq i} b_j^e}{n_{R_i}-1} > 0$, the claim follows.

To compute welfare under a non-segregated scenario, we need to compute $\mathbb{E}[(\mu_{ij} - \theta_j)^2]$. Take, for example, a player j with biases in $\{b - p + 1/2, b + p - 1/2\}$. We showed that this player always sends the high signal if $b_i = b + p - 1/2$ if at least one player of the other group is in his room. The most informative messaging strategy of such a player in such a room is therefore truthtelling when $b_i = b - p + 1/2$ and sending the high message otherwise. This implies $\lambda^h = 1$ and $\lambda^l = 1/2$ and therefore $\mu_{ij}^h = (1+p)/3$ and $\mu_{ij}^l = 1-p$. In this case,

$$\begin{aligned}
\mathbb{E}[(\mu_{ij}-\theta_j)^2] &= \frac{1}{2}\left[\frac{1}{2}\left\{p\left(\frac{1+p}{3}-1\right)^2+(1-p)(-p)^2\right\}+\frac{1}{2}\left\{p(1-p)^2+(1-p)\left(\frac{1+p}{3}\right)^2\right\}\right] \\
&\quad +\frac{1}{2}\left[\frac{1}{2}\left(\frac{1+p}{3}-1\right)^2+\frac{1}{2}\left(\frac{1+p}{3}\right)^2\right] \\
&= \frac{1}{4}\left[(1+p)\frac{p^2-4p+4}{9}+(1-p)p^2+p(1-p)^2+(2-p)\frac{1+2p+p^2}{9}\right] \\
&= \frac{1}{4}\left[\frac{2}{3}+\frac{4}{3}p-\frac{4}{3}p^2\right].
\end{aligned}$$

Following the derivations of player i's utility in a room that contains players of both groups, see the proof of proposition 1, we can write player i's utility if all players are in the same fully integrated room – and choose the best possible messaging strategy corresponding to $\lambda^h = 1$ ($\lambda^h = 1/2$) and $\lambda^l = 1/2$ ($\lambda^l = 1$) for players with expected bias $b_i^e = b$ ($b_i^e = b$) – as

$$\begin{aligned}
U_i^{int} &= -\alpha\sum_{j\neq i}\left\{(b_j-b_i)^2\right\}-\left[n+\alpha(n-1)n\right]/4+(1/4-p(1-p))(1+\alpha(n-1)) \\
&\quad +\left(1/4-[2/3+p4/3-p^2 4/3]/4\right)\left[n-1+\alpha\sum_{j\neq i}\{n-1\}\right]
\end{aligned}$$

while his expected payoff under full segregation is

$$U_i^{seg} = -\alpha\sum_{j\neq i}\left\{(b_j-b_i)^2\right\}-\left[n+\alpha(n-1)n\right]/4+(1/4-p(1-p))(n/2+\alpha(n-1)n/2).$$

U_i^{seg} exceeds U_i^{int} if and only if

$$\begin{aligned}
&(1/4-p(1-p))(1+\alpha(n-1))(n/2-1) \geq \left(1/4-[2/3+p4/3-p^2 4/3]/4\right)\left[n-1+\alpha(n-1)^2\right] \\
\Leftrightarrow\ &(1-4p+4p^2))(1+\alpha(n-1))(n/2-1) \geq \left(1/3-p4/3+p^2 4/3\right)\left[n-1+\alpha(n-1)^2\right] \\
&\Leftrightarrow 3(1+\alpha(n-1))(n/2-1) \geq n-1+\alpha(n-1)^2 \\
&\Leftrightarrow \frac{3}{2}(1+\alpha(n-1))\frac{n-2}{n-1} \geq 1+\alpha(n-1) \\
&\Leftrightarrow \frac{n-2}{n-1} \geq \frac{2}{3}
\end{aligned}$$

which is true for $n \geq 4$. As the payoffs do not differ across players in each of the two scenarios, welfare is higher under segregation than under integration given that $n \geq 4$.

To see that other room configurations cannot improve welfare, start from full segregation. Moving k players from room 1 to room 2 will lead to less information for the remaining players in room 1. Suppose nevertheless that this move was welfare increasing. Then players in the new room 2 must have better information than under segregation. Note that by assumption the most informative strategy players could possibly adopt in the new room is $\lambda^h = 1$ ($\lambda^h = 1/2$) and $\lambda^l = 1/2$ ($\lambda^l = 1$) for players with expected bias $b_i^e = b$ ($b_i^e = b$). Assume that this strategy is an equilibrium in the new room 2 (if it is not, this step increases the welfare gain over segregation). But then it is clearly optimal to move the remaining players from room 1 to room 2 as well (if this strategy remains an equilibrium): This improves information for all players. But this would imply $U_i^{int} > U_i^{seg}$ which contradicts what we showed above. □

6. Public information

Here we add a public information component. Our main interest is in the comparative statics of the weight of this public information component, i.e. if more information becomes publicly available, how will communication be affected? The main mechanisms of our model remain unchanged in such an extension. Public information, however, crowds out incentives to tell the truth: If we increase the importance of public information, it becomes more tempting to mislead players with different biases. As private information is relatively less important if there is much public information, other players respond less strongly to one's message (if the message is believed to be truthful) and consequently players are less disciplined by the danger of misleading their audience "too much".

More formally, we consider in this extension a state of the world $\theta = \tau\theta_0 + (1 - \tau)\sum_{i=1}^{n}\theta_i$; that is, we add – compared to the main model of the paper – an element $\theta_0 \in \{0, 1\}$ which receives a weight τ. As for all other θ_i, there is also binary a signal σ_0. This signal has accuracy $p_0 > 1/2$, i.e. $Pr(\sigma_0 = \sigma^h|\theta_0 = 1) = Pr(\sigma_0 = \sigma^l|\theta_0 = 0)$, and is observed by all players. Consequently, all players share the same belief about θ_0 which is denoted by μ_0. Everything else is as in the main model of the paper.[8]

The optimal choice of action is now

$$a_i^* = b_i + \mathbb{E}[\theta] = b_i + \tau\mu_0 + (1 - \tau)\sum_{j=1}^{n}\mu_{ij}.$$

Note that the proof of lemma 1 goes through with straighforward adaptations. In

[8]One way to interpret the weights is the following: Suppose there is a continuum of $\tilde{\theta}$ of unit length, say $[0, 1]$. Each $\tilde{\theta} \in [0, 1]$ is either 0 or 1 and agents try to match the average $\tilde{\theta}$ (plus their bias) with their action. For a set $\Theta_0 \subset [0, 1]$ of measure τ, a public signal is available (there could be several public signals which are then aggregated; the only thing that matters is that everyone has the same expectation about the value of the average $\tilde{\theta}$ in Θ_0). Each player i has a private signal about the average value of the $\tilde{\theta} \in \Theta_i$ where Θ_i has measure $(1-\tau)/n$ and we assume that all Θ_i are pairwise disjoint. The comparative static with respect to τ answers then the question: What happens if information/signals that used to be privately held by some expert are now publicly available?

particular,

$$U_i(m_i) = \mathbb{E}\left[const - \alpha \sum_{j\in R_i, j\neq i} \left(a_j(m_i, m_{-i,R_i}, \sigma_j) - b_i - \tau\theta_0 - (1-\tau)\sum_{k=1}^{n}\theta_k \right)^2 \middle| \sigma_i \right].$$

which leads to

$$\begin{aligned}
\Delta U_i(\sigma_i) &= (U_i(m^h) - U_i(m^l))/\alpha \\
&= -\sum_{j\in R_i, j\neq i} \mathbb{E}\Big[(1-\tau)^2 {\mu^h_{ji}}^2 - (1-\tau)^2 {\mu^l_{ji}}^2 \\
&\quad +2(1-\tau)(\mu^h_{ji} - \mu^l_{ji}) \left(b_j - b_i + \tau(\mu_0 - \theta_0) + (1-\tau)\sum_{k\neq i}(\mu_{jk} - \theta_k) - (1-\tau)\theta_i \right) \Big| \sigma_i \Big] \\
&= -2(1-\tau)(\mu^h_{ji} - \mu^l_{ji}) \sum_{j\in R_i, j\neq i} \left[(1-\tau)\frac{\mu^h_{ji} + \mu^l_{ji}}{2} + b_j - b_i - (1-\tau)\mathbb{E}\left[\theta_i | \sigma_i\right] \right] \\
&= 2(1-\tau)(\mu^h_{ji} - \mu^l_{ji})(n_{R_i} - 1) \left[-(1-\tau)\frac{\mu^h_{ji} + \mu^l_{ji}}{2} - \frac{\sum_{j\in R_i, j\neq i} b_j}{n_{R_i} - 1} + b_i + (1-\tau)\mathbb{E}\left[\theta_i | \sigma_i\right] \right].
\end{aligned}$$

Using this expression instead of (15) in the paper the proof of lemma 1 applies and we can concentrate on pure strategy equilibria.

A result similar to theorem 1 in the paper now follows immediately from the expression above:

Theorem 4. *Let $\bar{b} = \frac{\sum_{k\in R} b_k}{n_R}$ be the mean bias of players in room R. In the most informative equilibrium in this room, a player i tells the truth if and only if*

$$b_i \in \left[\bar{b} - \frac{n_R - 1}{n_R}(p - \frac{1}{2})(1-\tau),\ \bar{b} + \frac{n_R - 1}{n_R}(p - \frac{1}{2})(1-\tau) \right]$$

and babbles otherwise.

Proof. Consider the difference between lying and truth-telling for player i, i.e. ΔU_i as derived above. For every non-babbling player $\mu^h_{ji} = p$ and $\mu^l_{ji} = 1 - p$ (as we can concentrate on pure strategy equilibria) which implies that the necessary equilibrium condition $\Delta U_i(\sigma^h) \geq 0$ simplifies to

$$\begin{aligned}
b_i - \frac{1}{n_R - 1}\sum_{j\in R_i, j\neq i} b_j &\geq (\frac{1}{2} - p)(1-\tau) \\
\frac{n_R}{n_R - 1} b_i - \frac{1}{n_R - 1}\sum_{k\in R_i} b_k &\geq (\frac{1}{2} - p)(1-\tau) \\
b_i &\geq \bar{b} - \frac{n_R - 1}{n_R}\left(p - \frac{1}{2} \right)(1-\tau).
\end{aligned}$$

If this inequality does not hold, player i will not use the truthful strategy in the most informative equilibrium and therefore he will babble in the most informative equilibrium.

We can analogously solve for $\Delta U_i(\sigma^l) \leq 0$ and get the interval in the theorem. □

Theorem 4 implies that a higher weight on public information reduces the length of the truthtelling interval. That is, public information crowds out communicated private information in a given room. This implies that also under the welfare optimal room allocation less private information will be communicated for higher τ. [9]

Proposition 12. *Let $\tau^h > \tau^l$. In the welfare optimal room allocation the total amount of communicated information is (weakly) less under τ^h than under τ^l*

Proof. Take the welfare optimal room allocation under τ^h. Using the same room allocation under τ^l will create at least as much information as under τ^h by theorem 4. (Adapting the room allocation may increase the number of communicated pieces of information further.) □

Note, however, that welfare is not necessarily decreasing in τ. The positive effect of more public information counteracts the negative effect of less private information. The overall effect is generally ambiguous.[10]

To think about segregation, let us focus on the binary bias case, i.e. $b_i \in \{0, b\}$ for $i = 1, \dots n$. Recall that in this setting the welfare optimal room allocation has to fall into one of the following four categories, see section 1:

1. full integration (either with everyone truthtelling or only the majority)
2. full segregation
3. a mixed room in which only majority players are truthtelling and one room with some minority players
4. a mixed room in which everyone is truthtelling and an extra room with some majority players.

[9] Note that welfare – for a given τ – is, as in the paper, proportional to the number of communicated pieces of information. The derivation goes through with the obvious adaptations leading to

$$\begin{aligned} W &= -\alpha \sum_{i=1}^{n} \sum_{j \neq i} \{(b_j - b_i)^2\} - \frac{1}{4}(1-\tau)^2 n^2 \left[1 + \alpha(n-1)\right] \\ &\quad + (1-\tau)^2 (p - \frac{1}{2})^2 (1 + \alpha(n-1)) \sum_i \zeta_i - n(1 + \alpha(n-1)) p_0 (1 - p_0) \tau^2. \end{aligned}$$

[10] To construct an example where welfare is locally decreasing in τ it is enough to choose parameter values such that truthtelling in a fully integrated room is just possible, i.e. some player is indifferent between truthtelling and not. Marginally increasing τ in this situation will discretely lower the $\sum_i \zeta_i$ in the welfare function while affecting all other terms continuously. Hence, a slightly higher τ will decrease welfare.

The detrimental effect of higher τ on communication, now immediately implies that an increase in τ leads to more segregation in one of the following ways: First, full segregation might become optimal for higher τ (as truthful communication in a mixed or fully integrated room is possible to a lesser degree). Second, less minority players can remain in a mixed bias room in which only majority players are truthtelling (as otherwise majority players babble). Third, less majority players can remain in a mixed room in which all players are truthtelling (as otherwise minority players lose their incentive to be truthful).

The main intuition behind these results is that more public information implies less influence of i's message on j's decision because i holds less relevant information privately. With less influence lying is less costly as j will "overshoot" less (when communicating a high message instead of a low message). This intuition also suggests that for sufficiently high τ communication between players of different biases is impossible and therefore segregation by bias is welfare optimal and an equilibrium. The following result states this formally for generic configurations of biases $\mathcal{B}$ (not necessarily binary). Assume that $\mathcal{B}$ is generic in the sense that player i's bias is not the average of other players' biases (whose biases are distinct from b_i).[11]

Theorem 5. *For generic $\mathcal{B}$, there exists a $\bar{\tau} < 1$, such that full segregation based on biases is both welfare optimal and an equilibrium if $\tau \geq \bar{\tau}$.*

Proof. Theorem 4 together with our genericity assumption implies that for a given room in which at least two players differ in their bias there is a $\bar{\tau}_R < 1$ such that babbling is the unique equilibrium of the messaging game in this room if $\tau \geq \bar{\tau}_R$. By the finiteness of the set of players, the number of possible room configurations is finite and therefore $\max_R \bar{\tau}_R$ exists and is strictly less than 1. Take $\bar{\tau} = \max_R \bar{\tau}_R$. Then babbling is the unique equilibrium of all non-segregated rooms and it is clear that segregation maximizes the number of communicated pieces of information and therefore welfare. Also no player wants to deviate from a segregated room to another room as this would lead to babbling by all players in the room he deviates to (and also deprives the players of the segregated room of his truthful message). □

The results of this extension suggest an additional mechanism for why segregation occurs and how it may differ over time and between settings. In communication settings, both private or professional, where almost all relevant information is private information of the participants, it may be easier to achieve communication and hence segregation is less useful. But when the discussion is about national politics, for example, where almost all information is public and people's private knowledge and experiences are only a small facet of a larger whole, more segregation may be desirable.

[11]More precisely, the assumption is that $b_i \neq \sum_{b_j \in \mathcal{B} \setminus \{b_i\}} \frac{\tilde{n}_{b_j}}{\sum_k \tilde{n}_{b_k}} * b_j$ for any $\tilde{n}_{b_j} \in \{0, 1, \ldots, n_{b_j}\}$ where n_{b_j} is the number of players with bias b_j.

The results also suggest that progress in information technologies, which make information publicly accessible that in earlier times was held only by experts, may lead to less (truthful) private communication and more segregation. Note, however, that this does not necessarily imply that players make less informed decision as the additional public information can more than outweigh the informational loss from less private communication.

7. States correlated within bias groups

This section considers a variation of the model in which players with a similar bias have similar information. This feature implies that communication across bias groups is even more desirable from a welfare perspective. However, we will show that for the same reasons as in the paper such communication is infeasible in equilibrium if bias differences are large.

We will focus on a model setup with two bias groups, i.e. $\mathbb{B} = \{0, b\}$. Without loss of generality let players $i = 1, 2, \dots, n_0$ have bias $b_i = 0$ and players $i = n_0 + 1, \dots, n_0 + n_b$ have bias $b_i = b$. We will introduce similarity of information by assuming that θ_i and θ_j are positively correlated if either $i, j \in \{1, \dots, n_0\}$ or $i, j \in \{n_0 + 1, \dots, n_0 + n_b\}$. However, we maintain the assumptions that (i) θ_i and θ_j are uncorrelated if $i \in \{1, \dots, n_0\}$ and $j \in \{n_0 + 1, \dots, n_0 + n_b\}$, (ii) signal σ_i is noisy and independent of σ_j and θ_j conditional on θ_i, (iii) that $\theta_i \in \{0, 1\}$ and the marginals are such that $\mathbb{E}[\theta_i] = 1/2$ (this latter assumption is for covenience of notation only). We will not be more specific about the correlation but want to point out the two extreme cases within this framework: First, perfect correlation within bias groups. In this case, all players with the same bias receive effectively information about the same underlying variable. Second, complete independence which is the case we analyze in the paper.

Given the flexible formulation, it is unsurprising that a closed form solution no longer exists. However, we will be able to show a result similar to the one in the main text in this setup. In a given room including players of both bias groups essentially no information transmission is possible if b if sufficiently large.

The following proposition states that the amount of information transmitted in a given room with players of both biases is less than an arbitrary $\varepsilon > 0$ if b is large enough.[12]

Proposition 13. *Let R be a room containing at least one player with bias 0 and at least one player with bias b. For every $\varepsilon > 0$, there exists a b_ε such that $\mathbb{E}_{m_{-i},\sigma_j}[\mu_j(h) - \mu_j(l)|\sigma_i] < \varepsilon$ for every player $i \in R$ in every equilibrium of the communication stage.*

Proof of proposition 13: Clearly, it is still optimal to choose action $a_i = b_i +$

[12] We denote here the action of player j when i sends message l by $a_j(l)$. This action depends also on σ_j and messages of other players but we suppress this dependence in the interest of readability. Similarly $\mu_j(l)$ is j's belief about θ if i sends message l which again depends also on σ_j and other players' messages.

$\mathbb{E}[\theta|\sigma_i, m_{R_i}]$ where m_{R_i} are the messages observed by player i.

For concreteness take a player i with bias $b_i = 0$ and compare the difference in expected utility of this player when sending message h and message l (we neglect that the expectation is conditional on σ_i to avoid cluttering of notation):

$$\begin{aligned}\Delta U_i &= \alpha \sum_{j\neq i, j\in R_i} \mathbb{E}\left[a_j(l)^2 - a_j(h)^2 - 2\theta(a_j(l) - a_j(h))\right] \\ &= \alpha \sum_{j\neq i, j\in R_i} \mathbb{E}\left[\mu_j(l)^2 - \mu_j(h)^2 + 2b_j(\mu_j(l) - \mu_j(h)) - 2\theta(\mu_j(l) - \mu_j(h))\right] \\ &= -2\alpha \sum_{j\neq i, j\in R_i} \mathbb{E}\left[(\mu_j(h) - \mu_j(l))\left(\frac{\mu_j(h) + \mu_j(l)}{2} - \theta + b_j\right)\right].\end{aligned}$$

As $-n_0 - n_b \leq (\mu_j(h) + \mu_j(l))/2 - \theta \leq n_0 + n_b$, choosing $b_\varepsilon = (n_0 + n_b) + (n_0 + n_b)^2/\varepsilon$ is sufficient for $\Delta U_i < 0$ regardless of σ_i. □

8. Alternative signal technologies

In this section, we consider three variations of the model in the paper. The first is a straightforward extensions in which we allow for more signals than just the binary signal structure considered in the paper. (We can also allow for more states but this is relatively immaterial in our setting.) The second variation considers goes a bit further by considering a continuum of signals. The third changes the signal structure such that no longer each player receives a signal about "his state" θ_i as in the main text but instead all players receive a noisy signal about the same one-dimensional state θ. For all variations we show that our main result that integration is optimal and an equilibrium if there is little polarization while segregation is optimal and an equilibrium if there is a lot of polarization continue to hold. The main shortcoming of the first and third variation is that for intermediate values of polarization it is no longer possible to determine the most informative equilibrium of the messaging game as we can no longer rule out that this equilibrium involves mixed strategies. The second variation allows only a closed form solution of the most informative messaging equilibrium for particular distributions, e.g.the uniform distribution. This makes each variation less tractable than the model of the main paper.

8.1. Larger signal and state space

Now allow for an arbitrary finite number of states, biases and signals. We keep the assumption that states and signals of different players are independent and that player i receives a signal that is partially informative about state θ_i (but independent about all other states). We also keep the utility function, i.e. the additive structure. The message space equals the signal space and we assume that lower signals lead to a lower expected value of θ_i. For notational simplicity let the signal be the posterior it leads to, i.e. $\sigma_i = \mathbb{E}[\theta_i|\sigma_i]$.

Following similar steps as in the main text, we can derive the expected utility difference between sending two messages labeled as high (h) and low (l). Let μ_{ij}^h denote the expected value that j assigns to θ_i upon receiving message h (given some equilibrium messaging strategy by i). The expected utility difference can then, similarly to above, be derived as

$$\begin{aligned}
\Delta U_i(\sigma_i) &= \sum_{j\in R_i, j\neq i} (\mu_{ji}^l)^2 - (\mu_{ji}^h)^2 + 2(\mu_{ji}^l - \mu_{ji}^h)(b_j - b_i - \mathbb{E}[\theta_i|\sigma_i]) \\
&= \sum_{j\in R_i, j\neq i} \left(\mu_{ji}^l - \mu_{ji}^h\right)\left[\left(\mu_{ji}^l + \mu_{ji}^h\right) + 2(b_j - b_i - \sigma_i)\right] \\
&= -2(n_{R_i} - 1)\left(\mu_{ji}^h - \mu_{ji}^l\right)\left[\frac{\mu_{ji}^l + \mu_{ji}^h}{2} + \sum_{k\in R_i, k\neq i}\left\{\frac{b_k}{n_{R_i} - 1}\right\} - b_i - \sigma_i\right]
\end{aligned}$$

This expression implies that a truthtelling equilibrium exists if and only if for every player i and every $\sigma_i^l < \sigma_i^h$

$$\sigma_i^l \leq \frac{\sigma_i^l + \sigma_i^h}{2} + \sum_{k\in R_i, k\neq i}\left\{\frac{b_k}{n_{R_i} - 1}\right\} - b_i \leq \sigma_i^h$$

$$\Leftrightarrow \left|\sum_{k\in R_i, k\neq i}\left\{\frac{b_k}{n_{R_i} - 1}\right\} - b_i\right| \leq \frac{\sigma^h - \sigma^l}{2}.$$

If we assume that all players have the same signal space, this condition is tightest for the player whose bias b_i is furthest away from the other players' average bias, $\sum_{j\in R_i, j\neq i} b_j/(n_{R_i} - 1)$, and for the two signals that are closest together.

It is immediate from the expression above that (i) truthtelling is impossible if bias differences are too high, (ii) adding moderates can establish truthtelling as it can move the average of the other players closer to each player's bias (e.g. consider a room with 2 people with differing biases, then adding a player with the average bias can only help). To state this formally, consider first the expected payoff of player i when choosing room

R_i and expecting a given (e.g. equilibrium) room allocation:

$$
\begin{aligned}
-\mathbb{E}\Bigg[& \Bigg(\sum_{j \in R_i^{truth}, j \neq i} (\mu_{ij} - \theta_j) + \sum_{j \notin R_i, j \in R_i^{bab}} (\bar{\mu}_j - \theta_j) \Bigg)^2 \\
& + \alpha \sum_{j \in R_i, j \neq i} \Bigg(b_j - b_i + \sum_{k \in R_i^{truth} \cup \{j\}} (\mu_{jk} - \theta_k) + \sum_{k \notin R_i, k \in R_i^{bab} \setminus \{j\}} (\bar{\mu}_k - \theta_k) \Bigg)^2 \\
& + \alpha \sum_{j \notin R_i} \Bigg(b_j - b_i + \sum_{k \in R_j^{truth} \cup \{j\}} (\mu_{jk} - \theta_k) + \sum_{k \notin R_i, k \in R_j^{bab} \setminus \{j\}} (\bar{\mu}_k - \theta_k) \Bigg)^2 \Bigg]
\end{aligned}
$$

where we denote $\mathbb{E}[\theta_j]$ as $\bar{\mu}_j$, the set of players babbling in room R_j in the messaging equilibrium of the given room allocation as R_j^{bab} and the set of players sending truthful messages in room R_j in the messaging equilibrium of the given room allocation as R_j^{truth}. Note that most of the terms drop out in the expression above as signals are assumed to be independent and therefore $\mathbb{E}[\mu_{ij} - \theta_j] = 0$ and also $\mathbb{E}[(\mu_{ij} - \theta_j)(\mu_{ik} - \theta_k)] = 0$. Consequently, the expression above can be rewritten as

$$
\begin{aligned}
& - \sum_{j \in R_i^{truth}, j \neq i} \mathbb{E}\left[(\mu_{ij} - \theta_j)^2\right] - \sum_{j \notin R_i, j \in R_i^{bab}} \mathbb{E}\left[(\bar{\mu}_j - \theta_j)^2\right] \\
& - \alpha \sum_{j \in R_i, j \neq i} (b_j - b_i)^2 - \alpha \sum_{j \in R_i, j \neq i} \sum_{k \in R_i^{truth} \cup \{j\}} \mathbb{E}\left[(\mu_{jk} - \theta_k)^2\right] - \alpha \sum_{j \in R_i, j \neq i} \sum_{k \notin R_i, k \in R_i^{bab} \setminus \{j\}} \mathbb{E}\left[(\bar{\mu}_k - \theta_k)^2\right] \\
& - \alpha \sum_{j \notin R_i} (b_j - b_i)^2 - \alpha \sum_{j \notin R_i} \sum_{k \in R_j^{truth} \cup \{j\}} \mathbb{E}\left[(\mu_{jk} - \theta_k)^2\right] - \alpha \sum_{j \notin R_i} \sum_{k \notin R_i, k \in R_j^{bab} \setminus \{j\}} \mathbb{E}\left[(\bar{\mu}_k - \theta_k)^2\right]
\end{aligned}
$$

As we cannot rule out mixed strategies, this expression will not simplify as neatly as in the main text. However, we can already see from here that a player's payoff is higher if another player is truthtelling than when he is babbling or mixing. This observation will be enough for our purposes.

To state our results we first introduce some notation. Let $\underline{\sigma} = \min\{|\sigma^j - \sigma^k| : j \neq k, \sigma^j, \sigma^k \in \Sigma\}$ and $\bar{\sigma} = \max\{|\sigma^j - \sigma^k| : j \neq k, \sigma^j, \sigma^k \in \Sigma\}$ and furthermore, $\bar{b} = \max_i\{|nb_i - \sum_j b_j|\}$. We will denote by $\mathcal{B}_\eta$ the set of biases scaled by η; that is, it contains all the elements ηb_i. We will use this to talk about more spread out biases. If the set of biases is $\mathcal{B}_\eta$ with $\eta > 1$, then biases are more spread out.

Proposition 14. *If $\underline{\sigma} \geq 2\bar{b}/(n-1)$, then a single room in which all players are truthtelling is both welfare maximizing and an equilibrium.*
Let the set of biases be $\mathcal{B}_\eta$ and fix all parameter values apart from η. Generically, full separation is welfare maximizing and an equilibrium if η is sufficiently high.

Proof of proposition 14: Recall that a truthtelling equilibrium exists if and only if for all players i $\left|\sum_{k\neq i}\{b_k/(n-1)\} - b_i\right| \leq (\sigma^h - \sigma^l)/2$ for every $\sigma^h > \sigma^l$ in Σ. This can be rewritten as $\left|\sum_k\{b_k\} - nb_i\right|/(n-1) \leq (\sigma^h - \sigma^l)/2$. The condition in the proposition ensures that this inequality holds for all players and all signals. Clearly, having all players in one room and telling the truth is welfare optimal whenever it is feasible.

If $\left|\sum_{k\in R_i, k\neq i}\{b_k/(n-1)\} - b_i\right| > (\sigma^h - \sigma^l)/2$, then i will not be truthful when receiving either signal σ^l or σ^h. Generically, $\left|\sum_{k\in R_i, k\neq i}\{b_k/(n-1)\} - b_i\right| \neq 0$ for any room configuration containing players from more than one bias group. (This follows from the finiteness of players which obviously implies that the number of such room configurations is finite.) Now observe that the left hand side of the non-truthtelling inequality is scaled by η while the right hand side is not. That is, for η sufficiently high player i will report the highest (lowest) signal in Σ in all rooms in which $\sum_{k\in R_i, k\neq j} b_k < n_{R_i} b_i$ ($\sum_{k\in R_i, k\neq j} b_k > n_{R_i} b_i$). Put differently, any room that contains one or more players of a bias not equal to b_i will lead to totally uninformative messages by i if η is sufficiently high. For high enough η, this holds true for all players and it is then obvious that full separation is both welfare maximizing and an equilibrium. □

8.2. Continuum of signals

In this subsection we consider the messaging game in a setup that differs from the one in the paper by assuming that the signal space is not binary but a continuum. That is, we take the room allocation as given and analyze equilibrium messaging strategies. Room choice is considered briefly towards the end of the section. Occasionally, we will refer to a player's signal as his "type". Instead of replicating the derivation of $\Delta U_i(\sigma_i)$ from section 8.1 we simply refer the reader at some points to it.

The main reason why our model is so tractable is that we can consider equilibrium incentives player by player. That is, player j's messaging strategy does not affect player i's incentives when deciding which message to take. This can be nicely seen from (11) where the only factors influencing preferences over messages are the beliefs induced by the messages, the bias distribution in the room and player i's signal.

We will first derive a few results that apply to general finite as well as infinite signal spaces. Signals are without loss of generality viewed as posteriors, i.e. $\sigma_i^k = \mathbb{E}[\theta_i|\sigma_i^k]$. Similarly, messages can – in equilibrium – be equated with the beliefs they induce. The following lemma states that the support of player i's message strategy is quite small: In fact, each type mixes at most between two messages and these messages are in some sense "adjacent".

Lemma 8. *In equilibrium, the support of type σ_i^k's strategy consists of at most two elements. If type σ_i^k mixes between two messages, then there is no message inducing a belief in between the two beliefs induced by the messages in his support.*

Proof of lemma 8: The indifference condition requires that σ_i is indifferent between any two messages in his support. Denoting the messages as l and h which lead in equilibrium to beliefs μ^l and μ^h (by the other players concerning θ_i), this indifference condition can, as derived in section 8.1, be written as

$$\frac{\mu^l + \mu^h}{2} + \sum_{k \in R_i, k \neq i} \left\{ \frac{b_k}{n_{R_i} - 1} \right\} - b_i - \sigma_i = 0. \tag{11}$$

The crucial insight is that – given that σ_i is indifferent between μ^l and μ^h, σ_i strictly prefers inducing any belief $\tilde{\mu} \in (\mu^l, \mu^h)$ to either μ^l or μ^h. To see this note that

$$\frac{\mu^l + \tilde{\mu}}{2} + \sum_{k \in R_i, k \neq i} \left\{ \frac{b_k}{n_{R_i} - 1} \right\} - b_i - \sigma_i < 0 < \frac{\tilde{\mu} + \mu^h}{2} + \sum_{k \in R_i, k \neq i} \left\{ \frac{b_k}{n_{R_i} - 1} \right\} - b_i - \sigma_i$$

by the indifference condition. This implies that $\tilde{\mu}$ is strictly preferred to μ^l and μ^h (see $\Delta U_i(\sigma_i)$ as derived in section 8.1). It follows that a type can only mix between two messages μ^l and μ^h in equilibrium if these two beliefs are "adjacent", i.e. there is no message inducing a belief between μ^l and μ^h. □

In case of mixing, each message is only used by few signal types. Furthermore, there is a standard order property in the sense that higher types send higher messages.

Lemma 9. *Each message is used by at most two types that use truly mixed strategies. If a type σ_i mixes between μ^l and $\mu^h > \mu^l$, then σ_i^k is the highest (lowest) type using message μ^l (μ^h).*

Proof of lemma 9: Suppose to the contrary that three types σ_i^k with $k = 1, 2, 3$ (i) use truly mixed strategies and (ii) use a message inducing belief μ with positive probability. As each type mixes only over two adjacent messages (see lemma 8), this would imply that at least two of the three types have the same support. Clearly, indifference condition (11) cannot be satisfied for different types and the same support μ^l and μ^h. Consequently, each message is used at most by two types that mix.

From the indifference condition, (11), and the expression $\Delta U_i(\sigma_i)$ it is clear that all types below (above) σ_i strictly prefer μ^l over μ^h (μ^h over μ^l). □

The order property of the previous lemma can be extended. In equilibrium, higher signal types send weakly higher messages. This does not exclude the possibility that one signal type mixes or that several signal types pool on the same message.

Lemma 10. *The induced belief μ^k is weakly increasing in the received signal σ_i^k.*

Proof of lemma 10: Take two signal types σ_i^h and σ_i^l with $\sigma_i^h > \sigma_i^l$. Suppose contrary to the lemma that $\mu(\sigma_i^l) > \mu(\sigma_i^h)$. In equilibrium σ_i^l must prefer sending his message to sending the message that σ_i^h sends in equilibrium, i.e.

$$\frac{\mu(\sigma_i^l) + \mu(\sigma_i^h)}{2} + \sum_{k \in R_i, k \neq i} \left\{ \frac{b_k}{n_{R_i} - 1} \right\} - b_i - \sigma_i^l \leq 0.$$

If the previous inequality holds, then it holds strictly with σ_i^h in place of $\sigma_i^l < \sigma_i^h$. That is, σ_i^h strictly prefers $\mu(\sigma_i^l)$ over $\mu(\sigma_i^h)$ which contradicts that σ_i^h induces $\mu(\sigma_i^h)$ in equilibrium. □

After these preliminaries, we turn now to a model with a continuum of signals. Let signal σ_i, i.e. player i's ex post belief $\mathbb{E}[\theta_i]$, be distributed according to some distribution Φ with density $\phi > 0$ on an interval, say $[0, 1]$ for simplicity. Lemmas 9 and 10 imply then that the equilibrium is a partition of $[0, 1]$. Note that truthfulness, i.e. truthfully revealing one's signal no matter what the signal is, is only feasible if $\Delta_i \equiv b_i - \sum_{k \in R_i, k \neq i} b_k/(n_{R_i} - 1) = 0$. That is, truthfullness is only an equilibrium if all players in a room share the same bias. Furthermore, the partition will be finite with the maximal number of partition elements being less than $1 + 1/(2|\Delta_i|)$. Finiteness is straightforward: If the partition was not finite, there would be types σ_i for which $\mu(\sigma_i) \approx \sigma_i$ and messages arbitrarily close to σ_i exist. But for $\Delta_i \neq 0$, some of these types would clearly want to misrepresent. The upper bound on the number of partition elements follows from the following observation: Let σ_i^{t-1}, σ_i^t and σ_i^{t+1} be consecutive partition boundary type in an equilibrium partition. Then σ_i^t has to be indifferent between the two messages $\mu^t = \mathbb{E}\left[\sigma_i | \sigma_i \in [\sigma_i^{t-1}, \sigma_i^t]\right]$ and $\mu^{t+1} = \mathbb{E}\left[\sigma_i | \sigma_i \in [\sigma_i^t, \sigma_i^{t+1}]\right]$ which means

$$\mu^{t+1} + \mu^t = 2\sigma_i^t + 2\Delta_i.$$

As $\mu^{t+1} \leq \sigma_i^{t+1}$ and $\mu^t \leq \sigma_i^t$, this implies that $\sigma_i^{t+1} + \sigma_i^t \geq 2\sigma_i^t + 2\Delta_i$ or equivalently $\sigma_i^{t+1} - \sigma_i^t \geq 2\Delta_i$. Hence, every partition element (with exception of the first) has length of at least $2\Delta_i$ if $\Delta_i > 0$ (for $\Delta_i < 0$ a similar argument using lower instead of upper bounds for $\mu^\cdot$ works analogously). If a partition equilibrium with T partition elements exists, then it can be computed similarly to Crawford and Sobel (1982): Say $\Delta_i > 0$ and denote the partition by $(\sigma_i^0 = 0, \sigma_i^1, \ldots, \sigma_i^T = 1)$. For $t \in \{1, \ldots, T-1\}$, the indifference condition of σ_i^t determines σ_i^{t+1}, i.e.

$$\frac{\int_{\sigma_i^t}^{\sigma_i^{t+1}} \sigma_i \, d\Phi(\sigma_i)}{\Phi(\sigma_i^{t+1}) - \Phi(\sigma_i^t)} = 2\sigma_i^t + 2\Delta_i - \frac{\int_{\sigma_i^{t-1}}^{\sigma_i^t} \sigma_i \, d\Phi(\sigma_i)}{\Phi(\sigma_i^t) - \Phi(\sigma_i^{t-1})}. \tag{12}$$

That is, as soon as σ_i^1 is fixed, all other values are determined inductively by this condition. Note that not any σ_i^1 belongs to an equilibrium partition as eventually the indifference condition for σ_i^{T-1} has to yield $\sigma_i^T = 1$. If the following monotonicity condition (M) holds, then there is an essentially unique equilibrium with T partition elements for all T up to some $\bar{T}$.

(M): Partition cutoff types obtained from some σ_i^1 through induction by (12) are increasing in σ_i^1, i.e. $\sigma_i^t(\sigma_i^1) > \sigma_i^t(\sigma_i^{1\prime})$ if and only if $\sigma_i^1 > \sigma_i^{1\prime}$.

To give an example, suppose Φ is the uniform distribution on $[0, 1]$. Then $\mu^t =$

$(\sigma_i^{t-1} + \sigma_i^t)/2$ and (12) becomes

$$\sigma_i^{t+1} = \sigma_i^t + (\sigma_i^t - \sigma_i^{t-1}) + 4\Delta_i$$

which clearly satisfies (M). For $t \geq 2$, this can be solved as

$$\sigma_i^t = t\sigma_i^1 + 4\Delta_i \sum_{j=1}^{t-1} j.$$

A T element partition has to satsify $\sigma_i^T = 1$ or $1 = T\sigma_i^1 + 4\Delta_i \sum_{j=1}^{T-1} j$ which means that $\sigma_i^1(T) = \left[1 - 4\Delta_i \sum_{j=1}^{T-1} j\right]/T$. If $1 - 4\Delta_i \sum_{j=1}^{T-1} j < 0$, then no equilibrium with T partition elements exists. This illustrates that a higher Δ_i leads to a less informative equilibrium in the sense that there are less partition elements. The derivation of the most informative equilibrium for the case $\Delta_i < 0$ is analogous.

Regarding room choice, we do not attempt a full characterization of the equilibrium. However, our result that sufficiently large polarization makes segregation generically optimal follows directly from the derivations above: Let the set of biases be such that no possible room has $\Delta_i = 0$ for some player i unless the room consists only of players sharing the same bias. This is satisfied for generic bias values. Consider a $\mathcal{B}_\eta$ scaling of the biases as in the paper. Note that Δ_i, now denoted as $\Delta_i(\eta) = \eta\Delta_i(1)$, scales linearly in η. For η sufficiently high the upper bound on the number of partition elements $1 + 1/(2|\Delta_i(\eta)|)$ will be below 2 and babbling will be the only equilibrium. This is true in all possible rooms not consisting of only players sharing the same bias. Note that the number of possible rooms is finite due to the finite number of the players and therefore there exists a $\bar{\eta}$ such that babbling is the unique equilibrium in all rooms in which players do not share the same bias for all $\eta \geq \bar{\eta}$. It follows immediately that full segregation is optimal and an equilibrium for $\eta \geq \bar{\eta}$. Similarly, it is straightforward that equilibrium partitions can be arbitrarily fine as $\eta \to 0$. Consequently, full integration is welfare optimal and an equilibrium for η sufficiently low.[13]

8.3. Single state

In this variation, a state of the world $\theta \in \Theta$ is distributed according to distribution F. The state is unobserved but each player i out of n players receives a noisy signal $\sigma_i \in \Sigma$ of the state where σ_i is conditional on θ distributed according to G_θ. The signals are private and – conditional on the state – independent across players. (The latter assumption is relaxed at the end of this subsection.) After observing his signal, a player can access one

[13]To see this it is sufficient to note that (i) full integration is the unique welfare optimal allocation for $\eta = 0$, i.e. in a situation in which all players have the same bias, and (ii) information (in the most informative messaging equilibrium) in any given room allocation approaches full information as $\eta \to 0$. This implies that welfare in a given room allocation approaches welfare under full information in this room allocation as $\eta \to 0$.

of $K \geq 2$ "rooms" and send a message $m_i \in \mathcal{M}$. The message is received by all players in the same room. Afterwards each player takes an action a_i.

The payoff of player i is $u(a, b_i, \theta) = -(a_i - b_i - \theta)^2 - \alpha \sum_{j \neq i}(a_j - b_i - \theta)^2$ where a denotes the vector of actions of all players and $b_i \in \mathcal{B}$ is a commonly known "bias" of player i. That is, player i would like that all players choose the action $b_i + \theta$. The parameter α measures the relative weight players assign to other players' behavior. Players are assumed to maximize expected utility.

The solution concept used is perfect Bayesian Nash equilibrium.

For simplicity, let $\Theta = \{\theta^h, \theta^l\}$ and $\Sigma = \{\sigma^l, \sigma^h\}$ and the signal structure is such that $prob(\sigma^j|\theta^j) = p > 1/2$. We let the message space be binary as well: $\mathcal{M} = \{h, l\}$. Furthermore, we let $\mathcal{B} = \{0, b\}$ and assume that there is at least one player with each of the two biases.

Action choice Denote the belief of player i that the state of the world is high by μ_i (after observing his signal and listening to all the messages in his room). The expected utility of player i can then be written as

$$\begin{aligned} U(a, \mu_i) &= -a_i^2 - \mathbb{E}\left[(b_i + \theta)^2\right] + 2a_i(b_i + \mathbb{E}[\theta]) - \alpha \sum_{j \neq i} \mathbb{E}\left[(a_j - b_i - \theta)^2\right] \qquad (13) \\ &= -a_i^2 - \mu_i(b_i + \theta^h)^2 - (1 - \mu_i)(b_i + \theta^l)^2 + 2a_i(b_i + \mu_i\theta^h + (1 - \mu_i)\theta^l) \\ &\quad -\alpha \sum_{j \neq i} \left[\mu_i(a_j - b_i - \theta^h)^2 + (1 - \mu_i)(a_j - b_i - \theta^l)^2\right] \end{aligned}$$

The optimal action choice of player i is then

$$a_i^* = b_i + \mathbb{E}[\theta] = b_i + \theta^l + \mu_i(\theta^h - \theta^l). \qquad (14)$$

Cheap talk The cheap talk game can – as usual – have several equilibria. There is always a babbling equilibrium where the message is independent of the observed signal and therefore nothing about the state of the world is learned, e.g. $m_i(\sigma_i) = \sigma^h$ for all $\sigma_i \in \Sigma$ and $\mu_i = p$ ($\mu_i = 1 - p$) if $\sigma_i = \sigma^h$ ($\sigma_i = \sigma^l$). We will focus on most informative equilibria, that is equilibria where $m_i(\sigma_i) = \sigma_i$ with as high probability as possible.

Truthful communication is an equilibrium for a given room if all players in this room have the same b_i. To see this, suppose player i could maximize his expected utility (1) not only over a_i but also over the a_j of all the players in his room. Clearly, he would choose the same action for everyone namely $b_i + \theta^l + \mu_i(\theta^h - \theta^l)$. Deviating from the truthful strategy is not profitable because by adhering to truthfulness player i ensures that all other players in the room choose precisely the action he would have chosen for them (while deviating changes the other players' beliefs and therefore their optimal action). Note that this argument depends on all players having the same bias and truthful communication is normally not an equilibrium if players in a given room have different biases. We state this result for future reference in the following lemma.

Lemma 11. *If all players in a given room have the same bias, truthful communication in this room is the most informative equilibrium of the cheap talk game (taking room choice as given).*

We will now analyze the cheap talk problem in rooms in which players with both types of biases are present. In particular, we will be interested in the case of strong differences in opinion, i.e. the case where b is sufficiently large.

Lemma 12. *Let $n_0 \geq 1$ players with bias $b_i = 0$ and $n_b \geq 1$ players with bias $b_i = b$ be in a room. There exists a $\bar{b}$ such that for $b \geq \bar{b}$ babbling is the only equilibrium of the cheap talk game.*

Proof of lemma 12: Suppose that there is a non-babbling equilibrium, i.e. an equilibrium where belief μ_j depends on the messages of players $i \neq j$. Let i be a player affecting j's belief. Without loss of generality, say μ_j is lower if i sends the message l and higher if i sends the message h. By Bayesian updating and independence of the signals, μ_k will then be lower when i sends message l than when he sends message h for all $k \neq i$. (Moreover two players that observe the same signal themselves and are in the same room will have the same belief because of Bayesian updating and independence of signals.) Hence it is without loss of generality to assume that $b_i \neq b_j$. For concreteness, let $b_i = 0$ and $b_j = b$ (the proof for the opposite case is analogous).

Now suppose i observes signal σ^h. We will show that it is optimal for i to send message l if b is sufficiently high. To see this, denote the change in i's expected utility (13) when sending message l instead of message h as ΔU_i [14]

$$
\begin{aligned}
\Delta U_i &= -\alpha \sum_{j \neq i} \mathbb{E}\left[a_j(l)^2 - a_j(h)^2 - 2\theta(a_j(l) - a_j(h))\right] \\
&= -\alpha \sum_{j \neq i} \mathbb{E}\left[\left(\mu_j(l)^2 - \mu_j(h)^2\right)\left(\theta^h - \theta^l\right)^2 + 2(\mu_j(l) - \mu_j(h))(\theta^h - \theta^l)(b_j + \theta^l - \theta)\right] \\
&= \alpha \sum_{j \neq i} \mathbb{E}\left[(\mu_j(h) - \mu_j(l))\left(\theta^h - \theta^l\right) * \left((\mu_j(h) + \mu_j(l))(\theta^h - \theta^l) + 2(b_j + \theta^l - \theta)\right)\right] \\
&= \alpha\left(\theta^h - \theta^l\right) n_b \mathbb{E}\left[(\mu_j(h) - \mu_j(l)) * \left((\mu_j(h) + \mu_j(l))(\theta^h - \theta^l) + 2(b + \theta^l - \theta)\right)\right] \\
&\quad + \alpha\left(\theta^h - \theta^l\right)(n_0 - 1)\mathbb{E}\left[(\mu_j(h) - \mu_j(l)) * \left((\mu_j(h) + \mu_j(l))(\theta^h - \theta^l) + 2(\theta^l - \theta)\right)\right] \\
&= \alpha\left(\theta^h - \theta^l\right)(n_b + n_0 - 1) \\
&\quad \mathbb{E}\left[(\mu_j(h) - \mu_j(l)) * \left((\mu_j(h) + \mu_j(l))(\theta^h - \theta^l) + 2\left(\frac{n_b b}{n_b + n_0 - 1}\theta^l - \theta\right)\right)\right].
\end{aligned}
$$

[14] For a more general proof, one could already go from the first line to $\alpha \sum_{j \neq i} \mathbb{E}\left[(a_j(h) - a_j(l)) * (a_j(h) + a_j(l) - 2\theta)\right]$ and then note that for b high enough even $a_j(l) > \theta^h$.

If $b \geq \theta^h(n_b + n_0 - 1)/(\theta^l n_b)$, the term inside the expectation is positive for any θ and therefore ΔU_i is definitely strictly positive. Hence, i strictly prefers sending message l to message h and i receives signal σ^h. This would imply that i sends message l with probability 1 if the signal is σ^h in this equilibrium. But this contradicts that $\mu_j(l) > \mu_j(h)$. Hence, choosing $\bar{b} = \theta^h N/\theta^l$ where N is the total number of agents implies $\bar{b} \geq \theta^h(n_b + n_0 - 1)/(\theta^l n_b)$ and gives the result. □

Lemma 12 implies that – given a finite number of players – the only way allowing meaningful communication if differences in opinion is high is to have only players with the same bias in a room.

If the differences in opinion are minimal, i.e. b is very low, truthful communication is an equilibrium for any room composition. The reason is the coarseness of the signal structure: Lying in the message game leads – in a truthful equilibrium – to a discrete reaction of all other players in the room. If the difference in bias is very small, this discrete reaction is "too high", i.e. even those players with a (slightly) different bias react more than the deviating player would wish for. The following lemma formalizes this generalization of lemma 11.

Lemma 13. *Let there be $n_0 \leq n$ players with $b_i = 0$ and $n_b \leq n - n_0$ players with $b_i = b$ in a room. There exists a $\underline{b} > 0$ such that for $b \leq \underline{b}$ truthful communication is an equilibrium.*

Proof of lemma 13: For $b = 0$, truthtelling is strictly better than lying (given that all other players tell the truth). Note that i's expected utility is continuous in a_j and a_j^* is continuous in b_j, see (14). Hence, U_i is continuous in b_j. However, μ_j and therefore a_j^* reacts discretely to lying. Consequently, truth-telling is still a best response to truth-telling for $b_j > 0$ sufficiently small. □

From lemma 12 and lemma 13 we know that for b low the most informative equilibrium in a room with a given configuration is truth-telling and for b sufficiently high the "most informative" equilibrium is babbling if players with different biases are present. It seems most likely that $\bar{b} > \underline{b}$. In this case, there are mixed strategy equilibria for $b \in (\underline{b}, \bar{b})$.

Room choice equilibria We claim that separation is an equilibrium if differences in opinion, i.e. the parameter b, are sufficiently high.

Proposition 15. *If $b \geq \bar{b}$, the following strategies constitute an equilibrium:*

1. *Players with bias 0 (b) go to room 0 (1).*

2. *A player sends truthful messages if only players of the same type are in his room and babbles otherwise.*

3. *Actions are taken according to (14) and beliefs μ_i are formed using Bayes' rule (given the equilibrium strategies in 1 and 2).*

This equilibrium is the most informative equilibrium in the sense that no player has more precise information about the state θ in any other equilibrium.

Proof of proposition 15: Given lemma 12, unilateral deviations to other rooms are not profitable: Any such deviation would either lead to being alone in a room or babbling. In either case, the deviating player does not have any information beyond his own signal about the state of the world. This reduces his expected utility directly. Furthermore, deviations lead to less information for other players which again lowers the deviating player's payoff: Less information for players with the same bias as player i implies that their actions are further away from $b_i + \theta$ in expectation. Furthermore, the players with $b_j \neq b_i$ choose actions further away to $b_j + \theta$ if they have less information, i.e. variance of their choice is increased while the expected value stays the same. Given the strictly concave loss function, player i looses from this as well.

Lemmas 11 and 12 imply that no profitable deviation in the cheap talk stage exists. As (14) gives the optimal action (given one's beliefs), no deviation in choosing one's action is profitable either.

By lemma 12, a given player i cannot observe more "non-babbling" messages than in the suggested equilibrium in any other equilibrium. Given that communication is truthful in the suggested equilibrium, player i can therefore not have more precise information about θ in any other equilibrium. $\square$

For $b \leq \underline{b}$, the most informative equilibrium is clearly that every player goes to the same room and truthfully reports his signal.

Welfare optimal room allocation Suppose a social planner could allocate players to rooms. After being assigned a room, players play the same game as above; that is, the planner has no influence on messages or actions. We claim that for $b \geq \bar{b}$ the welfare optimal allocation is to assign everyone with bias 0 in one room and everyone with bias b in another room, i.e. the equilibrium described in proposition 15 is welfare optimal. The idea is the following: For $b \geq \bar{b}$, the cheap talk game in a room where players with both bias types are present will only have a babbling equilibrium by lemma 12. Consequently, any room allocation that assigns players with different biases to the same room will lead to completely uninformative messages and is therefore equivalent to putting every player to a separate room. By assigning players with the same bias to the same room, the planner achieves the most informative equilibrium. That is, truthful communication is possible in each room. The additional information ensures that player with the same bias as player i choose actions closer to $b_i + \theta$. Furthermore, the players with $b_j \neq b_i$ choose actions closer to $b_j + \theta$, i.e. the variance is reduced while the expected value stays the same. Given the strictly concave loss function, player i gains from this as well. Note that the welfare notion can be chosen quite strict in the sense that the described allocation maximizes the welfare of every agent. That is, if agent i could dictatorially decide the room allocation (without having any influence on the messages or actions taken by other players), the same allocation would result.

Similarly, the most informative equilibrium is welfare optimal in the strong sense established above if $b \leq \underline{b}$.

Correlated signals Finally, we want to discuss an extension to this model: People with similar biases might be similar in other respects and therefore have similar information. More precisely, one could imagine that the signals of people with the same bias are positively correlated conditional on the state. The following paragraphs shows that similar results as before hold when signals are correlated.

The main difficulty is to show a result similar to lemma 12 all other results go through without change. The following lemma states that in the limit as b grows large no information can be transmitted in equilibrium. The result is somewhat weaker than lemma 12 but similar in nature.

Lemma 14. *Let $n_0 \geq 1$ players with bias $b_i = 0$ and $n_b \geq 1$ players with bias $b_i = b$ be in a room. Let the signal technology be such that signals are not perfectly correlated and such that all signal vectors have strictly positive probability. For every $\varepsilon > 0$, there exists a b_ε such that $\mathbb{E}_{m_{-i},\sigma_j}\left[\mu_j(m_i = \sigma^h) - \mu_j(m_i = \sigma^l)|\sigma_i\right] < \varepsilon$ in every equilibrium.*

Proof of lemma 14: Suppose that there is a non-babbling equilibrium, i.e. an equilibrium where belief μ_j depends on the messages of players $i \neq j$. Let i be a player affecting j's belief. Without loss of generality, say μ_j is lower if i sends the message l and higher if i sends the message h. By Bayesian updating, μ_k will then be lower when i sends message l than when he sends message h for all $k \neq i$. First, let $b_i \neq b_j$. For concreteness, let $b_i = 0$ and $b_j = b$ (the proof for the opposite case is analogous).

Now suppose i observes signal σ^h. To make an informative equilibrium possible, the change in i's expected utility (1) when sending message l instead of message h, ΔU_i, must not be strictly positive:

$$
\begin{aligned}
\Delta U_i &= -\alpha \sum_{j \neq i} \mathbb{E}\left[a_j(l)^2 - a_j(h)^2 - 2\theta(a_j(l) - a_j(h))\right] \\
&= -\alpha \sum_{j \neq i} \mathbb{E}\left[\left(\mu_j(l)^2 - \mu_j(h)^2\right)\left(\theta^h - \theta^l\right)^2 + 2(\mu_j(l) - \mu_j(h))(\theta^h - \theta^l)(b_j + \theta^l - \theta)\right] \\
&= \alpha \sum_{j \neq i} \mathbb{E}\left[(\mu_j(h) - \mu_j(l))\left(\theta^h - \theta^l\right) * \left((\mu_j(h) + \mu_j(l))(\theta^h - \theta^l) + 2(b_j + \theta^l - \theta)\right)\right] \\
&> \alpha \sum_{j \neq i} \mathbb{E}\left[(\mu_j(h) - \mu_j(l))\left(\theta^h - \theta^l\right) * \left((\mu_j(h) + \mu_j(l))(\theta^h - \theta^l) + 2(b_j + \theta^l - \theta^h)\right)\right] \\
&= \alpha\left(\theta^h - \theta^l\right)\Big(n_b b \mathbb{E}\left[\mu_j(h) - \mu_j(l) | b_j = b, \sigma_i = \sigma^h\right] \\
&\quad + \sum_{j \neq i} \mathbb{E}\left[(\mu_j(h) - \mu_j(l))(\mu_j(h) + \mu_j(l) - 2)(\theta^h - \theta^l)\right]\Big) \\
&> \alpha\left(\theta^h - \theta^l\right)\left(n_b b \mathbb{E}\left[\mu_j(h) - \mu_j(l) | b_j = b, \sigma_i = \sigma^h\right] - 2N(\theta^h - \theta^l)\right)
\end{aligned}
$$

where $N = n_b + n_0$. Clearly, the last expression is greater than zero if $b\mathbb{E}\left[\mu_j(h) - \mu_j(l) | b_j = b, \sigma_i = \sigma^h\right] > 2N(\theta^h - \theta^l)/n_b$. Hence, $b_\varepsilon = 2N(\theta^h - \theta^l)/(n_b \varepsilon)$ gives the result in the lemma.

Second, let $b_j = b_i$. Note that the result above says that i's message contains no information in the limit as $b \to \infty$. It follows that given that the signal technology is (i) not perfectly correlated and (ii) puts strictly positive probability on all signal vectors, the result has to hold also for j with $b_j = b_i$.[15] □

9. Follower model

This section replicates our results for a slightly different model in which instead of choosing "rooms" in the first stage, players choose which other players to "follow". Players are unrestricted regarding the size and composition of the set of players they follow. In the second stage every player sends one cheap talk message to his "followers". We will adopt the convention that each player follows himself (which is immaterial for the results but allows us to proceed with some of the derivations analogously to the paper.) Signal technology and preferences are the same as in the paper.

Many of our main results carry over to this extension in a modified way. Similarly to theorem 1 of the paper, a player will now tell the truth if and only if his own bias is in a symmetric interval around the average bias of his followers. This means that player i always wants to follow player j unless the very act of following makes j babble. This feature of the best response implies that the notions of most informative equilibrium and welfare optimal follower-assignment coincide. We can again show that if polarization increases, segregation becomes more desirable and it becomes optimal for players to segregate more.

[15] The two assumptions avoid that lying leads to a zero probability event where beliefs cannot be determined by Bayes' rule.

If polarization is low, it is efficient and an equilibrium for everyone to follow everyone – similar to the fully integrated room in our main model.

This extension has the interesting feature that there are differences between players with moderate and extreme preferences in how isolated they are from others. Players with moderate preferences can in equilibrium be followed by much of the population but still tell the truth, because different players' influences on $\bar{b}$ at least partially neutralize each other. Extremist players, however, can only be followed by other extremists of the same persuasion, as they would babble if followed by too many moderates or even by extremists at the other end of the spectrum.

Let us now derive these results more formally. Clearly, the optimal action is still

$$a_i^* = b_i + \sum_{j=1}^{n} \mathbb{E}[\theta_j].$$

More importantly, lemma 1 still applies and we can concentrate on pure strategy equilibria.

Lemma 15. *Let $(m_1, \ldots, m_n)$ be equilibrium strategies. If m_i is a mixed strategy, then there also exists an equilibrium with strategies (m_i^t, m_{-i}), where m_i^t is the truthful strategy.*

Proof. Denoting i's followers by F_i, the set of player i is following by f_i and fixing some equilibrium $(m_1, \ldots, m_n)$, player i's expected payoff when sending message m_i to F_i can be written as

$$U_i(m_i|\sigma_i) = \mathbb{E}\left[-\left(a_i(m_{-i,R_i}, \sigma_i) - b_i - \sum_{k=1}^{n} \theta_k \right)^2 - \alpha \sum_{j \notin F_i} \left\{ \left(a_j(m_{-i,f_j}, \sigma_j) - b_i - \sum_{k=1}^{n} \theta_k \right)^2 \right\} \right.$$
$$\left. -\alpha \sum_{j \in F_i, j \neq i} \left\{ \left(a_j(m_i, m_{-i,f_j}, \sigma_j) - b_i - \sum_{k=1}^{n} \theta_k \right)^2 \right\} \middle| \sigma_i \right].$$

which can be split in a part that is independent of i's message m_i and a part that depends on m_i:

$$U_i(m_i) = \mathbb{E}\left[const - \alpha \sum_{j \in F_i, j \neq i} \left(a_j(m_i, m_{-i,f_j}, \sigma_j) - b_i - \sum_{k=1}^{n} \theta_k \right)^2 \middle| \sigma_i \right].$$

Specifically, sending message m^h gives expected payoff

$$U_i(m^h) = \mathbb{E}\left[const - \alpha \sum_{j \in F_i, j \neq i} \left(b_j - b_i + \mu_{ji}^h + \sum_{k \neq i} \mu_{jk} - \theta_i - \sum_{k \neq i} \theta_k \right)^2 \middle| \sigma_i \right]$$

where $\mu_{ji}^h = \mathbb{E}[\theta_i | m_i = m^h]$, i.e. μ_{ji}^h is the belief of a player j (following i) concerning θ_i if player i sends message m^h. Note that this belief is the same for all players $j \neq i$ following i. Sending message m^l gives

$$U_i(m^l) = \mathbb{E}\left[const - \alpha \sum_{j \in F_i, j \neq i} \left(b_j - b_i + \mu^l_{ji} + \sum_{k \neq i} \mu_{jk} - \theta_i - \sum_{k \neq i} \theta_k \right)^2 \middle| \sigma_i \right]$$

where $\mu^l_{ji} = \mathbb{E}[\theta_i | m_i = m^l]$. The difference in expected payoff is then

$$\begin{aligned}
\Delta U_i(\sigma_i) &= (U_i(m^h) - U_i(m^l))/\alpha \\
&= -\sum_{j \in F_i, j \neq i} \mathbb{E}\left[{\mu^h_{ji}}^2 - {\mu^l_{ji}}^2 + 2(\mu^h_{ji} - \mu^l_{ji}) \left(b_j - b_i + \sum_{k \neq i} \mu_{jk} - \theta_i - \sum_{k \neq i} \theta_k \right) \middle| \sigma_i \right] \\
&= -2(\mu^h_{ji} - \mu^l_{ji}) \sum_{j \in F_i, j \neq i} \left[\frac{\mu^h_{ji} + \mu^l_{ji}}{2} + b_j - b_i - \mathbb{E}\left[\theta_i | \sigma_i \right] \right] \\
&= 2(\mu^h_{ji} - \mu^l_{ji})(n_{F_i} - 1) \left[-\frac{\mu^h_{ji} + \mu^l_{ji}}{2} - \frac{\sum_{j \in F_i, j \neq i} b_j}{n_{F_i} - 1} + b_i + \mathbb{E}\left[\theta_i | \sigma_i \right] \right] \qquad (15)
\end{aligned}$$

where n_{F_i} denotes the number of elements in F_i. (For the transformation to line 3, we make use of the fact that $\mu^{\cdot}_{ji}$ is the same for all $j \in F_i \setminus \{i\}$.)

Player i is only willing to choose a mixed strategy after receiving signal σ_i if $\Delta U_i(\sigma_i) = 0$. From expression (15) it is clear that this can only be true for at most one signal as $\mathbb{E}[\theta_i | \sigma_i]$ varies in σ_i. Furthermore, $U_i(\sigma^h) = 0$ implies $U_i(\sigma^l) < 0$ and similarly $U_i(\sigma^l) = 0$ implies $U_i(\sigma^h) > 0$.

Now suppose i's equilibrium strategy m_i is mixed after signal σ^h. Then, $\Delta U_i(\sigma^h) = 0$ implies $\Delta U_i(\sigma^l) = 2(\mu^h_{ji} - \mu^l_{ji})(n_{F_i} - 1)(1 - 2p) < 0$ and therefore $m_i(\sigma^l) = m^l$ which implies $\mu^h_{ji} = p$ as a m^h is only sent by i after receiving signal σ^h. This implies $(\mu^h_{ji} + \mu^l_{ji})/2 \geq 1/2$ as $\mu^l_{ji} \geq 1 - p$. Now consider the equilibrium candidate (m^t_i, m_{-i}). With the truthful strategy m^t_i, $\mu^{th}_{ji} = p$ and $\mu^{tl}_{ji} = 1 - p$ and therefore $(\mu^{th}_{ji} + \mu^{tl}_{ji})/2 = 1/2$. This implies that $\Delta U_i(\sigma^h) > 0$ in the equilibrium candidate (m^t_i, m_{-i}), i.e. truthful reporting is optimal for i after receiving signal σ^h. In the equilibrium candidate (m^t_i, m_{-i}), truthful messaging is still optimal after signal σ^l as well: From $p > 1/2$, $\mu^h_{ji} \leq p$ and $\mu^l_{ji} \leq 1/2$ it follows that $-1/2 + (1 - p) < -\left(\mu^h_{ji} + \mu^l_{ji}\right)/2 + p$. As in the original equilibrium (m_i, m_{-i}) we had $\Delta U_i(\sigma^h) = 0$ and therefore $-\left(\mu^h_{ji} + \mu^l_{ji}\right)/2 + p = \sum_{j \in R_i, j \neq i} b_j/(n_{F_i} - 1) + b_i$, we get that $-1/2 + 1 - p < \sum_{j \in F_i, j \neq i} b_j/(n_{F_i} - 1) + b_i$ and therefore $U_i(\sigma^l) < 0$ in the truthful equilibrium candidate (m^t_i, m_{-i}). Hence, truthful messaging is i's best response in the equilibrium candidate (m^t_i, m_{-i}). Finally, note that the $\Delta U_j(\sigma_j)$ for $j \neq i$ is not affected by changing i's strategy from m_i to m^t_i. Hence, (m^t_i, m_{-i}) is an equilibrium.

The argument in case i's strategy is mixed after signal σ^l is analogous. □

The previous lemma (and its proof) allow a characterization of the equilibrium messaging strategy in the most informative equilibrium and an analogue to theorem 1.

Theorem 6. *Let $\overline{b} = \frac{\sum_{k \in F_i} b_k}{n_{F_i}}$ be the mean bias of i's followers. In the most informative equilibrium in this room, a player i tells the truth to his followers if*

$$b_i \in \left[\overline{b} - \frac{n_{F_i} - 1}{n_{F_i}}(p - \frac{1}{2}),\ \overline{b} + \frac{n_{F_i} - 1}{n_{F_i}}(p - \frac{1}{2}) \right]$$

and babbles otherwise.

Proof. As in theorem 1 in the paper. □

We proceed by turning to stage 1. Take some follower allocation as fixed, then the expected payoff of player i following players in f_i while having followers F_i equals

$$\begin{aligned} U_i \quad = \quad & -\mathbb{E}\left[\left(\sum_{j \in f_i^{truth} \cup \{i\}} (\mu_{ij} - \theta_j) + \sum_{j \notin f_i^{truth} \cup \{i\}} (\frac{1}{2} - \theta_j) \right)^2 \right. \\ & \left. + \alpha \sum_{j \neq i} \left(b_j - b_i + \sum_{k \in f_j^{truth} \cup \{j\}} (\mu_{jk} - \theta_k) + \sum_{k \notin f_j^{truth} \cup \{j\}} (\frac{1}{2} - \theta_k) \right)^2 \right] \end{aligned}$$

where f_i^{truth} are the players in f_i that send truthful/informative messages in equilibrium and $f_i \setminus f_i^{truth}$ are those players in f_i that are babbling.

For any $i \neq j$, the two values of θ_i and θ_j are independent; the same is true for μ_{ij} and μ_{ik}. Hence $\mathbb{E}\left[\mu_{ij} - \theta_j\right] = 0$ and $\mathbb{E}\left[(\mu_{ij} - \theta_j)(\mu_{ik} - \theta_k)\right] = 0$, which means that the above expression can be rewritten as

$$\begin{aligned} U_i \quad = \quad & - \sum_{j \in f_i^{truth} \cup \{i\}} \mathbb{E}\left[(\mu_{ij} - \theta_j)^2\right] - \sum_{j \notin f_i^{truth} \cup \{i\}} \mathbb{E}\left[(\frac{1}{2} - \theta_j)^2 \right] \\ & - \alpha \sum_{j \neq i} (b_j - b_i)^2 - \alpha \sum_{j \neq i} \sum_{k \in f_j^{truth} \cup \{j\}} \mathbb{E}\left[(\mu_{jk} - \theta_k)^2\right] - \alpha \sum_{j \neq i} \sum_{k \notin f_j^{truth} \cup \{j\}} \mathbb{E}\left[(\frac{1}{2} - \theta_k)^2 \right]. \end{aligned}$$

Now note that $\mathbb{E}\left[(\mu_{jk} - \theta_k)^2\right]$ can have two possible values in the most informative equilibrium: If $k \in f_j^{truth} \cup \{j\}$, i.e. if j has received information about θ_k, then $\mathbb{E}\left[(\mu_{jk} - \theta_k)^2\right] = p(1-p)$. If j has not received information about θ_k, then $\mathbb{E}\left[(\mu_{jk} - \theta_k)^2\right] = \frac{1}{4}$. (We can check that information always reduces variance and increases welfare since $p > \frac{1}{2}$ and hence $p(1-p) < \frac{1}{4}$.) This allows to denote utility in the notation of the paper using pieces of information

$$U_i = -\alpha \sum_{j \neq i} \{(b_j - b_i)^2\} - 1/4\left[n + \alpha(n-1)n\right] + (1/4 - p(1-p)) \left[\zeta_i + \alpha \sum_{j \neq i} \zeta_j \right]$$

and express welfare as

$$W = \sum_i U_i = \sum_i \left[-\alpha \sum_{j\neq i} \{(b_j - b_i)^2\} - 1/4\left[n + \alpha(n-1)n\right] + (1/4 - p(1-p))\left[\zeta_i + \alpha \sum_{j\neq i} \zeta_j\right]\right]$$
$$= -\alpha \sum_{i=1}^{n} \sum_{j\neq i} \{(b_j - b_i)^2\} - \frac{1}{4} n^2 \left[1 + \alpha(n-1)\right] + (p - \frac{1}{2})^2 (1 + \alpha(n-1)) \sum_i \zeta_i.$$

In this expression, all terms are model parameters except for the sum over all ζ_i, which shows that welfare is linearly increasing in $\sum_i \zeta_i$.

We are now ready to analyze equilibrium follow decision. The following describes player j's best response: Hold arbitrary stage 1 decisions of players other than j fixed. From the expression for U_i in terms of pieces of information, it is clear that it is uniquely optimal for j to follow i if i tells the truth given the stage 1 decisions of the other players and j's decision to follow. Furthermore, j is indifferent between following i and not following i if i babbles regardless of j's choice. Last but not least, j optimally does not follow i if following leads to babbling by i while not following allows informative messages by i. This leads to the following result which is in line with the empirically found homophily.

Proposition 16. *It is weakly dominant for j to follow i if $|b_j - b_i| \leq (p - 1/2)/2$.*

Proof. From the reasoning of the previous paragraph, it is sufficient to show that j following i will not cause i to babble (given some arbitrary first stage choices of the other players) if $|b_j - b_i| \leq (p - 1/2)/2$. If j follows i, then $n_{F_i} \geq 2$ (recall that by convention $i \in F_i$) which implies $(n_{F_i} - 1)/n_{F_i} \geq 1/2$. Consequently, i will tell the truth if j is the only player following i. Now consider the case where some players (other than j) are following i. If i is truthtelling without j following him, then, by theorem 6, $b_i \geq \bar{b} - (p-1/2)(n_{F_i} - 1)/n_{F_i}$ where $\bar{b}$ is the average bias of players other than j following i. By the condition of the proposition $b_i \geq b_j + (p-1/2)/2$ and bringing the previous two inequalities together yields

$$b_i \geq \frac{(n_{F_i} - 1)\bar{b} + b_j}{n_{F_i}} - (p - 1/2)\frac{n_{F_i}^2 - 3n_{F_i}/2 + 1}{n_{F_i}^2}$$

which implies

$$b_i \geq \frac{(n_{F_i} - 1)\bar{b} + b_j}{n_{F_i}} - (p - 1/2)\frac{n_{F_i}}{n_{F_i} + 1}.$$

Similarly, $b_i \leq \bar{b} + (p - 1/2)(n_{F_i} - 1)/n_{F_i}$ and $b_i \leq b_j + (p - 1/2)/2$ imply

$$b_i \leq \frac{(n_{F_i} - 1)\bar{b} + b_j}{n_{F_i}} + (p - 1/2)\frac{n_{F_i}^2 - 3n_{F_i}/2 + 1}{n_{F_i}^2}.$$

Consequently, i will be truthtelling when j follows him if i is truthtelling when j does not follow him. $\square$

The simple characterization of equilibria above makes it straightforward to characterize the structure of the most informative and therefore welfare maximizing equilibrium in stage 1. The welfare optimal set of followers for i can be determined independently of the welfare optimal set of followers of other players. In fact, it is given by simple maximization problem:

Proposition 17. *The welfare optimal set of i's followers, F_i^*, is given by the maximization problem maximizing the number of elements of F_i subject to the truthtelling constraint*

$$b_i \in \left[\frac{\sum_{j\in F_i} b_j}{n_{F_i}} - \frac{n_{F_i}-1}{n_{F_i}}(p-1/2), \frac{\sum_{j\in F_i} b_j}{n_{F_i}} + \frac{n_{F_i}-1}{n_{F_i}}(p-1/2)\right] \tag{16}$$

where $n_{F_i} = \sum_{j\in F_i} \mathbb{1}_{j\in F_i}$. The welfare optimal follower allocation $(F_1^*, \ldots, F_n^*)$ is an equilibrium.

Proof. Welfare is increasing in the pieces of information provided in equilibrium. The maximal number of pieces of information provided by i is given by the results of the maximization problems in the proposition. As there are no constraints on how many players to follow, $(F_1^*, \ldots, F_n^*)$ is feasible. It is also an equilibrium: No player i wants to follow an additional player j as – by the definition of $(F_1^*, \ldots, F_n^*)$ – this would lead to babbling by j. As each player is only following players that are truthtelling, no player $i \in F_j^*$ benefits from not following j. $\square$

Note two implications of the previous proposition. First, a pure strategy equilibrium exists. Second, the welfare optimal follower allocation always coincides with the follower allocation in the welfare optimal equilibrium.

Let $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$ be a bias configuration. (Note that this is not a set, as several people can have the same bias.) Assume that $\mathcal{B}$ is generic in the sense that no bias is the average of any set of other biases (except in cases where several people have the same bias). Now we can consider an alternative bias configuration $\mathcal{B}_\eta$, with $\eta \in (0, \infty)$, which for every b_i in $\mathcal{B}$ contains ηb_i. Then the following is true:

Theorem 7. *(i) If η is sufficiently close to 0, full integration, i.e. $F_i = \{1, \ldots, n\}$ for all $i = 1, \ldots, n$, is welfare-optimal for bias configuration $\mathcal{B}_\eta$.*

(ii) If η is sufficiently large, full segregation by bias types is generically welfare-optimal for bias configuration $\mathcal{B}_\eta$.

Proof. Note that the truthtelling constraint (16) for set of biases $\mathcal{B}_\eta$ can be written as

$$b_i \in \left[\frac{\sum_{j\in F_i} b_j}{n_{F_i}} - \frac{1}{\eta}\frac{n_{F_i}-1}{n_{F_i}}(p-1/2), \frac{\sum_{j\in F_i} b_j}{n_{F_i}} + \frac{1}{\eta}\frac{n_{F_i}-1}{n_{F_i}}(p-1/2)\right].$$

For $\eta \to 0$, this constraint is arbitrary slack while for $\eta \to \infty$ it is arbitrarily strict. The latter implies that for $\mathcal{B}_\eta$ such that no element is a convex combination of other elements (not all of which equal to the initial element) no F_i apart from full segregation can satisfy the constraint. □

Example 1. *As a straightorward example consider the binary case where $b_i \in \{0, b\}$ for all players. Let n_b (n_0) be the number of players with $b_i = b$ ($b_i = 0$). The welfare optimal follower allocation is then as follows: F_i consists of all players j with $b_j = b_i$ and k players with $b_j \neq b_i$ where k is the highest integer such that i's truthtelling constraint still holds. This implies the following: A majority player has more followers than a minority player. A majority player has (weakly) more followers of a different bias type than a minority player. As b grows larger, players have less and less followers of the other bias type. For b above some critical $\underline{b}$ each minority player is only followed by the other members of the minority. For b above some critical $\tilde{b} \geq \underline{b}$ each majority player is only followed by the other members of the majority.*

Moving away from the welfare optimal equilibrium note that other equilibria exist in stage 1. In particular there are equilibria in which players babble. From the best response structure we immediately get the following result.

Lemma 16. *If player i babbles given followers F_i, then i would still babble if his set of followers was $F_i \setminus \{j\}$ for $j \in F_i$.*

Proof. Suppose there existed a $j \in F_i$ such that i would not babble with set of followers $F_i \setminus \{j\}$. In this case, j has a profitable deviation: Not following i will increase the number of pieces of information of some other players while it will not reduce the number of pieces he has himself. By $\alpha > 0$ the deviation is profitable. □

The lemma indicates that in equilibrium there can be players that babble because they are followed by too many other players. These player are however so much over-subscribed by players with very different biases that they could still not tell the truth if an arbitrary single player decided not to follow them anymore. That is, they are, so to speak, far away from being tempted to tell what they know.

Some interesting comparison between extremists and centrists can be made based on the best response structure of the cheap talk stage. Consider for instance "extremists", i.e. players with an unusual high or low bias. These players can send truthful messages if they are followed by similarly extreme players. These will typically be only a few people given that only a minority can have "extreme", i.e. unusually high or low, biases. Now consider a centrist, i.e. someone whose bias is close to the average of the population. He can be followed by (nearly) everyone and he can still be truthtelling. In the extreme case where his bias equals the average bis in the population, indeed everyone will follow him in the welfare optimal equilibrium and he will be truthtelling. Compare this to the extremist: If (sufficiently) many people follow an extremist, he will be babbling. The following proposition uses the same intuition to show that in the welfare optimal equilibrium centrists have more followers than extremists if the distribution of biases is single peaked and symmetric.

Proposition 18. *Let biases be distributed on an equally spaced finite grid and let the distribution of biases be single-peaked and symmetric around the mean. Then the number of followers in the welfare maximal equilibrium is lower, the farther a player's bias is away from the mean bias.*

Proof of proposition 18: Denote the mean bias in the population by μ_b and order – without loss of generality – players according to their biases, i.e. $b_1 \leq b_2 \leq \cdots \leq b_n$. By proposition 17, F_i^* is given by $\max n_{F_i}$ subject to $|b_i - \sum_{j \in F_i,\, j \neq i} b_j/(n_{F_i} - 1)| \leq (p - 1/2)$. Given the assumptions in the proposition, the solution to this maximization problem is straightforward: If $b_i > \mu_b$, then F_i is the set of players $\{\underline{i}, \underline{i}+1, \ldots, n\}$ where $\underline{i} \leq i$ is determined such that $b_i - \sum_{j \geq \underline{i}, j \neq i} b_j/(n_{F_i} - 1) \leq (p - 1/2)$ and $b_i - \sum_{j \geq \underline{i}-1, j \neq i} b_j/(n_{F_i} - 1)| > (p - 1/2)$. Similarly, if $b_i < \mu_b$, then F_i is the set of players $\{1, 2, \ldots, \bar{i}\}$ where $\bar{i} \geq i$ is determined such that $-b_i + \sum_{j \leq \bar{i}, j \neq i} b_j/(n_{F_i} - 1) \leq (p - 1/2)$ and $-b_i + \sum_{j \leq \bar{i}+1, j \neq i} b_j/(n_{F_i} - 1)| > (p - 1/2)$. For $b_i = \mu_b$, clearly $F_i^* = \{1, \ldots, n\}$. From the definitions of $\underline{i}$, $\bar{i}$ and the single peakedness of the bias distribution, the result follows directly. □

References

Crawford, V. P. and J. Sobel (1982). Strategic information transmission. *Econometrica 50*(6), 1431–1451.