When Are Echo Chambers Useful?*

Ole Jann — CERGE-EI

Christoph Schottmüller — University of Cologne and TILEC

January 27, 2026

Abstract

We model a society with dispersed information and polarized preferences in which individuals choose with whom to communicate. Segregation into small, homogeneous groups can maximize truthful communication, and can thus be individually rational and Pareto-efficient. We characterize the welfare-optimal communication structure, provide conditions under which it arises in equilibrium, and identify an informational externality that can lead to inefficiently little segregation in equilibrium. Our framework can accommodate uncertainty, public information, and different network structures.

JEL: D72 (Political Processes), D82 (Asymmetric Information), D83 (Learning, Communication), D85 (Network Formation and Analysis)

Keywords: asymmetric information, echo chambers, polarization, debate, cheap talk, information aggregation

*Jann: CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Prague; `ole.jann@cerge-ei.cz`. Schottmüller: Department of Economics, University of Cologne; `c.schottmueller@uni-koeln.de`. We are grateful for helpful comments by Rachel Bernhard, James Best, Benjamin Blumenthal, Ben Brooks, Vince Crawford, Marcelo Fernandez, Ben Golub, Sanjeev Goyal, Bård Harstad, Paul Klemperer, Vasily Korovkin, Nenad Kos, Meg Meyer, Kirill Pogorelskiy, David Ronayne, Larry Samuelson, Bill Sandholm, Karl Schlag, Jakub Steiner, Peter Norman Sørensen, Kyle Woodward and Peyton Young, as well as audiences at Bar-Ilan, Bielefeld, Bonn, CERGE-EI, Cologne, Copenhagen, Groningen, Konstanz, Munich, Oxford, Royal Holloway, Sussex, Vienna, Warwick and Wisconsin-Madison and at various conferences and workshops. This research was supported by Charles University Primus project 20/HUM/019 and GAČR grant 22-33162M.

Large parts of our professional and private lives, online and offline, are organized around the non-market exchange of information. But while people constantly seek out others' views and knowledge, they do not seek out a wide range of them. Instead, they tend to segregate into homogeneous communities and limit their exposure to different viewpoints or diverse backgrounds.[1]

This poses a puzzle: If people constantly seek and exchange information, why do they voluntarily restrict the diversity and amount of information they receive? Can this only be explained by behavioral biases and conflict avoidance, or are there informational reasons? More practically: Is segregation into "echo chambers" necessarily socially harmful – and should policy seek to influence who communicates with whom?[2]

In this paper, we study a society with two problems: dispersed information and polarized preferences. The former means that no single individual knows enough to make successful choices; the latter means that even perfectly informed individuals would disagree on optimal actions.[3] We develop a general framework in which people communicate strategically within groups and choose groups, while anticipating how communication will function within each group.

We show that segregation into small, homogeneous groups can arise from rational choice and can *maximize* information transmission. Such segregation may be welfare-maximizing and even Pareto-optimal when the degree of preference polarization is high.

The key mechanism behind this result is the interaction between information dispersion and preference polarization. Their effects on communication are opposite: Dispersion creates incentives to learn from others, while polarization creates incentives to misrepresent information to influence others' actions. As polarization within a group increases, truthful communication breaks down even though incentives to learn remain. Segregation into more homogeneous groups reduces these incentives to mislead: It limits the amount of *potential* communication but enables more *actual* communication.

The paper makes three main contributions. First, we develop a tractable framework of many-to-many communication among agents with heterogeneous information and preferences (sections 1 to 3). The framework admits a simple graphical characterization of truthful communication in equilibrium, and welfare can be evaluated by counting pieces of information that are exchanged. We also show how the framework can be extended to settings with uncertainty and endogenous network formation (section 5).

Second, we use this framework to characterize the first-best communication structure

[1]See evidence on segregation in blogs (Lawrence et al., 2010), Facebook (Del Vicario et al., 2016; Quattrociocchi et al., 2016), Twitter (Barberá et al., 2015) and online and offline contexts more generally (Gentzkow and Shapiro, 2011).

[2]Consider claims that echo chambers are "dangerous" (Grimes, 2017), have "Balkanised society" (Itten, 2018) and contribute to populist movements such as Brexit (Chater, 2016) and the election of Donald Trump (Hooton, 2016).

[3]Throughout, we use "polarization" to refer exclusively to exogenous preference dispersion, similar to the distribution measure of Esteban and Ray (1994), and not to disagreement in beliefs or information.

and show that segregation into “echo chambers” can increase the amount of information actually exchanged when preferences are sufficiently polarized (section 4).

Third, we characterize sufficient conditions under which efficient group allocations arise in an equilibrium of the group-choice game (section 4). This is the case when individuals care sufficiently about the decisions of others, or when preference polarization is either very large or very small relative to information dispersion. We show that people may segregate *too little* in equilibrium compared to the social optimum, because group choice generates an informational externality: individuals choose groups to maximize their own learning without fully internalizing the informational benefits that their choices have on others. In societies with bipolar polarization, insufficient segregation is the only way in which the best equilibrium can deviate from the welfare-optimum.

Our model is deliberately stylized and omits many features commonly associated with echo chambers. Nonetheless, we argue that it captures a “smallest common denominator” of their core mechanisms, namely systematic sorting into communication structures. We discuss this interpretation in section 6.

We do not claim that echo chambers are unambiguously beneficial, and our results call for a nuanced view: Whether segregation facilitates honest debate or narrows perspectives depends on the underlying structure of polarization and mutual mistrust. Overall, ideological segregation is better understood as a symptom rather than a root cause, as polarization and mistrust do the primary damage while segregation can be a rational response that mitigates it.

Our analysis also sheds light on the role of social networking platforms, which provide infrastructure for users to sort and segregate rather than merely to connect (see section 6).

The paper proceeds as follows. We begin with a non-technical overview and review the literature. Sections 1–3 develop the model and analyze communication and group formation. Section 4 studies polarization and segregation, section 5 summarizes extensions, section 6 discusses applications, and section 7 concludes.

A non-technical overview of our model and results We study a model in which individuals face uncertainty about the state of the world and differ in their preferences over actions. The state of the world, preferences and actions are all real numbers, and each individual has an ideal action: she prefers all actions – her own and others’ – to be as close as possible to the sum of the state of the world and her preference. Individuals may therefore disagree both because they have different information and because they have different preferences. It is useful to think of individuals as being characterized by their preference parameter b_i; panel (i) of figure 1 shows an example with six individuals and their positions on the preference line.

Each individual receives private information about the state of the world, while preferences are common knowledge. Before choosing actions, individuals sort into separate "rooms". They then engage in "cheap talk"-style communication within each room, where each message can be heard by anyone else within the room, but not by people in other rooms. If a person finds herself in a room with very diverse preferences, she thus faces a trade-off: On the one hand, she wants to correctly inform those with preferences close to her own, as it will bring their actions closer to her own ideal point. On the other hand, she wants to mislead those with very different preferences, as this will move their (far-away) actions closer to her ideal point. If most of her audience has a much lower preference parameter than her, for example, she wants to make them believe that the state of the world is high, to move their action closer to her ideal point.

Each person's signal is informative about the state of the world, but not about the signals of others – intuitively, different people observe different aspects of the world. This simplifying assumption allows us to derive a set of communication equilibria within each room that is well-ordered by informativeness, and in which the most informative equilibrium is given by a simple graphical solution. In particular, the welfare of an equilibrium is linear in the sum of the pieces of information that all players have (where each player's signal constitutes one "piece of information").

This is illustrated in panel (ii) of figure 1. In the most informative communication equilibrium, those indviduals with preferences close enough to the average preference in the room will truthfully reveal their information, and all others reveal no information. If all individuals are in one room (as shown in the top panel), there are therefore $3+3+2+2+3+3 = 16$ pieces of information in this example.

Panel (iii) of figure 1 illustrates our main result about efficient segregation: If the individuals are segregated into two rooms instead of being in one room as in the top panel, individuals 3 and 4 are no longer able to exchange any information with each other – but every individual now reveals their information to everyone within their own room. In total, the individuals have $3+3+3+3+3+3 = 18$ pieces of information, meaning that welfare has increased (and in fact all individuals are weakly better off).

Panel (iv) of figure 1 illustrates one of our results on whether efficient segregation can arise in equilibrium. Individual 4 has a profitable deviation from the efficient segregation shown in panel (iii), namely joining the room in which individuals 1 to 3 are communicating. This increases the information available to individual 4, but deprives individuals 5 and 6 of 4's information – and thus lowers welfare, since the total number of pieces of information is now 17.

In Section 4.2 we study polarization as clustering around preference types. We show that high polarization makes full segregation both welfare-optimal and an equilibrium, while low polarization makes full integration optimal and an equilibrium (Theorem 3).

Our analysis disentangles the welfare effects of polarization and segregation. While greater polarization both increases segregation and lowers welfare, the welfare loss is

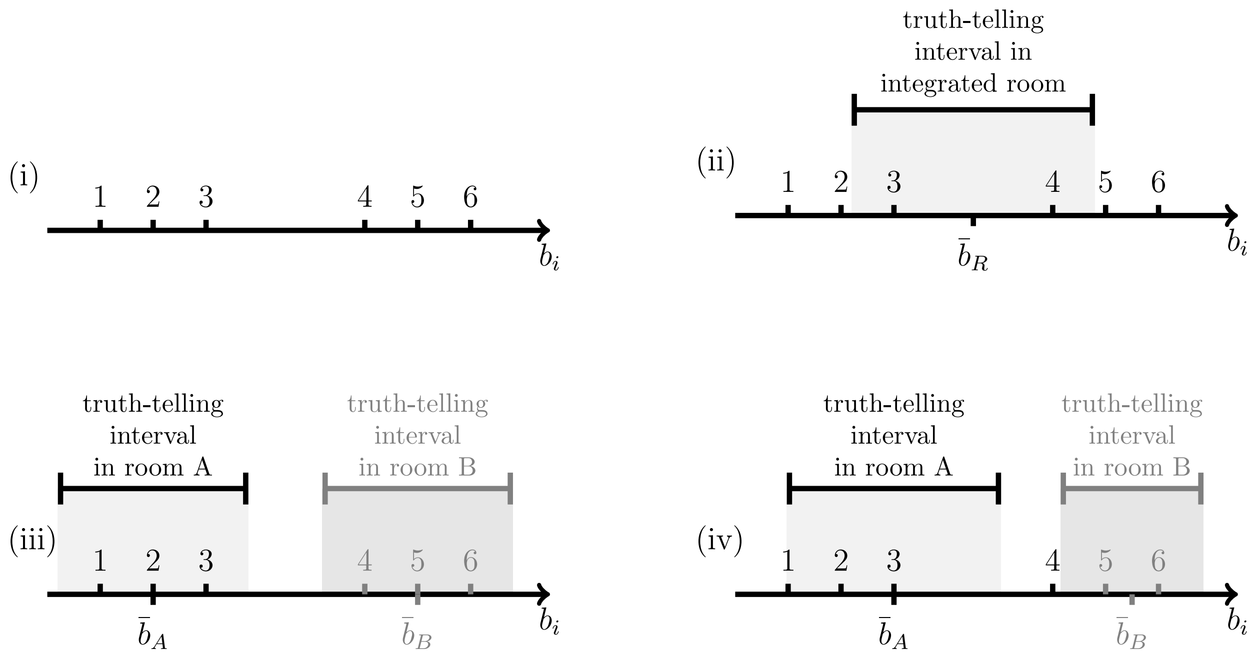

Figure 1: An illustration of our main results. (i): Six players characterized by their preferences b_i. (ii): In the fully integrated room, only players 3 and 4 reveal their information. (iii): In two segregated rooms, all players reveal their information. (iv) Player 4 has a profitable deviation from the optimal segregation.

caused by polarization, not by segregation. On the contrary, segregation mitigates the harmful effects of polarization: preventing segregation in a polarized society would further reduce welfare. Segregation into echo chambers can therefore be understood as a decentralized response to the welfare losses created by polarized preferences.

Relation to other research We contribute to the literature on strategic communication and segregation, and its applications, by developing a tractable many-to-many cheap-talk framework with endogenous group choice. Galeotti et al. (2013) study one-to-one communication in networks with correlated signals, which prevents equilibria from being ordered by informativeness as in our lemma 1, and does not allow for an analysis of strategic group choice. Their framework has been applied to alliance formation (Penn, 2016), party factions (Dewan and Squintani, 2016), and deliberation design (Patty, 2022). While Patty is closest to our setting, room composition is exogenous and there is no strategic room choice in his paper.

Hagenbach and Koessler (2010) also study one-to-one communication in networks with coordination motives. While their truth-telling result also involves distance between biases, it arises from coordination incentives. Our main contribution to the literature on many-player communication is to combine a welfare measure based on information content with a strict informativeness ranking of equilibria, which allows a joint analysis of individual and socially optimal room choice.

Our paper adds a different perspective to a substantial literature on "echo chambers" and related concepts; see Barberá et al. (2018); Levy and Razin (2019); Arguedas et al. (2022) for overviews. Our focus on strategic senders and receivers distinguishes our work from studies that consider "echo chambers" through the choice of exogenous news sources or algorithm designers (Che and Mierendorff, 2019; Acemoglu et al., 2021; Martinez and Tenev, 2024; Williams, 2024) or focus on non-Bayesian learning mechanisms (Levy and Razin, 2019). Studies on the benefits of homophily when learning in exogenous networks (Lobel and Sadler, 2016; Mostagir and Siderius, 2023) also highlight advantages of small homogeneous groups, but do not have our focus on strategic communication and endogenous group formation.

Concerns about echo chambers are prominently articulated by (Sunstein, 2001, 2017), who argues that self-segregation undermines democratic deliberation. Empirically, Gentzkow and Shapiro (2011) show that ideological segregation online is comparable in magnitude to offline contexts. We propose to view segregation as an endogenous outcome that is a reaction to polarization rather than its cause. Segregation can make communication more effective in polarized societies and fighting segregation rather than polarization may worsen communication.

1. Model

Overview The state of the world is $\theta = \sum_{i=1}^{n} \theta_i$, where each $\theta_i \in \{0, 1\}$ is drawn with equal probability. Each player $i \in \{1, \dots, n\}$ observes a private binary signal $\sigma_i \in \{\sigma^l, \sigma^h\}$ about θ_i with accuracy $p > 1/2$, i.e. $Pr\left(\sigma_i = \sigma^h | \theta_i = 1\right) = Pr\left(\sigma_i = \sigma^l | \theta_i = 0\right) = p$. Before observing signals, players each choose to enter one of n rooms. Entering a room is costless and rooms have no capacity constraints, but each player can only be in exactly one room. After observing their signals, players send a binary cheap-talk message to others in their room and then choose an action $a_i \in \mathbb{R}$.

Payoffs Player i's payoff is

$$\begin{aligned} u_i(a, \theta) &= -(a_i - b_i - \theta)^2 - \alpha \sum_{j \neq i} (a_j - b_i - \theta)^2 \\ &= -\left(a_i - b_i - \sum_{k=1}^{n} \theta_k\right)^2 - \alpha \sum_{j \neq i} \left(a_j - b_i - \sum_{k=1}^{n} \theta_k\right)^2 \end{aligned} \tag{1}$$

where $b_i \in \mathbb{R}$ is i's commonly known *bias* and $\alpha > 0$ measures how much players care about the actions of others. Thus, i prefers all players' actions to equal $b_i + \theta$. We interpret b_i as i's preference and σ_i as i's private information. Players maximize expected payoffs.

Timing of the game

1. Players simultaneously choose a room.[4]
2. Players observe their signals σ_i and simultaneously send messages m_i. Message m_i is observed by all players in the same room as i.
3. Players simultaneously choose actions a_i; payoffs are realized.

Independence We assume independence of substates and signals: for $i \neq j$, θ_i is independent of θ_j and σ_i is independent of σ_j. These assumptions are mainly made for tractability and imply that player i's signal contains no information about player j's information. One can interpret θ_i as a part of the world about which player i has exclusive information. This does not rule out additional public information observed by everyone, see the supplementary material (section 6) for an explicit extension along those lines. σ_i can hence be thought of as i's residual private information, after we have taken account of all public information. Many of our results still hold in setups where signals are correlated (cf. section 8.3 in the supplementary material). However, a closed form characterization of the most informative communication equilibrium becomes intractable – cf. the discussion after lemma 1.

Notation We denote the room in which player i is by R_i and let n_{R_i} be the number of players in room R_i.[5]

Strategies A room choice strategy in stage 1 is a number in $\{1, \ldots, n\}$, so that all players who pick the same number are in the same room. A (potentially mixed) messaging strategy in stage 2 is a map $m_i : \{\sigma^l, \sigma^h\} \to \Delta\left(\{m^l, m^h\}\right)$ where the restriction to a binary message space is without loss of generality due to the binary signal space. Stage 3 action strategies map own signal σ_i and messages sent in R_i into an action $a_i \in \mathbb{R}$. While messaging or action strategies could in principle depend on the composition of other rooms, it is straightforward to show that players cannot gain from such a dependence and we therefore ignore this possibility without loss of generality.

Solution method We solve the model by backward induction: We solve for actions given messages, for messages given room composition, and finally for room choice. The solution concept is Perfect Bayesian Equilibrium. We follow the usual convention and focus on the most informative equilibrium in the messaging stage; this means that all messages occur in equilibrium and there are no off-path beliefs to consider.

[4] In equilibrium, this is equivalent to a stage in which players can enter and leave rooms until nobody wants to switch rooms anymore.

[5] Note that R_i is *not* the ith room *but* the room player i is in, so e.g. $R_2 = R_3$ if players 2 and 3 are in the same room.

2. Equilibrium Behavior Within a Room

2.1. Choice of Action

Only the first term in (1) affects i's optimal action a_i^*. The first-order condition yields

$$a_i^* = b_i + \mathbb{E}[\theta | m_{R_i}, \sigma_i] = b_i + \sum_{j=1}^{n} \mathbb{E}[\theta_j | m_{R_i}, \sigma_i], \tag{2}$$

where m_{R_i} denotes the profile of messages sent in room R_i. Thus, i's optimal action equals his bias plus his posterior expectation of the state.

In the following, we will denote by $\mu_{ij} = \mathbb{E}_i[\theta_j | m_{R_i}, \sigma_i]$ i's expectation about θ_j, so that expression (2) becomes $a_i^* = b_i + \sum_{j=1}^{n} \mu_{ij}$.[6]

2.2. Choice of Message

We now analyze message choice within a given room. Since messages and actions in other rooms are irrelevant, we can focus on finding an equilibrium among the players in a specific room.

Definition 1. *We call a messaging strategy m_i ...*

- babbling *if m_i is independent of i's observed signal σ_i.*
- truthful *if $m_i(\sigma^l) = m^l$ and $m_i(\sigma^h) = m^h$.*
- lying *if $m_i(\sigma^h) = m^l$ or $m_i(\sigma^l) = m^h$.*
- pure *if m_i is either babbling or truthful.*
- mixed *if for some signal σ^k, $k \in \{l, h\}$, both messages are sent with positive probability and the strategy is not babbling.*

The cheap talk game, as usual, admits multiple equilibria, including babbling equilibria. As in the cheap talk literature, we focus on the most informative equilibrium. In our model, unlike in general multi-sender games, this notion is well defined and characterized below.[7] The following lemma structures the set of equilibria:

Lemma 1. *If $(m_1, \ldots, m_n)$ is an equilibrium and m_i is mixed, then there exists another equilibrium (m_i^t, m_{-i}), where m_i^t is the truthful strategy.* *(Proof on page 19.)*

[6] i's optimal action is independent of j's optimal action. Our model is not about coordination, see Hagenbach and Koessler (2010) for a model of communication with coordination motives.

[7] Information transmission can sometimes be increased if players commit to playing specific, less-informative equilibria in certain rooms in order to affect room choice – see supplementary material for an example. Since this would require players committing to equilibria that would leave everyone strictly worse off, we do not consider such schemes.

Suppose player i mixes after observing σ^h, and is therefore indifferent between sending m^h, which induces higher actions, and m^l, which induces lower actions. Thus, actions induced by m^h are too high from i's perspective, while those induced by m^l are too low. After σ^l, i always sends m^l, since desired actions are increasing in the signal.

In such an equilibrium, m^h thus perfectly reveals σ^h. Now consider a truthful equilibrium. Sending m^h after σ^h induces the same actions as before, while sending m^l induces strictly lower actions. Since i was previously indifferent, he now strictly prefers m^h after σ^h, so truth-telling is optimal. Since i can only have been mixing after σ^h if i's bias is below average, he also benefits from σ^l inducing lower actions in the truth-telling equilibrium. Switching i from a mixed strategy to truthtelling does not affect other players' incentives, because states and signals are additive and independent.[8]

Lemma 1 implies that the most informative equilibrium is always pure: starting from any mixed equilibrium, we can stepwise replace mixed strategies by truthful ones without destroying equilibrium. Since truthful strategies are more informative (in the Blackwell sense), this strictly improves information transmission.

Corollary 1. *The most informative equilibrium in a room is always in pure strategies.*

We now characterize the most informative equilibrium. Player i's incentive to tell the truth depends on the distance between b_i and the biases of others in the room, since larger differences increase incentives to distort messages. The formal result in the following theorem is illustrated in Figure 2.

Theorem 1. *Let $\bar{b}_{R_i} = \frac{\sum_{k \in R_i} b_k}{n_{R_i}}$ be the mean bias in room R_i. In the most informative equilibrium, i tells the truth if and only if*

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

and babbles otherwise. *(Proof on page 20.)*

The truth-telling interval increases with room size n_{R_i} and signal precision p. Larger rooms matter because incentives depend on i's distance from the average bias of the *other* players – and this average is closer to $\bar{b}_{R_i}$ when n_{R_i} is large. Higher signal precision strengthens the disciplinary effect of lying, since a false message moves others' actions more.

3. Room Choice

We now analyze room choice, assuming that the most informative equilibrium is played in each room. This section characterizes the welfare-optimal room allocation; section 4

[8] If players receive signals about the same state and therefore have correlated information, no version of lemma 1 holds and much of the tractability is lost – but some of our results can still be derived, see section 8.3 in the supplementary material.

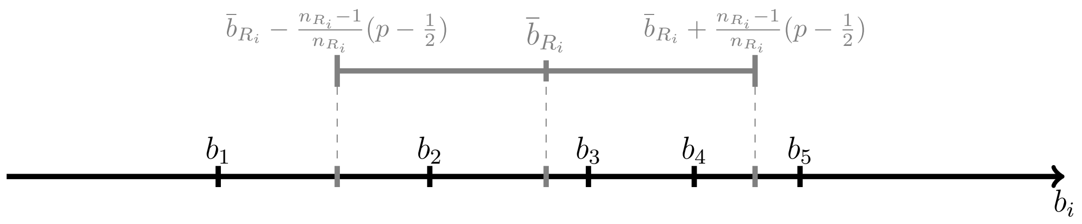

Figure 2: We find the most informative equilibrium in the room containing players 1 to 5 by constructing a symmetric interval around the average bias. Players 1 and 5 babble; players 2, 3 and 4 tell the truth.

studies when it arises in equilibrium.

Given expression (1), overall welfare is given by

$$W(a,b,\theta) = \sum_{i=1}^{n} u_i(a,b_i,\theta) = -\sum_{i=1}^{n}\left(a_i - b_i - \sum_{k=1}^{n}\theta_k\right)^2 - \alpha\sum_{i=1}^{n}\sum_{j\neq i}\left(a_j - b_i - \sum_{k=1}^{n}\theta_k\right)^2.$$

but this expression is not directly useful for comparing room allocations. Instead, we show that in our model, welfare can be expressed in terms of the amount of information players hold after communication.

Because payoffs are quadratic, they can be additively separated into losses from preference difference and losses from residual uncertainty. The former are unavoidable in equilibrium, while the latter depend on how much information is exchanged. These losses can be further decomposed into the theoretical loss from complete ignorance (if nobody had any information) and the informational gain from signals and communication. Only this gain depends on room allocation. Since the most informative equilibrium is in pure strategies, we can measure this gain by counting the pieces of information each player has after communication has taken place.

Consider the information available to a single player. i always receives his own signal, which we count as one piece of information. Each truthful message from another player adds one additional piece. Let $\zeta_i \in \{1, 2, \ldots, n\}$ denote the total number of such pieces of information available to i. Since each signal is binary, ζ_i in fact measures information in *bits*. The following result shows that all welfare comparisons reduce to informational accounting in bits:

Proposition 1. *If the most informative equilibrium is played within all rooms,*

1. *player i's payoff is affine and increasing in $\zeta_i + \alpha\sum_{j\neq i}\zeta_j$.*
2. *welfare is affine and increasing in $\sum_{i=1}^{n}\zeta_i$.*

(Full expressions in the proof on page 21.)

Room choice therefore becomes a purely informational problem: player i maximizes a weighted sum of his own information and that of others. When switching rooms, i trades off how much he learns against how much his move changes others' information. The weight of this "informational externality" is given by α. If $\alpha = 1$, there is no informational externality and individual incentives maximize welfare:

Corollary 2. *If $\alpha = 1$, the welfare-optimal room allocation is an equilibrium of the room choice game.*

Proposition 1 allows for rapid welfare comparisons. In figure 2, full integration yields $4+3+3+3+4 = 17$ information pieces, which cannot be exceeded by any partition into smaller rooms. Hence the illustrated allocation is welfare-optimal.

For larger n, such exhaustive comparison becomes computationally hard due to the rapid growth of the number of possible partitions of a set (given by the Bell sequence). We therefore derive general results on optimal and equilibrium room allocations in the next section.

4. Polarization and Segregation

Using the geometric characterization of messaging and the informational formulation of room choice, we now analyze how preference polarization shapes which room allocations are welfare-optimal and which arise in equilibrium.

Figure 1 in the introduction already gave an intuitive example in which segregation is efficient but not an equilibrium. Section 4.1 generalizes the insights from this example by characterizing the relationship between parameters, efficiency and equilibrium for all models in which there are two bias types. Section 4.2 generalizes some insights from this model further to all conceivable generic bias configurations with an arbitrary number of biases and players; in the supplementary material we derive more specific results for particular bias configurations within this space. Finally, section 4.3 shows that welfare losses from polarization occur despite segregation, not through it.

4.1. Bipolar Polarization

We begin with the case where there are two bias groups, i.e. $b_i \in \{0, b\}$ for some $b > 0$. n_0 players have bias 0 and $n_b\,(= n - n_0)$ have bias b. For the full parameter space, we characterize (i) the welfare-optimal room allocation, (ii) whether it is an equilibrium and (iii) if not, the welfare-maximizing equilibrium.

The full characterization is in Theorem 2 below. Because it is case-heavy, we first state two high-level results that summarize the structure and mechanisms; proofs and derivations are in the supplementary material to this paper.

Result 1. *If all $b_i \in \{0, b\}$, then for small b the welfare optimum is full integration, and for large b it is full segregation by type; in both cases the welfare-optimal allocation is also a room-choice equilibrium.*

With full integration, everyone is truthful when the average bias is close enough to either type, e.g. when $\overline{b} = \frac{n_b b}{n} < \frac{n-1}{n}(2p-1)$. So integration is welfare-optimal (and an equilibrium) if b is sufficiently small. At the other extreme, any mixed room induces babbling by all occupants when one type is "too far" from the other, e.g. when $b > n_0(p-\frac{1}{2})$ (and symmetrically for n_b). Then segregation is welfare-optimal and an equilibrium.

For intermediate b, the welfare optimal room allocation need not be an equilibrium. Perhaps counterintuitively, the best equilibrium then features *too little segregation* compared to the welfare optimum.

Result 2. *If all $b_i \in \{0, b\}$ and the welfare-optimal room allocation is not an equilibrium, then equilibrium segregation is below the welfare-optimal level: welfare could be improved by moving players from mixed rooms into same-type rooms.*

The gap arises because players may ignore informational externalities from their room choice if $\alpha < 1$. They integrate more than a planner would; this is the case shown in panel (iv) of figure 1 in the introduction. To pin down all cases, we next state a theorem that exhaustively characterizes equilibrium and optimal allocations.

Theorem 2. *Let all $b_i \in \{0, b\}$ and (w.l.o.g.) $n_0 \geq n_b$.*
(Welfare-optimality) *The welfare maximizing room allocation and messaging equilibrium are:*

1. *For $b/(p-1/2) \leq (n-1)/n_0$, all players are in one room with universal truth-telling.*
2. *For $(n-1)/n_0 < b/(p-1/2) \leq (n-1)/n_b$, let $n_{m0} = \lfloor (n_b - 1)/(b/(p-1/2) - 1) \rfloor$.*
   (a) *If $(n_b + n_{m0})^2 + (n_0 - n_{m0})^2 \leq n_0(n_0 + n_b) + n_b$ and $n_{m0} \geq n_b$, all players are in one room where $b_i = 0$ types are truth-telling and $b_i = b$ types babble.*
   (b) *Otherwise, one room contains $n_0 - n_{m0}$ players with $b_i = 0$, and the other contains all remaining players; universal truthtelling.*
3. *For $b/(p-1/2) > (n-1)/n_b$, let $n_{mb} = \lfloor (n_0 - 1)/(b/(p-1/2) - 1) \rfloor$.*
   (a) *If $n_0(n_0+n_{mb})+n_{mb}+(n_b-n_{mb})^2 \geq n_0^2+n_b^2$, one room contains n_b-n_{mb} players with $b_i = b$, and the other contains all other players in which only players with $b_i = 0$ are truth-telling,*
   (b) *Otherwise, full segregation with universal truth-telling.*

(Equilibrium) *The welfare optimal room allocation is an equilibrium in cases (1), (2a) and (3a) and also*

- *in case (2b) if*

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

- *in case (3b) if* $\alpha \geq (1 + n_0 - n_b)/(n_b - 1)$.

Otherwise, the welfare optimal equilibrium features too little segregation: In case (2b), it has one room instead of two, and in case (3b) it is not fully segregated. (Proofs and a step-by-step derivation in the supplementary material.)

Figures 3 and 4 illustrate the welfare-equilibrium relationship for equally-sized groups and one case of different-sized groups.

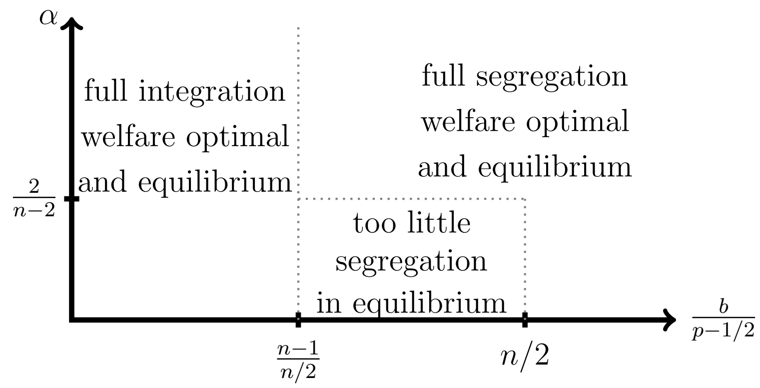

Figure 3: The relationship between welfare-optimum and optimal equilibrium in a model with bipolar polarization and equal-sized bias groups. The horizontal axis describes the polarization of preferences relative to the dispersion of information.

This result shows that information externalities can take two possible forms:

Entering externality: In case (2b), a player reduces information in the room she enters. Welfare would be increased by removing some majority-type players to a small separate room to balance a mixed room, but these players prefer to rejoin the large room because it yields them more information, even though their entry causes the minority type to babble.

Leaving externality: In case (3b), a player reduces information in the room she leaves. Full segregation is welfare-optimal, but minority players prefer to join the other room to improve their own learning since it does not destroy truth-telling by majority players in that room. The minority player who has switched, however, will not communicate truthfully, and this deprives all other players of her information.

One implication of theorem 2 is that in the welfare-optimal allocation, no player in the mixed room has an incentive to move to a less mixed room, since the mixed room always contains the largest number of truthful communicators.

As Figure 4 shows, equilibrium optimality can be non-monotonic in b: in an intermediate range between cases (2b) and (3b), it is not possible to "balance" the fully integrated room by removing b-types, but a fully segregated room would be worse than a fully integrated room in which only one type communicates informatively.

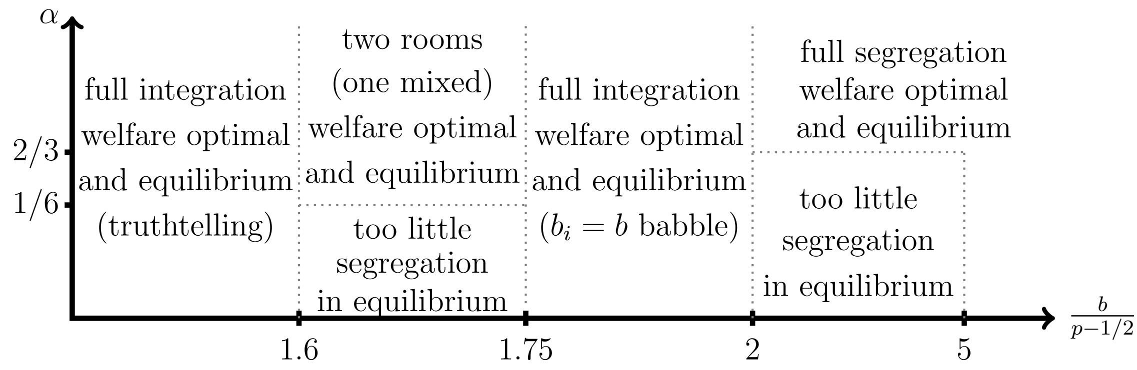

Figure 4: The relationship between welfare-optimum and optimal equilibrium in a model with bipolar polarization and bias groups of unequal size ($n_0 = 5 > 4 = n_b$). The horizontal axis describes the polarization of preferences relative to the dispersion of information. (Axis values are not to scale to improve readability.)

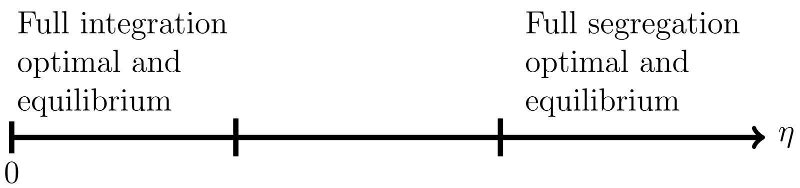

Figure 5: Illustration of theorem 3: The welfare-optimal room allocation is an equilibrium if preference polarization (measured by η) is small or large.

4.2. When is Segregation Optimal?

We have seen that integration and segregation are optimal, respectively, under low and high polarization. We now extend this result to arbitrary bias configurations. Let $\mathcal{B} = \langle b_1, \ldots, b_n \rangle$ be a bias configuration (this is not a set since repeated values are allowed). Assume $\mathcal{B}$ is generic in the sense that no bias equals the average of any subset of the others.[9] Define $\mathcal{B}_\eta = \langle \eta b_1, \ldots, \eta b_n \rangle$ with $\eta > 0$, which scales preference polarization relative to information dispersion. Larger η correspond to greater effective polarization.[10]

Theorem 3. *(i) For sufficiently small η, full integration is welfare-optimal and a room-choice equilibrium for bias configuration $\mathcal{B}_\eta$.*

(ii) For sufficiently large η, full segregation by bias types is generically welfare-optimal and a room-choice equilibrium for bias configuration $\mathcal{B}_\eta$. (Proof on page 23.)

Figure 5 summarizes the result. When biases are close relative to informational differences, full integration results in universal truth-telling, which maximizes welfare and all individual payoffs. When biases are very dispersed relative to information and no special

[9] 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 vector of $\tilde{n}_{b_j} \in \{0, 1, \ldots, n_{b_j}\}$ where n_{b_j} is the number of players with bias b_j.

[10] η is the polarization measure proposed by Esteban and Ray (1994) (theorem 1) with an appropriate scaling parameter.

symmetry holds, truth-telling is impossible in any mixed room. Full segregation is therefore welfare-optimal and stable, since no deviation can increase any player's information.

For intermediate polarization, no general ordering of bias configurations is possible without imposing more structure. Welfare-best equilibria may exhibit either too much or too little segregation; we provide an example of excessive equilibrium segregation in the supplementary material. Our framework can nonetheless analyze specific bias distributions, as we also show in the supplementary material. For uniformly distributed biases, (almost) full integration is welfare-optimal and full integration is an equilibrium. For single-peaked bias distributions, full integration can be welfare-optimal and an equilibrium, but only if the concentration around the center is sufficiently strong.

4.3. Polarization Destroys Welfare

Although segregation is a welfare-increasing response to polarization, polarization itself reduces welfare. Under the η-parameterization of the previous section, both welfare and total information transmission weakly decrease in η.

Proposition 2. *Let $W(\eta)$ be welfare and $\mathcal{Z}(\eta)$ total pieces of information in the welfare-optimal allocation under $\mathcal{B}_\eta$. Both $W(\eta)$ and $\mathcal{Z}(\eta)$ are decreasing in η. (Proof on page 24.)*

It follows from theorem 1 that increasing η weakly reduces communication in every room and therefore lowers welfare. Additional segregation can partially restore communication, but cannot fully offset this loss. Polarization lowers welfare through communication breakdown, not through segregation – even though it may look that way, since more polarization is accompanied by more segregation in equilibrium. Segregation is a response that mitigates, rather than causes, the communication breakdown, even though it cannot restore communication across separated groups. Echo chambers are therefore a decentralized defense mechanism to polarization rather than its cause. Reducing polarization improves welfare, while reducing segregation alone may not.

5. Extensions: Uncertainty, Public Information, and Follower Networks

Our model can be extended in various ways. We briefly discuss three extensions; further details can be found in the supplementary material.

Our framework easily generalizes to a setting with **uncertainty** about others' preferences, such as in anonymous online settings. Unlike Morgan and Stocken (2003) and Li and Madarász (2008), we consider uncertainty in the sign as well as size of preferences, and whether uncertainty helps or hurts depends on the exact shape of the probability distributions. For several partial distribution orderings, we show that large increases in uncertainty erode all truth-telling. Even in a setting where full integration is optimal if preferences are known, segregation may be optimal if there is preference uncertainty.

We allow for **public information** in addition to private signals. This does not change the basic mechanisms, but weakens incentives for truth-telling: as public information becomes more important, private messages matter less for actions, making it more tempting to mislead. When public information is sufficiently important, full segregation is optimal and arises in equilibrium. This suggests that settings dominated by public information (such as national political debates) lead to less truth-telling and more segregation, and advances in information technology may crowd out truth-telling incentives.

Many of our main results also apply in a modified setting where players can form **follower networks** by choosing whose message to receive. Player i tells the truth if her preferences are close enough to the average preferences of her followers; players therefore want to follow others unless doing so would make them stop communicating truthfully. Equilibrium network formation maximizes welfare and greater polarization leads to more segregation. The size of followership varies: Moderate players can be followed by many (since the average preference of their audience is close to their own), while extreme players are only followed by like-minded extremists.

6. Discussion

Social media and "rooms" In reality, segregated spaces such as the "rooms" in our model are often scarce or costly. If segregation is beneficial, people are willing to pay for institutions that allow selective inclusion and exclusion. This suggests that the business model of social media platforms is not merely to connect users, but to provide tools to structure and segregate audiences in ways that make communication possible. These group structures are also less portable than files or contacts lists, which contributes to platforms' market power.

Political Parties and Safe Spaces Political parties and similar organizations can be understood as self-created "rooms" that allow like-minded individuals to communicate honestly in polarized environments. Such spaces make it possible to discuss weaknesses and doubts about one's own position that one would otherwise hide or deny in the presence of outsiders. The same logic explains the informational value of "safe spaces", which can enable marginalized or minority groups to exchange information.

Room Choice as Communication Design Models of information design usually assume that a designer can set disclosure rules (Bergemann and Morris, 2019) or that senders can commit to them (Kamenica and Gentzkow, 2011). Our model shows that truthful communication can instead emerge from group composition, because people want to speak honestly to certain audiences. Room allocation therefore functions as a natural commitment device by making players *want* to tell the truth, without commitment power or external enforcement. Such "communication design" does not necessarily need external designers, but can emerge from individuals' own sorting behavior.

Is this a model of echo chambers? The academic and social debate has not agreed on a unified definition of "echo chambers". Our model does not aim to capture all significant features, but focuses on two elements that are common to all interpretations: communication in segregated spaces and systematic sorting into those spaces. We argue that the effects we describe would also be present in any richer model that includes strategic communication. If e.g. individual information was correlated with preferences, segregation has additional costs but retains its benefit (see supplementary material).

The equilibria of our model may not look, at first glance, as if players "echo" each other's messages, since they truthfully reveal their information. But the message space is arbitrary, and it is a (realistic and outcome-equivalent) equilibrium that all truth-telling players simply communicate their preferred action (given their information). An outside observer would then see an "echo" structure, in which messages in one room are systematically higher than those in another room.

When are echo chambers bad? Echo chambers can be useful through the mechanism we study, but their overall effects depend on whether this benefit outweighs other, potentially harmful forces. Even additional benefits to *diversity* would not change the main trade-off of our model and some segregation may remain efficient. If people in echo chambers learn poorly due to *correlation neglect or other inference mistakes*, exposure to opposing views would not automatically improve learning if communication remains strategic. *Endogenous polarization* would occur if preference evolve through past communication and would create a trade-off between present information exchange and future polarization.

Our model also implies that even *taste-based homphily* can improve information exchange when communication across groups would otherwise break down. Malicious actors who benefit from misinformation or confusion, however, could find it easier to spread misleading content in segregated societies, especially when audiences cannot correctly infer their motives.

7. Conclusion

Democratic societies aggregate information through debate, markets, and voting – but only debate requires no formal rules or framework beyond the ability to speak and listen. We have shown that when people differ in preferences and information, segregation into like-minded groups can be rational and Pareto-efficient. Echo chambers are thus better understood as consequences of polarization and mistrust rather than their cause. While society has much to gain from debate across divides, our argument implies that this cannot be achieved by forcing interactions that would not otherwise take place. Debate is more than putting people into a room and expecting them to come out smarter – it needs sufficient trust and common ground.

Appendix

A. Proofs

This appendix contains only the proofs for results that are explicitly given in the main text; all other results and their proofs can be found in the supplementary material.

Proof of lemma 1 on page 8.

Let $(m_1^*, \dots, m_n^*)$ be an equilibrium where m_i is a (possibly mixed) equilibrium strategy. We denote by m_{-i,R_i} the messages of all players in room R_i apart from i and by $a_i(m_{-i,R_i}, \sigma_i)$ player i's optimal action as derived in section 2.1. Player i's expected payoff when sending message m_i to players in room R_i can be written as

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

where expectations are taken over other players' signals. This $U_i(m_i|\sigma_i)$ can be split in a part that is independent of i's message m_i and a part that depends on m_i:

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

Specifically, sending message m^h gives expected payoff

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

where $\mu^h_{ji} = \mathbb{E}[\theta_i | m_i = m^h]$, i.e. μ^h_{ji} is the expectation of a player j in the same room as i concerning θ_i if player i sends message m^h. Note that this expectation is the same for all players $j \neq i$ in the same room as i. Sending message m^l gives

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

where $\mu_{ji}^l = \mathbb{E}[\theta_i | m_i = m^l]$. The difference in expected payoff (divided by α) is then

$$\begin{aligned}
\Delta U_i(\sigma_i)/\alpha &= (U_i(m^h) - U_i(m^l))/\alpha \\
&= -\sum_{j \in R_i, j \neq i} \mathbb{E}\left[{\mu_{ji}^h}^2 - {\mu_{ji}^l}^2 + 2(\mu_{ji}^h - \mu_{ji}^l)\left(b_j - b_i + \sum_{k \neq i} \mu_{jk} - \theta_i - \sum_{k \neq i} \theta_k \right) \middle| \sigma_i \right] \\
&= -2(\mu_{ji}^h - \mu_{ji}^l) \sum_{j \in R_i, j \neq i} \left[\frac{\mu_{ji}^h + \mu_{ji}^l}{2} + b_j - b_i - \mathbb{E}\left[\theta_i | \sigma_i\right] \right] \\
&= 2(\mu_{ji}^h - \mu_{ji}^l)(n_{R_i} - 1)\left[-\frac{\mu_{ji}^h + \mu_{ji}^l}{2} - \frac{\sum_{j \in R_i, j \neq i} b_j}{n_{R_i} - 1} + b_i + \mathbb{E}\left[\theta_i | \sigma_i\right] \right] \qquad (3)
\end{aligned}$$

where n_{R_i} denotes the number of players in room R_i. (For the transformation to line 3, we make use of the fact that μ_{ji} is the same for all $j \in R_i$.)

Player i is only willing to choose a mixed strategy after receiving signal σ_i if $\Delta U_i(\sigma_i) = 0$. From expression (3) 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 $\Delta U_i(\sigma^l) = 2(\mu_{ji}^h - \mu_{ji}^l)(n_{R_i} - 1)(1-2p) < 0$ and therefore $m_i(\sigma^l) = m^l$ which implies $\mu_{ji}^h = p$ as a m^h is only sent by i after receiving signal σ^h. Consequently, $(\mu_{ji}^h + \mu_{ji}^l)/2 \geq 1/2$ as $\mu_{ji}^l \geq 1 - p$. Now consider the equilibrium candidate (m_i^t, m_{-i}). With the truthful strategy m_i^t, $\mu_{ji}^{th} = p$ and $\mu_{ji}^{tl} = 1 - p$ and therefore $(\mu_{ji}^{th} + \mu_{ji}^{tl})/2 = 1/2$. This implies that $\Delta U_i(\sigma^h) > 0$ in the equilibrium candidate (m_i^t, m_{-i}), i.e. truthful reporting is optimal for i after receiving signal σ^h. In the equilibrium candidate (m_i^t, m_{-i}), truthful messaging is still optimal after signal σ^l as well: From $p > 1/2$, $\mu_{ji}^h \leq p$ and $\mu_{ji}^l \leq 1/2$ it follows that $-1/2 + (1-p) < -\left(\mu_{ji}^h + \mu_{ji}^l\right)/2 + p$. As in the original equilibrium (m_i, m_{-i}) we had $\Delta U_i(\sigma^h) = 0$ and therefore $-\left(\mu_{ji}^h + \mu_{ji}^l\right)/2 + p = \sum_{j \in R_i, j \neq i} b_j/(n_{R_i} - 1) - b_i$, we get that $-1/2 + 1 - p < \sum_{j \in R_i, j \neq i} b_j/(n_{R_i} - 1) - b_i$ and therefore $\Delta U_i(\sigma^l) < 0$ in the truthful equilibrium candidate (m_i^t, m_{-i}). Hence, truthful messaging is i's best response in the equilibrium candidate (m_i^t, 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_i^t. Hence, (m_i^t, m_{-i}) is an equilibrium.

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

Proof of theorem 1 on page 9.

Consider again the difference between lying and truth-telling for player i that we considered in equation (3) in the proof of lemma 1. Following corollary 1, we only consider pure strategies and therefore for every non-babbling player $\mu_{ji}^h = p$ and $\mu_{ji}^l = 1 - p$ which implies that $\Delta U_i(\sigma^h) \geq 0$ simplifies to

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

If this inequality does not hold, player i will not use the truthful strategy in the most informative equilibrium and by corollary 1 this implies that he will babble in the most informative equilibrium.

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

Proof of proposition 1 on page 10.

Denote the sets of babbling and truthful players in room R_j as R_j^{bab} and R_j^{truth}, respectively. For a given room allocation, the expected payoff of player i in room R_i is

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

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

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

Now note that $\mathbb{E}\left[(\mu_{jk}-\theta_k)^2\right]$ can have two possible values: If $k \in R_j^{truth} \cup \{j\}$, 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 $p(1-p) < \frac{1}{4}$.) Furthermore, $\mathbb{E}\left[(1/2-\theta_k)^2\right] = 1/4$.

This means that if i is telling the truth, we can write (the discussion in the next paragraph explains how we rearranged terms):

$$\begin{aligned} U_i^{truth} &= -\alpha \sum_{j\neq i} \{(b_j-b_i)^2\} - \frac{1}{4}\left[n+\alpha(n-1)n\right] \\ &+ (\frac{1}{4}-p(1-p))\left[n_{R_i}^{truth} + \alpha \sum_R \left\{n_R^{truth} n_R^{truth} + (n_R - n_R^{truth})(1+n_R^{truth})\right\} - \alpha n_{R_i}^{truth}\right]. \end{aligned} \tag{4}$$

The first term represents the loss that i suffers because other players choose a decision that is by $b_j - b_i$ too high from i's point of view. The second term represents the (theoretical) loss that would result if no player had any information and all μ's were simply $\frac{1}{2}$. The factors n and $(n-1)n$, which sum up to n^2, represent the total number of possible pieces of information in the model: If everybody's signal was available to everyone, n people would receive n pieces of information. The term hence represents, for each potential piece of information, the loss to i of that information not being available.

This loss is mitigated by information, which we see in the second line: i receives his signal and $n_{R_i}^{truth} - 1$ truthful messages, which means that instead of $\frac{1}{4}$, on each of these pieces of information i loses only $p(1-p) < \frac{1}{4}$. Other players, about whose decisions i cares with weight α, also receive some signals/messages: in any given room R, n_R^{truth} players receive their own signal and $n_R^{truth} - 1$ truthful messages while $n_R - n_R^{truth}$ players (those that babble in R) receive n_R^{truth} truthful messages and their own signal. (We have to subtract the correction term $-\alpha n_{R_i}^{truth}$ for room R_i in which there are only $n_{R_i}^{truth} - 1$ other players who tell the truth – in other words, i cannot count himself again as one of the players who receive information.) Analogously, we can write

$$\begin{aligned} U_i^{bab} &= -\alpha \sum_{j\neq i} \{(b_j-b_i)^2\} - 1/4\left[n+\alpha(n-1)n\right] \\ &\quad +(1/4-p(1-p))\left[1 + n_{R_i}^{truth} + \alpha \sum_R \left\{n_R^{truth} n_R^{truth} + (n_R - n_R^{truth})(1+n_R^{truth})\right\}\right. \\ &\quad \left. -\,\alpha(1+n_{R_i}^{truth})\right]. \end{aligned} \tag{5}$$

In both the expressions for U_i^{truth} and U_i^{bab}, the second lines are adjusting the (pessimistic) expression in the first line for the reduction in variance by information. We can simplify both expressions by simply writing

$$U_i = -\alpha \sum_{j\neq i}\{(b_j - b_i)^2\} - 1/4\,[n + \alpha(n-1)n] + (1/4 - p(1-p))\left[\zeta_i + \alpha \sum_{j\neq i} \zeta_j\right] \quad (6)$$

and express welfare as

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

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

Proof of theorem 3 on page 14.

A truth-telling equilibrium exists if $\Delta U_i(\sigma^h) \geq 0$ and $\Delta U_i(\sigma^l) \leq 0$ for each player i. Using the expressions as in the proof of theorem 1 for the case of one room R, $U_i(\sigma^h) = b_i - (\sum_{j\neq i} b_j/(n-1)) - 1/2 + p$ and $U_i(\sigma^l) = b_i - (\sum_{j\neq i} b_j/(n-1)) + 1/2 - p$. A truth-telling equilibrium therefore exists if and only if for every player i it is

$$\left|\sum_{k\neq i}\{b_k/(n-1)\} - b_i\right| \leq p - \frac{1}{2}.$$

This can be rewritten as $\left|\sum_k\{b_k\} - nb_i\right|/(n-1) \leq p - \frac{1}{2}$. If η is sufficiently small, 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, and no player can gain from leaving the room.

If $\left|\sum_{k\in R_i, k\neq i}\{b_k/(n-1)\} - b_i\right| > p - \frac{1}{2}$, 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 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 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. □

Proof of proposition 2 on page 15

Take two values of η, namely η' and $\eta'' > \eta'$. Denote a welfare optimal room assignment under η'' by R''. Consider the same room assignment R'' with biases η'. In each room the number of pieces of information is weakly higher with set of biases $B_{\eta'}$ than with set of biases $\mathcal{B}_{\eta''}$: By theorem 1 a player i is truthtelling if and only if $\eta\bar{b}_{R''_i} - \frac{n_{R''_i}-1}{n_{R''_i}}(p-\frac{1}{2}) \leq \eta b_i \leq \eta\bar{b}_{R''_i} + \frac{n_{R''_i}-1}{n_{R''_i}}(p-\frac{1}{2})$. Hence, player i will be truthtelling in room R''_i with biases in $\mathcal{B}_{\eta'}$ if he is truthtelling in R''_i with biases $\mathcal{B}_{\eta''}$ by $\eta' < \eta''$. Consequently, there is weakly more information transmitted in every room given assignment R'' under η' than under $\eta'' > \eta'$. This implies $W(\eta') \geq W(\eta'')$ by proposition 1. $\square$

References

Acemoglu, D., A. Ozdaglar, and J. Siderius (2021). Misinformation: Strategic sharing, homophily, and endogenous echo chambers. National Bureau of Economic Research Working Paper, No. 28884.

Arguedas, A., C. T. Robertson, R. Fletcher, and R. K. Nielsen (2022). Echo chambers, filter bubbles, and polarisation: A literature review. Report, Reuters Institute for the Study of Journalism, University of Oxford.

Barberá, P., J. T. Jost, J. Nagler, J. A. Tucker, and R. Bonneau (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science 26*(10), 1531–1542.

Barberá, P., J. A. Tucker, A. Guess, C. Vaccari, A. Siegel, S. Sanovich, D. Stukal, and B. Nyhan (2018). Social media, political polarization, and political disinformation: A review of the scientific literature.

Bergemann, D. and S. Morris (2019). Information design: A unified perspective. *Journal of Economic Literature 57*(1), 44–95.

Chater, J. (2016). What the EU referendum result teaches us about the dangers of the echo chamber. `https://www.newstatesman.com/2016/07/what-eu-referendum-result-teaches-us-about-dangers-echo-chamber`. Accessed: 2021-07-02.

Che, Y.-K. and K. Mierendorff (2019). Optimal dynamic allocation of attention. *American Economic Review 109*(8), 2993–3029.

Del Vicario, M., A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences 113*(3), 554–559.

Dewan, T. and F. Squintani (2016). In defense of factions. *American Journal of Political Science 60*(4), 860–881.

Esteban, J.-M. and D. Ray (1994). On the measurement of polarization. *Econometrica 62*(4), 819–851.

Galeotti, A., C. Ghiglino, and F. Squintani (2013). Strategic information transmission networks. *Journal of Economic Theory 148*(5), 1751–1769.

Gentzkow, M. and J. M. Shapiro (2011). Ideological segregation online and offline. *Quarterly Journal of Economics 126*(4), 1799–1839.

Grimes, D. R. (2017). Echo chambers are dangerous – we must try to break free of our online bubbles. https://www.theguardian.com/science/blog/2017/dec/04/echo-chambers-are-dangerous-we-must-try-to-break-free-of-our-online-bubbles. Accessed: 2021-07-02.

Hagenbach, J. and F. Koessler (2010). Strategic communication networks. *Review of Economic Studies 77*(3), 1072–1099.

Hooton, C. (2016). Social media echo chambers gifted Donald Trump the presidency. https://www.independent.co.uk/voices/donald-trump-president-social-media-echo-chamber-hypernormalisation-adam-curtis-pro
html. Accessed: 2021-07-02.

Itten, A. (2018). Coming undone: How echo chambers balkanised society. https://www.politics.co.uk/comment-analysis/2018/08/16/coming-undone-how-echo-chambers-balkanised-society. Accessed: 2021-07-02.

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*(6), 2590–2615.

Lawrence, E., J. Sides, and H. Farrell (2010). Self-segregation or deliberation? Blog readership, participation, and polarization in American politics. *Perspectives on Politics 8*(1), 141–157.

Levy, G. and R. Razin (2019). Echo chambers and their effects on economic and political outcomes. *Annual Review of Economics 11*(1), 303–328.

Li, M. and K. Madarász (2008). When mandatory disclosure hurts: Expert advice and conflicting interests. *Journal of Economic Theory 139*(1), 47–74.

Lobel, I. and E. Sadler (2016). Preferences, homophily, and social learning. *Operations Research 64*(3), 564–584.

Martinez, G. and N. H. Tenev (2024). Optimal echo chambers. arXiv preprint arXiv:2010.01249.

Morgan, J. and P. C. Stocken (2003). An analysis of stock recommendations. *RAND Journal of Economics 34*(1), 183–203.

Mostagir, M. and J. Siderius (2023). Social inequality and the spread of misinformation. *Management Science 69*(2), 968–995.

Patty, J. W. (2022). Designing deliberation for decentralized decisions. *American Journal of Political Science*, forthcoming.

Penn, E. M. (2016). Engagement, disengagement, or exit: A theory of equilibrium associations. *American Journal of Political Science 60*(2), 322–336.

Quattrociocchi, W., A. Scala, and C. R. Sunstein (2016). Echo chambers on Facebook. Available on SSRN.

Sunstein, C. R. (2001). *Republic.com.* Princeton University Press.

Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media.* Princeton University Press.

Williams, C. R. (2024). Echo chambers: Social learning under unobserved heterogeneity. *Economic Journal 134*(658), 837–855.