Feed for Good? On the Effects of Personalization Algorithms in Social Media Platforms*

Miguel Risco† Manuel Lleonart-Anguix‡

November 1, 2025

Abstract

We study how engagement-maximizing personalization shapes exposure and learning on social media. In a tractable model, users observe private signals, post, and consume a ranked feed; utility combines sincerity, conformity, and decision accuracy. Engagement is endogenous and users are heterogeneous: rational users choose it; naive users continue scrolling via a payoff-dependent hazard. Truthtelling is an equilibrium—unique for rationals—so platforms act through ordering alone. The optimal ranking is similarity-first, producing echo-chamber exposure; engagement is finite and, for naives, learning stalls at scale. Reverse-chronological and balanced-insertion counterfactual algorithms restore diversity; network effects may fail for naives.

Keywords: personalization algorithms, social learning, network effects, horizontal interoperability

JEL Codes: D43, D85, L15, L86.

*Risco acknowledges financial support from MUR PRIN 2022, Prot. 2022FB4LKK, "Finanziamento dell'Unione Europea—NextGenerationEU"—missione 4, componente 2, investimento 1.1, CUP J53D23004910001 and by the German Research Foundation (DFG) through CRC TR 224 (Project B05) and funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 949465). Lleonart-Anguix acknowledges financial support of the Spanish *Ministerio de Ciencia e Innovación* through grant PRE2021-099026. We thank Sven Rady, Pau Milán, Francesc Dilmé, Alexander Frug, Carl-Christian Groh, Francesco Decarolis, Rafael Jiménez-Durán, Martino Banchio, Marco Ottaviani, Justus Preusser, Muxin Li, David Jiménez-Gómez, Daniel Krähmer, Philipp Strack, Marc Bourreau, Mikhail Drugov, Manuel Mueller-Frank, Robin Ng, Raquel Lorenzo, Malachy Gavan, Francisco Poggi and Fernando Payro and the participants at the BSE Summer Forum in Digital Platforms, the EAYE 2024, the Paris Digital Economics Conference, the MaCCI Annual 2024, the CRC TR 224 retreat, the 6th Workshop on The Economics of Digitalization, the EDP Jamboree, the 2nd ECONtribute YEP Workshop, the CoED, the JEI, various CRC TR 224 YRWs as well as seminars and workshops at the University of Bonn, Bocconi University, UPF, TSE, UAB, UV, UA, EUI and UB for their helpful comments and discussions.

†IGIER, Università Bocconi. miguel.riscobermejo@unibocconi.it

‡Universitat Autònoma de Barcelona and Barcelona School of Economics. manuel.lleonart@bse.eu

1 Introduction

While entertainment may be the primary reason users open TikTok or Instagram, social media has become a pivotal source of news and opinion formation. More than half of U.S. adults now report getting news from social platforms, and the shares have risen steeply in recent years: between 2020 and 2024, the fraction of adults regularly accessing news on TikTok grew from 3% to 17%, and on Instagram from 11% to 20%.[1] This scale and centrality have sharpened concerns that engagement-maximizing algorithms amplify polarization, misinformation, and echo chambers. The public debate, energized since the 2016 U.S. election, initially focused on the spread of false content and platform incentives (Silverman, 2016; Allcott and Gentzkow, 2017) and has since broadened with empirical evidence on the welfare costs of social media use (Allcott et al., 2022; Braghieri et al., 2022; Bursztyn et al., 2023) and on algorithmic forces that trap users in ideologically narrow information environments (Levy, 2021). Internal platform documents further indicate awareness of such harms, especially for vulnerable groups (Horwitz et al., 2021). At the same time, social media affords sizable benefits—from access to timely information and social connection to professional networking—underscoring that policy must navigate genuine trade-offs (Allcott et al., 2020; Armona, 2023).

The feed is the critical design lever. Before roughly 2009, most platforms presented posts in reverse-chronological order.[2] Today, the feed is a personalized, ordered list, optimized to maximize attention (*engagement*). Because platform revenue is advertising-based, maximizing profits closely aligns with maximizing engagement, not with maximizing learning.[3] Empirical evidence links algorithmic personalization to higher time-on-platform and more addictive usage (Guess et al., 2023). These observations motivate a positive, micro-founded analysis of how an engagement-maximizing platform designs the feed, how users communicate and learn within it, and which regulatory tools improve outcomes.

We build a tractable model of communication and learning through personalized *feeds* (the ordered list of posts a user reads when scrolling down). Users receive private signals about an unknown state, post a message, and consume a feed whose order is strategically designed by the platform. Users' payoffs have two components. The first, within-the-platform utility, includes (i) direct gratification from reading, (ii) a taste for sincerity (reporting a message close to one's private signal), and (iii) disutility from disagreement (conformity costs).[4] The second, action utility, rewards taking a decision close to the true

1 Pew Research Center, Social Media and News Fact Sheet, data available here.

2 Facebook began introducing personalized feeds in 2009; Twitter (now X) and Instagram transitioned in 2015–2016; TikTok launched with a curated feed.

3 As an example, see here and Kamath et al. (2014) on RealGraph predictor, the base of Twitter's recommendation algorithm.

4 Conformity is a core force in social influence, defined as matching attitudes and behaviors to group norms (Cialdini and Goldstein, 2004). We model it as a reduced-form preference, consistent with

state after reading the feed (e.g., whether to vaccinate when facing contested information about vaccine efficacy).[5] Engagement is the number of posts read. We allow heterogeneous behavioral frictions: rational users choose how far to scroll; naive users continue according to a smooth, increasing function of their instantaneous within-platform utility, consistent with habit formation and present bias in digital consumption (Hoong, 2021; Allcott et al., 2022).

Methodologically, we impose improper priors so that each user treats her own signal as the local anchor. Our first result pins down the channel through which platforms exercise influence: *truthtelling is an equilibrium of the messaging game under any feed design*. Among rational users the equilibrium is unique, and for naive users it is the only equilibrium robust to any continuous, increasing continuation rule. Hence the platform cannot manipulate beliefs by inducing misreporting; it operates solely through *who appears and when* in the user's feed. This separability delivers closed-form posteriors and sharp comparative statics for ranking and engagement.

We leverage this benchmark to deliver three main findings. The first one is that engagement-optimal feeds generate echo chambers and, for naives, unravel the classic *wisdom-of-the-crowds* result (Golub and Jackson, 2010). For naive users, the structure of the platform-optimal algorithm is *similar first*: it orders others by expected similarity to the focal user. The resulting feed is a perfect echo chamber (Pariser, 2011). Because the implemented length is finite under any admissible continuation rule, feed positions concentrate on close copies as the platform's selection set grows, so the posterior variance fails to fall and learning vanishes as platform size grows large. Echo chambers thus arise from engagement incentives even when messages are truthful. For rational users, we show in a simplified two-type environment that the platform optimally *front-loads* same-type content to reduce early disagreement and implements a longer path that introduces cross-cutting signals later; among feeds that maximize implemented engagement, there is a front-loaded representative. Echo chambers are therefore not a mere artifact of behavioral naiveté.

The second main contribution consists on studying policy-relevant alternatives to personalization: the *reverse-chronological algorithm* and a minimal corrective, the *breaking-echo-chambers* algorithm. The reverse-chronological algorithm (our non-profiling benchmark, brought back to platforms through the EU's Digital Services Act) restores exposure diversity by randomizing order. For a fixed length, it converges to a wisdom-of-the-crowds variance benchmark, but it reduces within-platform utility when early disagreement is most costly and therefore shortens naive engagement; we derive the naive welfare comparison to the engagement-optimal feed and the rational user's stopping condition under

rational accounts of social conformity and herding (Bernheim, 1994; Chamley, 2004). See also Mosleh et al. (2021) for evidence on shared identity and conformity in online diffusion.

5 For example, Loomba et al. (2021) document that exposure to misinformation reduced vaccine acceptance in the U.S. and U.K. by about six percentage points.

random order. We then analyze a minimal breaking-echo-chambers tweak that inserts a maximally opposite account at the top of the platform-optimal algorithm for naives' feed (a *closest-first* feed) and leaves the rest intact. The insertion raises early disagreement (and can shorten engagement) but lowers posterior variance sharply; in large platforms, this modification beats closest-first whenever users place sufficiently high weight on learning.

The third main contribution is to show that personalization reshapes network effects and hence the scope for competition. Personalization allows finer matching as a platform grows. For rational users, implemented feeds increasingly front-load similarity of low conformity cost while still delivering enough later diversity, so expected utility rises with platform size. For naive users, the same force can reduce welfare: larger selection sets let the platform fill the top of the feed with close copies, eroding the informational value of additional posts; network effects need not arise. These forces bear directly on policy. Non-profiling defaults (reverse-chronological) temper echo chambers but rarely dominate engagement-optimal designs in overall welfare. A complementary market-design lever is *horizontal interoperability*: by sharing network effects across providers, rivalry shifts toward the non-interoperable dimension—the ranking algorithm itself—disciplining engagement-only designs and moving equilibrium feeds toward the utilitarian benchmark.
.

1.1 Related literature

Our paper connects the economics of recommendation and curation (e.g., Aridor et al., 2024) with social learning under correlated signals (Golub and Jackson, 2010) and conformity motives (Bernheim, 1994; Cialdini and Goldstein, 2004), but shifts the focus to the *timing* of exposure in personalized feeds. Unlike models in which platforms amplify misinformation to stimulate sharing (Acemoglu et al., 2023), our users truthfully communicate and value accuracy; echo chambers arise nonetheless because engagement incentives lead the platform to sequence like-minded content early, raising continuation among naive users and, in a two-type rational benchmark, lowering early disagreement at the cost of diversity. This temporal mechanism complements media-economics accounts of distortion through content selection or slant (Reuter and Zitzewitz, 2006; Ellman and Germano, 2009; Gentzkow and Shapiro, 2010; Abreu and Jeon, 2019; Kranton and McAdams, 2022): here the distortion is primarily in *when* rather than *what* users see. Closest to us, Acemoglu et al. (2023) study sequential sharing with reputational concerns and show that platforms benefit from homophily because it lowers scrutiny of misinformation; by contrast, we analyze simultaneous truthful posting, model learning explicitly via posterior precision, and derive echo chambers without misinformation or reputational motives. We are also related to models where profit-motivated platforms harm user welfare by complementing time spent or manipulating information flow (e.g., Mueller-Frank et al., 2022; Beknazar-Yuzbashev et al., 2024); we model continuation through a bounded-hazard,

addiction-driven process and separate within-platform utility from action utility to quantify learning losses.

Empirically, our predictions match empirical work showing reverse-chronological algorithms reduces usage while broadening exposure and weakening ideological clustering (Guess et al., 2023), consistent with evidence on algorithmic feeds and engagement (Kitchens et al., 2020; Gauthier et al., 2025) and with debates on "filter bubbles" (Pariser, 2011; Levy, 2021; Sunstein, 2017; Holtz et al., 2020). Finally, our policy counterfactuals relate to proposals that curb echo chambers through ranking changes rather than content moderation (Jackson et al., 2022; Guriev et al., 2023) and to market-level discussions of interoperability and contestability under the EU's Digital Markets Act (DMA) and DSA (Kades and Scott Morton, 2020; Bourreau and Krämer, 2022; Bourreau et al., 2023; Dhakar and Yan, 2024; Belleflamme and Peitz, 2020; Banchio et al., 2025).

The remainder of the paper proceeds as follows. Section 2 introduces the environment. Section 3 establishes truthtelling and characterizes platform-optimal personalization, showing when it generates echo-chamber exposure for naive and rational users. Section 4 quantifies the effects of personalization on learning and Section 5 evaluates two policy-relevant counterfactuals. Section 6 analyzes network effects and discusses the policy implications. Section 7 concludes.

2 A model of communication and learning through personalized feeds

We study a platform that curates personalized feeds and users who communicate and learn from them. The section first presents the model, then discusses and interprets the key assumptions.

Players. There is a social media platform p and a set, $\mathcal{U}$, of n users indexed i that visit it. We assume a complete network so that each user i's neighborhood is $N_i = \mathcal{U}\backslash\{i\}$. Expectations with respect to the platform's information are denoted $\mathbb{E}_p[.]$; user i's expectations are $\mathbb{E}_i[.]$.

Information structure. There is a state of the world, θ, for which we assume improper priors.[6] Each user i observes a private signal

$$\theta_i = \theta + \varepsilon_i, \qquad \varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)' \sim \mathcal{N}(0, \mathbf{\Sigma}),$$

where $\mathbf{\Sigma} = (\sigma_{ij})$ is symmetric positive definite and common knowledge. Under improper priors, $\theta \mid \theta_i \sim \mathcal{N}(\theta_i, \sigma_{ii})$, and $\theta_j \mid \theta_i \sim \mathcal{N}(\theta_i,\ \sigma_{ii} + \sigma_{jj} - 2\sigma_{ij})$ for all $j \neq i$ according to Lemma A.1.

6 Equivalently, one can take a normal (proper) prior with variance τ^2 and let $\tau^2 \to \infty$.

User choices. Each user i (i) posts a scalar message $m_i \in \mathbb{R}$, (ii) observes an endogenously determined number of feed posts $e_i \in \{1, \ldots, n\}$, and (iii) after reading such posts chooses an action $a_i \in \mathbb{R}$.

Platform choices. The platform designs an algorithm consisting of an assignment that, given a pair of users i, j, tells which position user j's message occupies in user i's feed. Hence, an *algorithm* $\mathscr{F} = (\mathscr{F}_i)_{i \in \mathcal{U}}$ is a collection of bijections $\mathscr{F}_i \in \text{Bij}\left(\{1, \ldots, n-1\},\, N_i\right)$. We interpret $\mathscr{F}_i(r)$ as the r-th position in i's feed. Given engagement e_i, the realized feed is

$$\mathscr{F}_i^{e_i} = \{\mathscr{F}_i(1), \ldots, \mathscr{F}_i(e_i))\}.$$

For notation purposes, we will assume that $\mathscr{F}_i(1) = i$ for every user.

Users preferences. Users derive utility from two streams: *within-the-platform utility* and *action utility*. Within-the-platform (instant) utility has three components: (i) a positive linear payoff coming from reading messages; (ii) *sincerity*: users dislike deviating from their own signals,[7] and (iii) *conformity*: disagreeing with others' opinions is taxing. Formally,

$$u_i(e_i, m_i, m_{-i}, \mathscr{F}_i, \theta_i) = \alpha e_i - \beta \overbrace{(\theta_i - m_i)^2}^{\text{Sincerity}} - (1-\beta) \overbrace{\sum_{j \in \mathscr{F}_i^{e_i}} (m_i - m_j)^2}^{\text{Conformity}}, \tag{1}$$

where $\alpha > 0$ and $\beta \in (0, 1)$. Total realized utility is the weighted sum of within-the-platform utility and action utility, which we define as the squared difference of the action a_i to the state of the world:

$$U_i(e_i, m_i, m_{-i}, a_i, \mathscr{F}_i, \theta_i, \theta) = \lambda\, u_i(e_i, m_i, m_{-i}, \mathscr{F}_i, \theta_i) - (1-\lambda) \overbrace{(a_i - \theta)^2}^{\text{Action utility}}. \tag{2}$$

User heterogeneity. Users differ in how they choose e_i.

- A fraction of users are *rational*, and choose (m_i, e_i, a_i) to maximize $\mathbb{E}_i[U_i]$.
- The remaining fraction of users are *naive*, and follow a myopic, addiction-driven sequential stopping rule: after reading k messages, a naive user i reads the next message with probability $g(u_i(k-1, m_i, m_{-i}, \mathscr{F}_i, \theta_i))$, where $g : \mathbb{R} \to [0, 1]$ is some continuous and increasing function such that $0 < \underline{g} < g(x) < \overline{g} < 1$ for some $\underline{g}$, $\overline{g}$ and all $x \in \mathbb{R}$, and exits otherwise.

Platform's revenues and information. The platform monetizes engagement. Given an increasing and positive function $\pi(.)$, the platform chooses $\mathscr{F}$ to maximize expected revenue

$$\max_{\mathscr{F}} \sum_{i=1}^{n} \mathbb{E}_p\Big[\pi(e_i)\Big].$$

7 Due to improper priors, $\mathbb{E}\big[(m_i - \theta)^2 \mid \theta_i\big] = (m_i - \theta_i)^2 + \sigma_{ii}$, so penalizing deviations from θ or from θ_i is equivalent up to an additive constant.

The platform observes user types (perfect personalization), knows $g(.)$ and $\mathbf{\Sigma}$, but not θ nor $\{\theta_i\}_{i=1}^n$. (All results extend if π depends on each user's type, for example, if there are different monetization rates for rational and for naive users.)

Timing. The game of communication and learning through personalized feeds described above consists of the following sequence of events:

1. The platform publicly commits to an algorithm $\mathscr{F}$.
2. Signals are realized; each user observes θ_i.
3. Each user i posts a message $m_i \in \mathbb{R}$.
4. Rational users choose e_i; naive users draw e_i from the sequential process above. Given $\mathscr{F}_i^{e_i}$, all users choose a_i.
5. The state of the world is revealed and payoffs are realized.

Equilibrium. The equilibrium concept is a version of Bayesian-Nash Equilibrium (BNE) that accounts for the behavior of naive users. A profile of strategies for users and the platform is an equilibrium if:

- For each naive i, (m_i, a_i) maximizes $\mathbb{E}_i[U_i]$ given the mechanical process for e_i and correct beliefs about others.
- For each rational j, (m_j, e_j, a_j) maximizes $\mathbb{E}_j[U_j]$ given correct beliefs about others.
- The platform's algorithm $\mathscr{F}$ maximizes $\sum_i \mathbb{E}_p[\pi(e_i)]$ given users' induced strategies and the naive users' engagement process.

2.1 Interpretation of the model

We conclude this section by examining the model's assumptions.

Improper priors. Users' prior distribution is uniform along $\mathbb{R}$. This is a tractable way to capture that users treat their own signal as locally "central".[8] Formally, $\mathbb{E}[\theta \mid \theta_i] = \theta_i$, so a user cannot diagnose whether her realization is extreme (Ross et al., 1977; Greene, 2004). We adopt this assumption for tractability. Under normal priors, we can only determine the users' optimal linear messaging strategies for naive users, but we cannot derive an explicit expression for the platform-optimal algorithm.

User heterogeneity and engagement. The empirical evidence (Hoong, 2021; Allcott et al., 2022) shows that a non-trivial share of users display self-control failures (present-biased, time inconsistent "keep-scrolling") on social media, deviating with their choices from the standard forward-looking benchmark. The engagement process we specify for naive users is precisely in line with these findings: continuation is addiction-driven and myopic (increases with instantaneous utility—dopamine kick). The fact that naives' engagement depends only on instantaneous utility formalizes an extreme present bias: while scrolling, the user heavily discounts the longer-run benefits of improved beliefs and action

8 For a general discussion of improper priors, see Hartigan (1983).

quality, placing near-zero weight on learning relative to immediate reward (Guriev et al., 2023). The boundedness $0 < \underline{g} \leq g(\cdot) \leq \overline{g} < 1$ matches the empirical fact that even very engaging (or very poor) content does not lead to infinite (or zero) sessions.

Perfect personalization. Platforms observe extremely rich traces of user behavior and can rapidly learn engagement propensities; we assume for this paper that personalization is perfect and our monopolist platform knows which type the user is.[9] In practice, large platforms come close to this benchmark via high-frequency interaction logs, look-alike modeling, and continuous A/B testing.

Platform's profits as a function of total engagement. Ads monetize exposure; more engagement yields more impressions. Modeling revenue as $\sum_i \pi(e_i)$ captures this directly. Allowing π to vary by type (e.g., naive users monetizing differently) leaves the analysis intact and only rescales the platform's objective.

3 Truthtelling and the Existence of Echo Chambers

This section establishes our first main result. For any algorithm $\mathscr{F}$ chosen by the platform, reporting the private signal—truthtelling—is an equilibrium. For rational users, this equilibrium is unique. For naives, in turn, it is the only equilibrium that arises for any specification of the engagement function g. We select truthtelling as the equilibrium of interest and stick to it from now on. As a consequence, once the platform commits to $\mathscr{F}$, every user i—naive or rational—behaves identically in equilibrium with respect to messaging: $m_i = \theta_i$. The platform can therefore influence behavior only through the composition and order of each user's feed, not through others' messages.

Proposition 3.1 (Truthtelling)**.** *For any algorithm $\mathscr{F}$, the profile $m_i^* = \theta_i$ for all $i \in \mathcal{U}$ is a Bayes–Nash equilibrium of the messaging game. Moreover, among rational users, truthtelling is the unique equilibrium.*

Proof. See Appendix A. □

The intuition behind truthful reporting is straightforward. Since users have improper priors, they lack an external anchor for their beliefs. Conditional on their own signal, they expect every other user's signal to coincide with theirs,

$$\mathbb{E}[\theta_j \mid \theta_i, \mathscr{F}] = \theta_i.$$

9 We assume a monopolist platform with all users aboard: arguably, most social media platforms in today's landscape are monopolists of their fields. For example, the Bundeskartellamt (the German competition protection authority) states in its case against Facebook (B6-22/16, "Facebook", p. 6): "The facts that competitors are exiting the market and there is a downward trend in the user-based market shares of remaining competitors indicate a market tipping process that will result in Facebook becoming a monopolist." (Franck and Peitz, 2023). See Garcia and Li (2024) for a discussion of monopoly platform strategy after market expansion.

This implies that, from user i's perspective, the platform's feed cannot shift first-order beliefs: when others report their signals truthfully, the expected content aligns with θ_i. In equilibrium, therefore, the best reply to everyone else reporting truthfully is to report truthfully as well.

Proposition A.2 shows that while certain specifications of the continuation rule g may sustain specific non-truthful equilibria, these depend on particular functional forms. Truthful reporting, in contrast, is the only equilibrium that holds for *any* function g satisfying our assumptions (i.e., continuous and increasing). In this sense, it is robust to how engagement reacts to within-platform utility. For this reason, we restrict attention to truthful reporting throughout the analysis: it is the only equilibrium independent of the exact form of g.

For rational users, the equilibrium is also unique. The first-order condition defining optimal messages generates a contraction mapping—each user's best response is a convex combination of her own signal and the expected reports of others. Iterating these best-response equations converges to a single fixed point, which coincides with $m_i = \theta_i$ for all i. Any deviation would require shifting expectations that are, by construction, pinned down by the user's own signal. As messages are truthful in equilibrium, the platform affects user i's payoffs only through the ordering of posts shown to i, i.e., through her feed $\mathscr{F}_i$. Under perfect personalization, this implies a separability property.

Corollary 3.2. *For the platform, maximizing aggregate expected engagement* $\sum_{i=1}^{n} \mathbb{E}_p[e_i]$ *is equivalent to maximizing* $\mathbb{E}_p[e_i]$ *for each user* i *separately.*

The platform can thus design, for each user type, an algorithm that maximizes that user's expected engagement. We denote by $\mathscr{P}$ the platform-optimal algorithm for naive users and by $\mathcal{P}$ the platform-optimal algorithm for rational users.

3.1 The Platform-Optimal Algorithm for Naive Users

Fix a naive user i. By Proposition 3.1, all messages equal their senders' signals. After $r-1$ messages, user i's within-platform utility is

$$\mathbb{E}_p[u_i(r-1, \theta_i, \theta_{-i}, \mathscr{F}, \theta_i)] = \mathbb{E}_p\left[\alpha(r-1) - (1-\beta) \sum_{j \in \mathscr{F}_i^{r-1}} (\theta_i - \theta_j)^2\right].$$

The (expected) probability that i stays to read the r-th message is $\mathbb{E}_p[g(u_i(r, \theta_i, \theta_{-i}, \mathscr{F}, \theta_i))]$. Since g is increasing and only the conformity term depends on the identity of the next user in the feed, the platform chooses, among the not-yet-shown accounts,[10] the j one

10 We use "user" and "account" interchangeably to denote the entities that post messages appearing in a feed.

minimizing the expected loss from conformity:[11]

$$j \in \arg \max_{l \in \mathcal{U} \setminus \mathscr{F}_i^{r-1}} \left\{ -\mathbb{E}_p\left[(\theta_i - \theta_l)^2\right] \right\}.$$

Iterating this choice yields the platform-optimal ranking $\mathscr{P}$ for naive user i:

$$\begin{aligned}
\mathscr{P}_i^1 &\in \arg\max_{j \in \mathcal{U}} \left\{ -\mathbb{E}_p\left[(\theta_i - \theta_j)^2\right] \right\}, \\
\mathscr{P}_i^2 &= \mathscr{P}_i^1 \cup \arg \max_{j \in \mathcal{U} \setminus \mathscr{P}_i^1} \left\{ -\mathbb{E}_p\left[(\theta_i - \theta_j)^2\right] \right\}, \\
&\vdots \\
\mathscr{P}_i^{e_i} &= \mathscr{P}_i^{e_i - 1} \cup \arg \max_{j \in \mathcal{U} \setminus \mathscr{P}_i^{e_i - 1}} \left\{ -\mathbb{E}_p\left[(\theta_i - \theta_j)^2\right] \right\}.
\end{aligned} \tag{3}$$

Proposition 3.3 (Naive Users Get Echo Chambers)**.** *In equilibrium, the platform chooses* $\mathscr{P}$ *as in* (3)*.*

Proof. See Appendix A. □

Thus, for each naive user i, the feed ranks others in reverse order of expected conformity loss with i. This is a *perfect* echo chamber in the sense of Pariser (2011). This might not be that surprising considering that naives' engagement depends on within-the-platform utility, and then the platform's goal is to minimize the loss on conformity.[12]

3.1.1 Naive Users' Optimal Action

Although the feed does not affect messages, it affects actions. Let $\boldsymbol{\Sigma}_{\mathscr{F}_i^{e_i}}$ denote the restriction of $\boldsymbol{\Sigma}$ to the accounts shown to i and $\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}$ the corresponding vector of signals.

Proposition 3.4. *For any algorithm* $\mathscr{F}$*, naive user* i*'s optimal action after reading* e_i *messages is*

$$a_i^* \;=\; \frac{\mathbb{1}\, \boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\, \boldsymbol{\theta}^{\,t}_{\mathscr{F}_i^{e_i}}}{\mathbb{1}\, \boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\, \mathbb{1}^{\,t}} \;=\; \mathbb{E}\left[\theta \;\middle|\; \boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}\right].$$

Proof. The action a_i^* minimizes $\mathbb{E}[(a_i - \theta)^2 \mid \boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}]$. The first-order condition yields the stated conditional expectation; see Lemma A.3. □

11 Because $u_i(r) = K - (1-\beta)(\theta_i - \theta_j)^2$ with K a constant independent of j, the random variables $u_i(r)$ are ordered by first–order stochastic dominance in the expected conformity loss. Since g is increasing, $\mathbb{E}_p[g(u_i)]$ preserves this order, so maximising $\mathbb{E}_p[g(u_i)]$ is equivalent to minimising $\mathbb{E}_p[(\theta_i - \theta_j)^2]$.

12 Note that a naïve user could, in principle, alter her message to influence engagement and thereby affect how much she learns. However, truthful reporting remains robust to such behavior: any deviation would distort the user's perceived similarity with others without improving expected utility.

3.2 The Platform-Optimal Algorithm for Rational Users

To show that echo chambers are not merely an artifact of the engagement process, we also consider a simplified population of rational users with two types (e.g., *democrats* and *republicans*) whose engagement depends on learning; even in this case, the platform may prefer to place same-type users first in a user's feed.

In equilibrium, a rational the user chooses an engagement level e_i^* that maximizes her expected utility given the algorithm $\mathscr{F}$,

$$e_i^* \in \arg\max_{e_i \geq 1} \ \mathbb{E}\Big[U_i(e_i, \theta_i, \theta_{-i}, \mathscr{F}) \,\Big|\, \theta_i, \mathscr{F}\Big].$$

Because the feed induced by the algorithm is ordered, scrolling further exposes the user to increasingly diverse content but also incurs higher disutility from disagreement. For large λ (low value of learning), the user stops early, consuming only same-type messages. When λ is small, the informational gain from reaching later, opposite-type content outweighs the discomfort cost, leading to a strictly positive interior optimum $e_i^* > 1$. In equilibrium, the platform exploits this forward-looking behavior by clustering similar messages at the top of the feed to keep the user engaged until she reaches the more informative, cross-cutting content that maximizes her overall expected utility.

Formally, the platform's problem is to select a feed that implements the highest possible engagement. Let $e_j^*(\mathscr{F}_j) \in \arg\max_{m \in \mathbb{N}} \mathbb{E}[U_j(m; \mathscr{F}_j)]$ be rational user j's best-response engagement under feed $\mathscr{F}_j$, and define the set of implementable engagement levels

$$K_j \ := \ \Big\{k \in \mathbb{N} : \exists\, \mathscr{F}_j \text{ such that } e_j^*(\mathscr{F}_j) = k\Big\}.$$

Proposition 3.5 (Platform-optimal algorithm for rationals)**.** *For each rational user j, the platform chooses a feed $\mathscr{F}_j^{k^*}$ that implements*

$$k^* \in \arg\max_{k \in K_j} k,$$

i.e., the largest engagement level that can be induced by some algorithm.

Proof. Fix any $k \in K_j$ and one implementing feed $\mathscr{F}_j^k$. Since the platform's expected payoff $\mathbb{E}_p[\pi(k)]$ is strictly increasing in k, it selects a feed attaining the largest element of K_j. By best response, $\mathbb{E}[U_j(k^*; \mathscr{F}_j^{k^*})] \geq \mathbb{E}[U_j(k; \mathscr{F}_j^{k^*})]$ for all $k \in \mathbb{N}$. □

Characterizing the platform-optimal algorithm for rational users, $\mathcal{P}$, in full generality is infeasible, as it requires closed forms for posterior variances under arbitrary correlation matrices. To isolate the core force, we analyze a two-bloc environment composed by *democrats* (D) and *republicans* (R) that captures the same tradeoff. Consider a focal democrat user j. Other users are either same-type (D) with correlation $\rho_{jk} = x \in (0, 1)$ or opposite-type (R) with $\rho_{j\ell} = -x$. Assume α is small enough so that each additional

message creates a conformity cost even when types coincide.[13]

A (finite) feed is an ordered list $\mathscr{F} = (j_1, \ldots, j_n)$; its prefix of length k is $\mathscr{F}^k = (j_1, \ldots, j_k)$, with $j \leq n$. Let $d(\mathscr{F}, k)$ and $r(\mathscr{F}, k)$ denote the numbers of same-type (D) and opposite-type (R) accounts in $\mathscr{F}^k$, respectively (so $d + r = k$). We will have $d(\mathscr{F}, 1) = 1$ because the first user in the feed is the user herself, a democrat. The user's expected utility after reading k messages depends on counts (of republicans and democrats), not on order:

$$U(\mathscr{F}, k) := U(d, r) = \lambda\Big[\alpha(d+r) - (1-\beta)\Big(d(1-x) + r(1+x)\Big)\Big] - (1-\lambda)\,\mathrm{Var}[\theta|d, r]. \quad (4)$$

The posterior variance under the two-bloc structure is order-irrelevant and has the closed form

$$\mathrm{Var}[\theta \mid d, r] = \sigma^2 \frac{(1-x)\Big(1 + x(d+r-1)\Big)}{(1-x)(d+r) + 4xdr}. \quad (5)$$

The platform evaluates an algorithm $\mathscr{F}$ by the engagement it implements, denoted by $e^*(\mathscr{F}) = \arg\max_k U(\mathscr{F}, k)$ (largest maximizer). Among all algorithms, the platform chooses $\mathcal{P} \in \arg\max_{\mathscr{F}} e^*(\mathscr{F})$. For reference, the user optimal algorithm is denoted by $\mathscr{F}^U \in \arg\max_{\mathscr{F}} U(\mathscr{F}, e^*(\mathscr{F}))$. Given a pair (d, r), we can compare the utility induced by adding a democrat to the pool (abusing notation, $U(d+1, r)$) to the utility induced by adding a republican ($U(d, r+1)$):

$$U(d+1, r) - U(d, r+1) = 2\lambda(1-\beta)x - (1-\lambda)\Big(\mathrm{Var}[d, r+1] - \mathrm{Var}[d+1, r]\Big).$$

Moreover,

$$\mathrm{Var}[\theta \mid d, r{+}1] - \mathrm{Var}[\theta \mid d{+}1, r] = \\ -\frac{4\sigma^2 x(1-x)(d-r)\Big(1 + x(d+r)\Big)}{\Big(4xdr - dx + d + 3rx + r - x + 1\Big)\Big(4xdr + 3dx + d - rx + r - x + 1\Big)}. \quad (6)$$

In particular, the learning gain from the marginal opposite-type signal is (weakly) larger when the user currently observes more same-type signals: if $d > r$ then $\mathrm{Var}[\theta|(d, r+1)] - \mathrm{Var}[\theta|(d+1, r)] > 0$, while it is zero at $d = r$ and negative if $d < r$.

The user–optimal algorithm, denoted by $\mathscr{F}^U$, satisfies two key properties. First, it is order independent: given $\mathscr{F}^U$, the engagement level $e_i^*(\mathscr{F}^U)$ maximizes the user's expected utility not only relative to other possible algorithms, but also relative to any reordering of the same content. In other words, once the set of messages in $\mathscr{F}^U(e_i^*)$ is fixed, their order does not affect the user's utility. Second, $\mathscr{F}^U$ contains a higher proportion of same–type than opposite–type accounts. Because the optimal learning profile combines exposure to both types, with the proportion of democrat and republican

13 This assumption only simplifies the exposition and does not affect the comparative statics emphasized below.

messages satisfying $r \in [d-1, d+1]$, the algorithm that maximizes learning balances confirmation and diversity. In the limiting case where the user cares only about learning ($\lambda = 0$), this balanced composition is precisely the user–optimal algorithm. As there is an extra cost associated with an opposing view signal, this tilts the balance towards more democrats than republicans in the feed. All the following results are stated for a democrat user; the republican version is similar.

Proposition 3.6 (Echo chambers arise even for rational users)**.** *Given an algorithm $\mathscr{F}$ and a level of engagement $e^*(\mathscr{F})$, let us assume that the number of republicans in a democrat user feed is larger than the number of democrats, i.e., $r(\mathscr{F}, e^*(\mathscr{F})) > d(\mathscr{F}, e^*(\mathscr{F}))$. Then, the platform can weakly increase engagement by choosing an algorithm $\mathscr{F}'$ showing more democrats than republicans in the democrat feed.*

Proof. Suppose, for contradiction, that $\mathscr{F}$ is platform-optimal with $r(\mathscr{F}, e^*(\mathscr{F})) > d(\mathscr{F}, e^*(\mathscr{F}))$. Since the focal user is a democrat, there exists $k < e^*(\mathscr{F})$ with $d(\mathscr{F}, k) = r(\mathscr{F}, k)$; let k_M be the largest such k. Construct $\mathscr{F}'$ by keeping the first k_M positions identical to $\mathscr{F}$ and flipping each label thereafter.

For any $t \leq k_M$, the prefixes are identical, so $U(\mathscr{F}', t) = U(\mathscr{F}, t)$. For $t > k_M$, write (d_t, r_t) for the counts under $\mathscr{F}$ in the first t positions. By construction, the counts under $\mathscr{F}'$ are (r_t, d_t) (because the increments after k_M are swapped). Using $\mathrm{Var}[\theta|d, r] = \mathrm{Var}[\theta|r, d]$ and

$$d(1-x) + r(1+x) = (d+r) + x(r-d),$$

we have

$$U(\mathscr{F}', t) - U(\mathscr{F}, t) = 2\lambda(1-\beta)\, x\, (r_t - d_t) \;\geq 0,$$

since $r_t \geq d_t$ for all $t \in \{k_M + 1, \ldots, e^*(\mathscr{F})\}$ by the choice of k_M and the assumption $r(\mathscr{F}, e^*(\mathscr{F})) > d(\mathscr{F}, e^*(\mathscr{F}))$. Hence $U(\mathscr{F}', t) \geq U(\mathscr{F}, t)$ for all t, with equality up to k_M. Therefore $\max_t U(\mathscr{F}', t) \geq \max_t U(\mathscr{F}, t)$ and $e^*(\mathscr{F}') \geq e^*(\mathscr{F})$. □

Even when users are rational, personalization brings into their feeds less diversity than they would randomly get in their neighborhood. Moreover, platform-optimal algorithms for rationals are front-loaded with same-type users. Thus, not only naive users receive echo chambers as their feed.

Proposition 3.7 (Front-loading for rationals)**.** *Fix user j and let $\mathcal{F}^\star$ be the set of feeds that maximize e_j^*. Then there exists $\mathscr{F}_j^\star \in \mathcal{F}_j^\star$ and some $k \in \{1, \ldots, e_j^*\}$ such that the first k positions of $\mathscr{F}^\star$ are democrats and the next $e_j^* - k$ are republicans.*

Proof. See the proof in the Appendix A □

The platform exploits the user's willingness to learn by front-loading low-conformity-cost, same-type content to keep the user engaged until the more informative opposite-type signals arrive. Because the learning benefit of an opposite-type signal is increasing in the stock of same-type signals (Equation (6)), a front-loaded feed can implement strictly

higher engagement than a balanced or alternating one. When learning matters more (small λ), the platform can tilt the implemented prefix further toward same-type content. Conversely, if willingness to learn is low (large λ), engagement is scarce and multiple platform-optimal algorithms may exist, including ones that feature less pronounced front-loading.

4 The Effects of Personalization on Learning

In this section, we evaluate how the platform-optimal algorithm affects learning both for naives and for rationals, i.e., how information acquired on the platform improves decision quality. First of all, let us assume homogeneous signal variances for tractability: $\sigma_{ii} = \sigma_{jj} =: \sigma^2$ for all $i, j \in \mathcal{U}$. Now, we define learning as the reduction in the expected squared error of the optimal action after reading messages. When user i chooses her optimal action after e_i messages, the expected loss equals the posterior variance:

$$\mathbb{E}\Big[(a_i - \theta)^2 \mid \boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}\Big] = \mathbb{E}\Big[\Big(\mathbb{E}[\theta \mid \boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}] - \theta\Big)^2 \mid \boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}\Big] = \mathrm{Var}\Big[\theta \mid \boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}\Big].$$

By Lemma A.3, this posterior variance admits the closed form

$$\mathrm{Var}\Big[\theta \mid \boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}\Big] = \frac{1}{\mathbb{1}\,\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\,\mathbb{1}^t}, \tag{7}$$

which allows us to compute learning for any algorithm and any $(\mathcal{U}, \boldsymbol{\Sigma})$. In particular, $\mathrm{Var}[\theta \mid \boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}] \leq \sigma^2$, so reading a feed weakly improves decision quality.

Under the platform-optimal algorithm for naives, the feed is ordered by similarity to minimize conformity losses. This ordering (weakly) raises engagement relative to the user–optimal ordering by front-loading like-minded content. Two forces determine learning. First, a higher engagement level k delivers more signals, which—holding composition and correlations fixed—reduces the posterior variance. Second, the way the platform raises k is by increasing similarity within the prefix, which lowers signal diversity and might dampen information. The net effect on learning for naives is therefore a priori ambiguous. These forces depend critically on the platform's selection set. With a small pool of available accounts (and a given $\boldsymbol{\Sigma}$), the platform's ability to shape both engagement and the composition of any prefix is limited. As the platform becomes large, the feed can be engineered more flexibly at each depth, and the dependence of learning on the specific covariance structure weakens. In the next subsection we formalize this "large platform" regime, but, before, let us analyze learning for rational users when platform size is fixed.

For rational users, platform and user objectives need not be aligned in terms of learning. By designing a sequence that induces a larger best-response engagement k—for instance, front-loading lower-cost same-type content so that the user keeps scrolling until opposite-type signals arrive—the platform can push the user beyond her stopping rule under $\mathscr{F}^U$. When learning is sufficiently valuable (small λ) and processing costs for dis-

sonant content are moderate, the additional exposure to opposite signals strictly reduces the posterior variance, potentially yielding higher learning than under $\mathscr{F}^U$. Intuitively, the marginal informational value of an opposite signal is largest when such signals are scarce in the user's history; by increasing k, the platform ensures enough disagreement to improve learning overall.

4.1 The Effects on Learning in Large Platforms

Motivated by the recent growth of social media usage,[14] we study learning as platform size grows. We show that, for naive users, learning vanishes in the limit. The mechanism is the echo chamber: as the selection set expands, the platform fills the user's finite implemented prefix with close copies of the focal user, so additional messages convey little extra information about θ. This contrasts sharply with classical *wisdom-of-the-crowd* results, highlighting the platform's strategic role in information design.

We define the expansion protocol as follows. Starting from a given user base $\mathcal{U}$, we add entrants whose covariances with incumbents are drawn from a continuous, symmetric distribution centered at zero with support on $[-\sigma^2, \sigma^2]$. The resulting covariance matrix $\mathbf{\Sigma}$ is symmetric and positive definite, ensuring a well-defined correlation structure across users. To formalize this argument, we assume that the pairwise correlations ρ_{ij}, $j \neq i$, are independent and identically distributed draws from a continuous distribution F on $[-1, 1]$ with strictly positive density $f(x) > 0$ for all $x \in [-1, 1]$.

4.1.1 Learning for Naives in Large Platforms

As platform size increases, the platform-optimal algorithm $\mathscr{P}$ selects the most similar neighbors to maximize naive engagement. The expected implemented length is nevertheless finite: because $g(\cdot) \in (0, 1)$, continuation is never guaranteed and the probability of an infinite reading path is zero. The following lemma formalizes bounded expected engagement.

Lemma 4.1. *There exist well-defined $k_{pi}, k_u \in \mathbb{N}$ such that $\mathbb{E}_p[e_i] \leq k_{pi}$ and $\mathbb{E}_i[e_i] \leq k_u$. In particular, there exists k_p with $\mathbb{E}_p[e_i] \leq k_p$ for all $i \in \mathcal{U}$.*

Proof. See Appendix A. □

Bounded implemented length implies that, as the platform grows, the naive user's observed prefix under $\mathscr{F}$ contains ever more highly correlated accounts. Consequently, diversity collapses within the prefix and learning disappears asymptotically.

Proposition 4.2. *Under the platform-optimal algorithm $\mathscr{F}$, user i's learning becomes negligible as $n \to \infty$:*

$$\operatorname*{plim}_{n\to\infty} Var\Big[\theta \mid \boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}\Big] = \sigma^2.$$

14 See, for example, the number of social media users from 2011 to 2028 (forecasted): https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/.

Proof. See Appendix A. □

This result is in stark contrast with classic environments where the wisdom of the crowd enhances learning as the population grows. Here, the platform's strategic feed design undermines diversity, creating the echo chambers and filter bubbles emphasized by Pariser (2011). The policy relevance is immediate. For instance, the DSA's requirement that platforms offer a non-profiling ranking has led to the reinstatement of reverse-chronological feeds. Section 5 analyzes such alternatives relative to personalization.

4.1.2 Learning for Rational Users in Large Platforms

In contrast to naives, rational users need not experience vanishing learning as the platform expands. Consider the expansion protocol above. For each entrant ℓ, if the correlation $\rho_{j\ell}$ is sufficiently high so that the within-the-platform benefit (scaled by α) outweighs conformity costs, including ℓ at the front of the feed weakly increases user j's expected utility while strictly reducing her posterior variance. Because the platform maximizes engagement, it will front-load all such nonnegative-marginal entrants; the rational user then reads them along the equilibrium path. Hence, as the set of utility-improving entrants grows with platform size, the rational user's posterior variance weakly decreases. In this sense, a large platform can generate a within-platform wisdom-of-the-crowd effect for rational users who can learn without incurring net conformity costs.

5 The reverse-chronological algorithm and other alternatives

This section evaluates ranking rules that can replace or modify the engagement-maximizing algorithms studied above. We consider two families that are especially salient in current debates. The first is the *reverse-chronological* algorithm, the canonical "non-profiling" option contemplated by the DSA. The second is a class of minimal *corrective* modifications to the platform-optimal algorithms that reintroduce diversity in exposure while preserving most of their engagement properties.

To build intuition for the subsequent results, we analyze in more detail how $\mathrm{Var}\left[\theta \mid \boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}\right]$ depends on the covariance structure $\boldsymbol{\Sigma}$. Recall that with homogeneous signal variances ($\sigma_{ii} = \sigma_{jj}$ for all i, j), the posterior variance after reading a prefix $\mathscr{F}_i^{e_i}$ admits the closed form

$$\mathrm{Var}\left[\theta \mid \boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}\right] = \frac{1}{\mathbb{1}\,\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\,\mathbb{1}^t}, \tag{8}$$

(see Equation (7)). Let $\boldsymbol{X} = \boldsymbol{\Sigma}^{-1}$ be the precision matrix. The partial correlation between θ_i and θ_j conditional on all other signals equals $-\frac{x_{ij}}{\sqrt{x_{ii}x_{jj}}}$. Hence, the informational value of adding user j to i's feed depends not only on the pairwise covariance σ_{ij}, but also on the pattern of *conditional* relationships among all accounts already in the prefix.

Differentiating $\text{Var} = (1^\top \Sigma_F^{-1} 1)^{-1}$ with respect to an off-diagonal covariance σ_{ij} yields

$$\frac{\partial}{\partial \sigma_{ij}} \text{Var}[\theta \mid \boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}] = 2 \left(\frac{1}{\mathbb{1}\, \boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\, \mathbb{1}^t} \right)^2 \left(\sum_\ell x_{i\ell} \right) \left(\sum_\ell x_{j\ell} \right),$$

so the effect of bringing in a more correlated user j is generally non-monotone: it increases variance when the precision row-sums $s_i = \sum_\ell x_{i\ell}$ and $s_j = \sum_\ell x_{j\ell}$ have the same sign, and decreases it when they have opposite signs. This non-monotonicity explains why, even though random exposure, and thus an increase in diversity, the effect on learning need not always be positive.

The phenomenon is illustrated in the next example. Consider a small network of four individuals ($n = 4$), and assume that $e_i = 3$ and that the distribution of signals, conditional on θ, is

$$\begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{pmatrix} \sim \mathcal{N} \left(\boldsymbol{\theta},\, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.8 & 0.7 & 0.5 \\ 0.8 & 1 & 0.3 & 0.6 \\ 0.7 & 0.3 & 1 & 0.4 \\ 0.5 & 0.6 & 0.4 & 1 \end{bmatrix} \right).$$

Define two possible feeds for user 1: $\mathscr{F}_1^{e_i} = \{1, 2, 3\}$, which includes the most correlated users, and $\mathscr{F}_2^{e_i} = \{1, 3, 4\}$, which features less correlated ones. The corresponding posterior variances are $\text{Var}[\theta \mid \mathscr{F}_1^{e_i}] = 0.58$ and $\text{Var}[\theta \mid \mathscr{F}_2^{e_i}] = 0.68$. Somewhat counterintuitively, $\mathscr{F}_1^{e_i}$—the more correlated feed—yields better learning. This illustrates that the covariance between user 1 and each of her peers is not the sole determinant of learning: the conditional dependencies among all users within the feed also shape the amount of information the user can extract.

5.1 Reverse-chronological (random) algorithm

Before the adoption of personalized feeds, most social media platforms displayed posts strictly in the order they were written, with the most recent appearing first. In the model, this corresponds to an algorithm that draws the next message uniformly at random from the remaining accounts, the *reverse-chronological algorithm*, denoted $\mathscr{R}$. Under $\mathscr{R}$, the next post is drawn uniformly from the remaining accounts. Conditional on $U \setminus \{i\}$, each neighbor j is equally likely to appear at any depth; the order is independent of the covariance structure. Unlike $\mathscr{P}$ or $\mathcal{P}$, $\mathscr{R}$ neither amplifies homophily nor minimizes disagreement in early positions. Its salient property is that it restores *exposure diversity*.

Diversity improves learning in large platforms. With homogeneous signal variances and a fixed implemented length e_i, a law-of-large-numbers argument implies

$$\mathbb{E}\Big[\operatorname*{plim}_{n\to\infty} \text{Var}\big[\theta \mid \mathscr{R}_i^{e_i}\big] \Big] \;=\; \sigma^2\, \mathbb{E}\Big[\frac{1}{e_i} \,\Big|\, \mathscr{R} \Big], \tag{9}$$

and in the large-population limit $\operatorname{plim}_{n\to\infty} \operatorname{Var}[\theta \mid \mathscr{R}_i^{e_i}] = \frac{\sigma^2}{e_i}$. Thus, conditional on reading e_i posts, $\mathscr{R}$ eventually delivers the classical wisdom-of-crowds benchmark. The cost is within-platform utility: random early exposure includes dissimilar posts precisely when continuation is most fragile, lowering engagement and platform revenue.

For naive users, $\mathscr{R}$ trades off learning gains from diversity against reduced gratification and higher disagreement costs in the early prefix. The next result compares $\mathscr{P}$ and $\mathscr{R}$ in large platforms (under the expansion protocol introduced above).

Proposition 5.1 (Naive users: $\mathscr{P}$ versus $\mathscr{R}$)**.** *Given the covariance structure* $\mathbf{\Sigma}$ *and for a large platform size, the platform–optimal algorithm* $\mathscr{P}$ *outperforms the reverse–chronological algorithm* $\mathscr{R}$ *in expected utility if and only if*

$$\lambda > \max_{i\in\mathcal{U}} \left\{ \frac{1 - \mathbb{E}\left[\frac{1}{e_i} \mid \mathscr{R}\right]}{\frac{\alpha}{\sigma^2}\left(\mathbb{E}[e_i \mid \mathscr{P}] - \mathbb{E}[e_i \mid \mathscr{R}]\right) + 2(1-\beta)\,\mathbb{E}[\,e_i - 1 \mid \mathscr{R}] + \left(1 - \mathbb{E}\left[\frac{1}{e_i} \mid \mathscr{R}\right]\right)} \right\}.$$

Proof. See Appendix A. □

Because $e_i \geq 1$ and $\mathscr{P}$ maximizes expected engagement, there is always some $\lambda \in (0,1)$ for which $\mathscr{P}$ dominates $\mathscr{R}$; in many configurations, $\mathscr{P}$ wins for most λ (see Figure **??**). Intuitively, the learning edge of $\mathscr{R}$ materializes only when users scroll sufficiently long; but random early disagreement lowers continuation, so naives consume fewer posts precisely when diversity would pay off.

Let us now characterize rational users' behavior under the reverse-chronological algorithm. Since the ranking imposed by $\mathscr{R}$ draws the next message uniformly at random from the remaining accounts, the rational user's decision at any depth k reduces to a standard optimal stopping problem: whether to continue reading one more randomly selected post or to stop and act based on the information already acquired. The user weighs the expected instantaneous benefit of continuing against the expected conformity cost generated by disagreement with the next, randomly drawn message. The following proposition formalizes this condition.

Proposition 5.2 (Rational users optimal behavior under $\mathscr{R}$)**.** *Consider a rational user at depth* k *who has already read the set of messages* $\mathscr{R}_i^k$*. Then it is optimal to continue to* $k{+}1$ *if and only if*

$$\lambda\,\alpha \;+\; (1-\lambda)\sigma^2\Big(\operatorname{Var}(\theta \mid \mathscr{R}_i^k) - \mathbb{E}\Big[\operatorname{Var}(\theta \mid \mathscr{R}_i^{k+1})\Big]\Big) \;>\; \frac{\lambda(1-\beta)}{n{-}k{-}1}\,\mathbb{E}\Big[(\theta_j - \theta_i)^2 \,\Big|\, \mathscr{R}_i^k\Big],$$

where the expectations are taken over the next random user j *among the* $n{-}k{-}1$ *accounts that have not yet appeared in the user's feed.*

Proof. See proof in Appendix A □

The takeaway here is that the reverse-chronological benchmark restores exposure diversity and attains crowd-wisdom asymptotically for a fixed e_i, but it sacrifices within-platform

utility and (for naives) implemented length. In practice, we have that $\mathscr{R}$ will rarely dominate engagement-optimal designs in overall welfare. This motivates hybrid designs that combine diversity with the continuation benefits of targeted similarity.

5.2 A minimal corrective: the breaking-echo-chambers algorithm

Several platforms have experimented with tools that surface corrective or contextual content (e.g., community notes, sponsored public-interest messages). We capture this idea with a minimal modification of the platform-optimal algorithm: the *breaking-echo-chambers* algorithm $\mathscr{B}$ inserts at the very top the account most negatively correlated with the focal user and then follows the platform-optimal order. Formally, for every $i \in \mathcal{U}$,

$$\mathscr{B}_i(1) \in \arg\min_{j \in \mathcal{U}\setminus\{i\}} \rho_{ij}, \qquad \mathscr{B}_i(k) = \mathscr{P}_i(k-1) \;\text{ for } k \geq 2.$$

In large platforms, the top insertion can be made nearly maximally opposite at the expense of a small conformity cost, while it sharply reduces posterior variance. The engagement effect is a small reduction in the probability of continuing from the first position. The net welfare comparison for naives is summarized next.

Proposition 5.3 (Naives' welfare comparison $\mathscr{P}$ vs $\mathscr{B}$)**.** *When platform size grows large, the above defined breaking-echo-chambers algorithm outperforms the platform-optimal algorithm $\mathscr{P}$ for user i if and only if*

$$\lambda \;\leq\; \frac{1}{1 + \frac{\alpha}{\sigma^2}\Big(\mathbb{E}[e_i \mid \mathscr{P}] - \mathbb{E}[e_i \mid \mathscr{B}]\Big)}.$$

Proof. See Appendix A. □

Thus, when users place sufficient weight on learning (small λ), the informational gain from a single early opposite post more than compensates for the small engagement loss relative to $\mathscr{P}$. In finite platforms the comparison is ambiguous: if a strongly opposite account exists, $\mathscr{B}$ delivers sizable learning improvements at modest conformity and engagement costs; otherwise $\mathscr{P}$ remains preferable.

For rational users, $\mathscr{B}$ typically brings limited benefits. Under $\mathcal{P}$ the platform already implements a balanced path by front-loading low-cost sameness to secure continuation and delivering cross-cutting content later; adding an opposite post at the top mainly lowers early gratification and may shorten engagement without providing commensurate learning gains.

5.3 Policy discussion

The DSA requires very large online platforms to offer a non-profiling algorithm. In practice, platforms satisfied this by (re)introducing reverse-chronological algorithms. In our framework, this corresponds to $\mathscr{R}$. Its effects are clear. As we have shown, on the one hand, $\mathscr{R}$ restores exposure diversity and, for any fixed implemented length, converges to the classical learning benchmark. On the other, early random disagreement lowers within-the-platform utility, shortens engagement for naive users, and typically yields lower welfare than engagement-optimal algorithms that time diversity. This mirrors the evidence that personalization was built to maximize engagement—and has succeeded (Guess et al., 2023): the platform-optimal $\mathscr{P}$ is tuned to deliver immediate gratification.

Hence $\mathscr{R}$ is unlikely to outperform $\mathscr{P}$ in practice. Rational users face a sharp trade-off: random exposure does not guarantee higher learning along the realized path and imposes conformity costs when continuation is most fragile. The reverse-chronological algorithm $\mathscr{R}$ thus fulfills the letter of the DSA while falling short as a welfare substitute for personalization.

Platforms have also tried corrective designs. Twitter (now $\mathbb{X}$) introduced Birdwatch (Community Notes) in January 2021 to crowd-source context. In our model, the breaking-echo-chambers tweak $\mathscr{B}$ captures this logic: inserting a maximally opposite post on top of the closest-first feed restores near first-best learning in large platforms at a small conformity cost (coming only from one user), while preserving the rest of $\mathscr{P}$. Implementation is the hurdle: it needs enforcement or platform cooperation, and users may ignore inserted opposite content.

Taken together, $\mathscr{R}$ and $\mathscr{B}$ do not fully resolve the engagement-learning tension. This motivates the market–design approach pursued next: under horizontal interoperability, network effects are shared and platforms compete on ranking quality; under some conditions, the utilitarian-optimal algorithm $\mathscr{U}$ becomes the implemented option by competing platforms.

6 Network Effects

This section analyzes how personalization interacts with network effects—how a larger user base amplifies engagement, reshapes exposure diversity and learning, and alters market structure and policy levers such as entry and interoperability. We say that user i experiences network effects under algorithm $\mathscr{F}$ if her expected utility weakly increases with platform size: $\mathbb{E}_i[U_i \mid n{+}1, \mathscr{F}(n{+}1)] \geq \mathbb{E}_i[U_i \mid n, \mathscr{F}(n)]$. For clarity, we write $\mathscr{F}(n)$ for the algorithm applied on a platform of size n, and $e_i(n)$ for the engagement induced for user i under $\mathscr{F}(n)$.

Proposition 6.1 (Network effects need not hold for naives)**.** *Fix a naive user i. Under*

the platform-optimal algorithm $\mathscr{P}$ define

$$\Delta e_i := \mathbb{E}\Big[e_i(n{+}1) - e_i(n) \,\Big|\, \theta_i, \mathscr{P}(n)\Big],$$

$$\Delta C_i := \mathbb{E}\left[\sum_{j \in \mathscr{P}_i^{e_i(n+1)}(n+1)} (\theta_i - \theta_j)^2 - \sum_{j \in \mathscr{P}_i^{e_i(n)}(n)} (\theta_i - \theta_j)^2 \,\middle|\, \theta_i, \mathscr{P}(n)\right],$$

$$\Delta V_i := \mathbb{E}\Big[Var\Big(\theta \,\Big|\, \boldsymbol{\theta}_{\mathscr{P}(n+1)}\Big) - Var\Big(\theta \,\Big|\, \boldsymbol{\theta}_{\mathscr{P}(n)}\Big) \,\Big|\, \theta_i, \mathscr{P}(n)\Big].$$

Whenever $\alpha\,\Delta e_i - (1-\beta)\,\Delta C_i + \Delta V_i \neq 0$, the platform-optimal algorithm $\mathscr{P}$ features network effects for naive user i if and only if

$$\lambda \;\geq\; \frac{\Delta V_i}{\alpha\,\Delta e_i - (1-\beta)\,\Delta C_i + \Delta V_i}.$$

If either $\alpha\,\Delta e_i - (1-\beta)\,\Delta C_i + \Delta V_i = 0$ or learning improves with platform size (i.e., $\Delta V_i \leq 0$), then $\mathscr{P}$ features network effects for naive user i for every $\lambda \in (0,1)$.

Proof. See Appendix A. □

Proposition 6.2 (Rational users always enjoy network effects)**.** *Under the platform-optimal algorithm $\mathcal{P}$, rational users always enjoy network effects.*

Proof. See Appendix A. □

The propositions formalize a sharp contrast. Personalization enlarges the platform's selection set and permits finer matching in the feed. For rational users, who internalize both conformity and learning when deciding how far to scroll, this translates into higher expected utility as n grows: the platform can front-load low-cost similar content while still implementing enough exposure to disagreement to sustain learning. By contrast, for naive users the platform over-exploits homophily: as n rises, the implemented feed becomes increasingly populated with close copies of the focal user, which raises conformity and lowers the informational value of each additional message. When the naive user places any weight on learning ($\lambda > 0$), the sign and magnitude of ΔV_i become pivotal; beyond a size threshold the learning loss can dominate, so expected utility can fall with n even if engagement rises.

This aligns with the evidence from Guess et al. (2023), which shows that switching users from Facebook's personalized feed to a chronological (non-personalized) feed significantly reduces engagement but also exposes users to more diverse content, mitigating echo chambers. This might be happening in 𝕏 and other social media platforms (Kitchens et al., 2020; Gauthier et al., 2025). These patterns suggest that naive users on highly personalized platforms may not realize the extent of the algorithm's filter bubble, continuing to engage more while actually seeing less variety in viewpoints. Such evidence underscores our model's implication that personalization, when taken to extremes, can erode the very network benefits that attracted users in the first place.

6.1 Network effects, competition, and personalization algorithms

Although our model does not endogenize platform competition, the network–effect results speak directly to current policy debates. Two observations are central. First, the presence of naive users challenges the conventional view that larger platforms unambiguously benefit participants in digital markets. For such users, stronger personalization may raise engagement while eroding learning, so scale can reduce welfare. This suggests that improving contestability in social-media markets may require tools beyond the standard repertoire (Banchio et al., 2025). Second, when user behavior does generate platform-specific network effects, the familiar winner-takes-most logic applies: large incumbents enjoy advantages that are difficult for entrants to overturn.

A leading proposal to address these frictions is horizontal interoperability—the ability of distinct platforms to interconnect so that users from one can interact with users from the other and *vice versa*(Kades and Scott Morton, 2020). Interoperability makes network effects no longer platform-specific, but shared across the whole market. Competition then shifts from for the market to within the market (Belleflamme and Peitz, 2020). This is standard in other communication layers (e.g., telephony and email), where users can communicate across providers without switching accounts.[15]

The EU's DMA is a first step in this direction. Article 7 mandates interoperability for number-independent interpersonal communication (messaging) services provided by gatekeepers. While some authors are skeptical about the magnitude of the resulting competitive gains in messaging (Bourreau and Kraemer, 2023; Bourreau et al., 2023; Dhakar and Yan, 2024), social media differs in a critical respect: personalization algorithms are the key competitive dimension once network effects are shared. In that environment, interoperability weakens lock-in and redirects rivalry toward the quality of feed design. Our results imply that such a shift would particularly benefit naive users: with interoperability, platforms cannot rely on captive scale and must compete on algorithms that balance engagement with learning. Rather than prescribing a specific ranking rule, interoperability harnesses competitive pressure to discipline engagement-maximizing designs that would otherwise entrench echo chambers.

7 Conclusion

This paper develops a tractable model of communication and learning through personalized feeds to study how an engagement-maximizing platform arranges exposure and how that arrangement feeds back into behavior. Two simple primitives—truthful messaging and correlated signals—deliver sharp implications. First, truthtelling is the robust equilibrium of the messaging game, so platforms exercise power solely through the composition and order of exposure. Second, the engagement-optimal ranking is

15 For example, a Yahoo user can seamlessly send an email to a Gmail user, and *vice versa*.

homophily-preserving: naives receive closest-first feeds (a perfect echo chamber), and rationals receive front-loaded sameness that sustains continuation until cross-cutting content arrives. In large platforms the naive learning benefit of additional scrolling vanishes, breaking the classical wisdom-of-the-crowds logic; rational users, by contrast, continue to enjoy network effects.

These positive results organize the welfare comparison among realistic ranking regimes. A reverse-chronological (non-profiling) feed restores exposure diversity and achieves the crowd-wisdom variance benchmark for a fixed length, but it reduces within-platform utility when early disagreement is most costly and hence curtails naive engagement. A minimal corrective—the insertion of a single opposite-type post at the top of a closest-first feed—creates a clean trade-off: it raises early disagreement (and may shorten engagement) but sharply lowers posterior variance; when users place sufficient weight on learning, the corrective dominates in expected utility. In short, timing of diversity matters as much as its level.

With respect to policy analysis, non-profiling defaults are a natural baseline but are not a panacea: absent additional design features, they underperform engagement-optimal personalization when users heavily value within-platform experience and rarely deliver first-best learning along realized paths. Corrective insertions or contextualization tools can move outcomes toward the utilitarian benchmark at modest engagement cost, especially on large platforms with rich selection sets. More broadly, competition policy that shares network effects—horizontal interoperability—redirects rivalry toward algorithmic quality, disciplining pure engagement maximization and aligning platform incentives with user welfare.

Two modeling choices invite further work. First, improper priors yield the clean truthtelling benchmark and closed-form posteriors. However, with normal (proper) priors, messaging and ordering would interact, potentially strengthening echo-chamber forces through both content *and* speech. Second, our within-the-platform utility formulation captures simple conformity motives. However, incorporating richer attention dynamics (e.g., U-shaped engagement in similarity) and heterogeneous conformity costs would let feeds exploit both outrage and affinity. Finally, endogenizing platform competition under interoperability would quantify how much algorithmic discipline can be achieved by market design rather than direct ranking mandates (but this is already being studied in Banchio et al. (2025)).

Taken together, our results show that when platforms optimize for engagement, echo chambers are not an accident—they are the optimal instrument. Restoring learning therefore requires either changing the instrument (altering algorithms to time in disagreement, which matters especially for naives) or changing the game (making platforms compete on algorithms rather than on captive network scale).

References

Abreu, Luis and Doh-Shin Jeon, "Homophily in social media and news polarization," working paper, available at https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=3468416, 2019.

Acemoglu, Daron, Asuman Ozdaglar, and James Siderius, "A model of online misinformation," NBER Working Paper no 28884, available at https://siderius.lids.mit.edu/wp-content/uploads/sites/36/2022/09/fake-news-July-20-2022.pdf, 2023.

Allcott, Hunt and Matthew Gentzkow, "Social media and fake news in the 2016 election," *Journal of Economic Perspectives*, 2017, *31* (2), 211–236.

_ , Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow, "The welfare effects of social media," *American Economic Review*, 2020, *110* (3), 629–76.

_ , Matthew Gentzkow, and Lena Song, "Digital addiction," *American Economic Review*, 2022, *112* (7), 2424–2463.

Aridor, Guy, Rafael Jiménez-Durán, Ro'ee Levy, and Lena Song, "The economics of social media," 2024.

Armona, Luis, "Online Social Network Effects in Labor Markets: Evidence from Facebook's Entry to College Campuses," *Review of Economics and Statistics*, 2023, pp. 1–47.

Banchio, Martino, Francesco Decarolis, Jiménez-Durán Rafael, Carl-Christian Groh, and Miguel Risco, "Contestability and the Optimal Regulation of Social Media Platforms," working paper, 2025.

Beknazar-Yuzbashev, George, Rafael Jiménez-Durán, and Mateusz Stalinski, "A Model of Harmful Yet Engaging Content on Social Media," in "AEA Papers and Proceedings," Vol. 114 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2024, pp. 678–683.

Belleflamme, Paul and Martin Peitz, "The competitive impacts of exclusivity and price transparency in markets with digital platforms," *Concurrences*, 2020, (1), 2–12.

Bernheim, B. Douglas, "A theory of conformity," *Journal of Political Economy*, 1994, *102* (5), 841–877.

Bourreau, Marc, Adrien Raizonville, and Guillaume Thébaudin, "Interoperability between Ad-Financed Platforms with Endogenous Multi-Homing," 2023.

_ and Jan Kraemer, "Interoperability in Digital Markets: Boon or Bane for Market Contestability?," working paper, available at https://ssrn.com/abstract=4172255, 2023.

_ **and Jan Krämer**, "Interoperability in digital markets," available at https://sitic.org/wordpress/wp-content/uploads/Interoperability-in-Digital-Markets.pdf, 2022.

Braghieri, Luca, Ro'ee Levy, and Alexey Makarin, "Social media and mental health," *American Economic Review*, 2022, *112* (11), 3660–3693.

Bursztyn, Leonardo, Benjamin Handel, Rafael Jiménez-Durán, and Christopher Roth, "When product markets become collective traps: the case of social media," working paper, available at https://ssrn.com/abstract=4596071, 2023.

Chamley, Christophe, *Rational herds: Economic models of social learning*, Cambridge University Press, 2004.

Cialdini, Robert B. and Noah J. Goldstein, "Social influence: Compliance and conformity," *Annual Review of Psychology*, 2004, *55* (1), 591–621.

Dhakar, Mudit and Jun Yan, "Interoperability & Privacy: A Case of Messaging Apps," 2024.

Ellman, Matthew and Fabrizio Germano, "What do the papers sell? A model of advertising and media bias," *Economic Journal*, 2009, *119* (537), 680–704.

Franck, Jens-Uwe and Martin Peitz, "Market power of digital platforms," *Oxford Review of Economic Policy*, 2023, *39* (1), 34–46.

Garcia, Filomena and Muxin Li, "Post-Merger Strategies of Multiplatform Monopolies," *mimeo*, 2024.

Gauthier, Germain, Roland Hodler, Ekaterina Zhuravskaya, and Philine Widmer, "The Political Effects of X's Recommender Algorithm," working paper, 2025.

Gentzkow, Matthew and Jesse M Shapiro, "What drives media slant? Evidence from US daily newspapers," *Econometrica*, 2010, *78* (1), 35–71.

Golub, Benjamin and Matthew O. Jackson, "Naive learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, 2010, *2* (1), 112–149.

Greene, Steven, "Social identity theory and party identification," *Social science quarterly*, 2004, *85* (1), 136–153.

Guess, Andrew M., Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow et al., "How do social media feed algorithms affect attitudes and behavior in an election campaign?," *Science*, 2023, *381* (6656), 398–404.

Guriev, Sergei, Emeric Henry, Théo Marquis, and Ekaterina Zhuravskaya, "Curtailing false news, amplifying truth," working paper, available at https://shs.hal.science/halshs-04315924/document, 2023.

Hartigan, John A., *Bayes Theory*, Springer Science & Business Media, 1983.

Holtz, David, Ben Carterette, Praveen Chandar, Zahra Nazari, Henriette Cramer, and Sinan Aral, "The engagement-diversity connection: Evidence from a field experiment on spotify," in "Proceedings of the 21st ACM Conference on Economics and Computation" 2020, pp. 75–76.

Hoong, Ruru, "Self control and smartphone use: An experimental study of soft commitment devices," *European Economic Review*, 2021, *140*, 103924.

Horwitz, Jeff et al., "The Facebook files," *The Wall Street Journal,* available online at: https://www.wsj.com/articles/the-facebook-files-11631713039, 2021.

Jackson, Matthew O., Suraj Malladi, and David McAdams, "Learning through the grapevine and the impact of the breadth and depth of social networks," *Proceedings of the National Academy of Sciences*, 2022, *119* (34).

Kades, Michael and Fiona Scott Morton, "Interoperability as a competition remedy for digital networks," *Washington Center for Equitable Growth Working Paper Series*, 2020.

Kamath, Krishna, Aneesh Sharma, Dong Wang, and Zhijun Yin, "Realgraph: User interaction prediction at twitter," in "user engagement optimization workshop@ KDD" number ii 2014.

Kitchens, Brent, Steven L Johnson, and Peter Gray, "Understanding echo chambers and filter bubbles: The impact of social media on diversification and partisan shifts in news consumption," *MIS quarterly*, 2020, *44* (4), 1619–1649.

Kranton, Rachel and David McAdams, "Social connectedness and information markets," working paper, available at https://sites.duke.edu/rachelkranton/files/2022/12/Social_Connectedness___Information_Markets-Dec-17_2022-final.pdf, 2022.

Levy, Ro'ee, "Social media, news consumption, and polarization: Evidence from a field experiment," *American Economic Review*, 2021, *111* (3), 831–870.

Loomba, Sahil, Alexandre de Figueiredo, Simon J. Piatek, Kristen de Graaf, and Heidi J. Larson, "Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA," *Nature Human Behaviour*, 2021, *5* (3), 337–348.

Mosleh, Mohsen, Cameron Martel, Dean Eckles, and David G. Rand, "Shared partisanship dramatically increases social tie formation in a Twitter field experiment," *Proceedings of the National Academy of Sciences*, 2021, *118* (7), e2022761118.

Mueller-Frank, Manuel, Mallesh M. Pai, Carlo Reggini, Alejandro Saporiti, and Luis Simantujak, "Strategic management of social information," working paper, available at *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4434685*, 2022.

Pariser, Eli, *The filter bubble: How the new personalized web is changing what we read and how we think*, Penguin, 2011.

Reuter, Jonathan and Eric Zitzewitz, "Do ads influence editors? Advertising and bias in the financial media," *Quarterly Journal of Economics*, 2006, *121* (1), 197–227.

Ross, Lee, David Greene, and Pamela House, "The "false consensus effect": An egocentric bias in social perception and attribution processes," *Journal of Experimental Social Psychology*, 1977, *13* (3), 279–301.

Silverman, Craig, "This analysis shows how viral fake election news stories outperformed real news on Facebook," *BuzzFeed,* available at https://newsmediauk.org/wp-content/uploads/2022/10/Buzzfeed-_This_Analysis_Shows_How_Viral_Fake_Election_News_Stories_Outperformed_Real_News_On_Facebook_-_BuzzFeed_News.pdf, November 16, 2016.

Sunstein, Cass R., "A prison of our own design: Divided democracy in the age of social media," *Democratic Audit UK*, 2017.

A Omitted proofs

Lemma A.1 (Variance of θ_j given θ_i)**.** *Let*

$$\theta_i = \theta + \varepsilon_i, \qquad \theta_j = \theta + \varepsilon_j,$$

where $(\varepsilon_i, \varepsilon_j)$ *is bivariate normal with variances* σ_{ii}, σ_{jj} *and correlation* σ_{ij}*. Assume improper prior for* θ*. Then,*

$$\theta \mid \theta_i \sim \mathcal{N}\Big(\theta_i,\ \sigma_{ii}\Big),$$

and

$$\theta_j \mid \theta_i \sim \mathcal{N}\Big(\theta_i,\ \sigma_{ii} + \sigma_{jj} - 2\sigma_{ij}\Big).$$

Proof. By Bayes with an improper prior, $f(\theta \mid \theta_i) \propto f(\theta_i \mid \theta)$, as $\theta_i \mid \theta \sim \mathcal{N}(\theta, \sigma_{ii})$,

$$\theta \mid \theta_i \sim \mathcal{N}(\theta_i, \sigma_{ii}).$$

Then, using conditional probability,

$$f(\theta_j|\theta_i, \theta) = \frac{f(\theta_j, \theta_i|\theta)}{f(\theta_i|\theta)},$$

which gives a standard result from multivariate normal:

$$\theta_j|\theta_i, \theta \sim N\left(\theta + \rho_{ij}\sqrt{\frac{\sigma_{jj}}{\sigma_{ii}}}(\theta_i - \theta),\ \sigma_{jj}(1-\rho_{ij}^2)\right), \tag{2}$$

where $\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$. Fix θ_i and write the joint conditional density

$$f(\theta, \theta_j \mid \theta_i) = f(\theta_j \mid \theta, \theta_i)\, f(\theta \mid \theta_i).$$

Up to a normalizing constant, the exponent is

$$-\frac{1}{2}\left[\frac{\Big(\theta_j - \theta - \rho_{ij}\frac{\sqrt{\sigma_{jj}}}{\sqrt{\sigma_{ii}}}(\theta_i - \theta)\Big)^2}{\sigma_{jj}(1-\rho_{ij}^2)} + \frac{(\theta - \theta_i)^2}{\sigma_{ii}}\right].$$

The term inside the parentheses is

$$\Big(\theta_j - \theta - \rho_{ij}\tfrac{\sqrt{\sigma_{jj}}}{\sqrt{\sigma_{ii}}}(\theta_i - \theta)\Big)^2 = (\theta_j - \theta)^2 - 2\rho_{ij}\tfrac{\sqrt{\sigma_{jj}}}{\sqrt{\sigma_{ii}}}(\theta_j - \theta)(\theta_i - \theta) + \rho_{ij}^2\tfrac{\sigma_{jj}}{\sigma_{ii}}(\theta_i - \theta)^2.$$

Using this, the exponent becomes

$$-\frac{1}{2(1-\rho_{ij}^2)}\left[\frac{(\theta_j - \theta)^2}{\sigma_{jj}} - \frac{2\rho_{ij}}{\sqrt{\sigma_{jj}\sigma_{ii}}}(\theta_j - \theta)(\theta_i - \theta) + \frac{(\theta_i - \theta)^2}{\sigma_{ii}}\right].$$

Expand the squares

$$(\theta_j-\theta)^2 = \theta_j^2 - 2\theta_j\theta + \theta^2, \quad (\theta_i-\theta)^2 = \theta_i^2 - 2\theta_i\theta + \theta^2, \quad (\theta_j-\theta)(\theta_i-\theta) = \theta_j\theta_i - \theta_j\theta - \theta_i\theta + \theta^2.$$

Collect the terms that depend on θ:

$$-\frac{1}{2}\left[A\,\theta^2 - 2B\,\theta\right] - \frac{1}{2}C$$

To simplify notation, we will define

$$A = \frac{1}{1-\rho_{ij}^2}\left(\frac{1}{\sigma_{jj}} + \frac{1}{\sigma_{ii}} - \frac{2\rho_{ij}}{\sqrt{\sigma_{jj}\sigma_{ii}}}\right),$$

$$B = \frac{1}{1-\rho_{ij}^2}\left(\frac{\theta_j}{\sigma_{jj}} + \frac{\theta_i}{\sigma_{ii}} - \frac{\rho_{ij}}{\sqrt{\sigma_{jj}\sigma_{ii}}}(\theta_j + \theta_i)\right),$$

$$C = \frac{1}{1-\rho_{ij}^2}\left(\frac{\theta_j^2}{\sigma_{jj}} - \frac{2\rho_{ij}\theta_j\theta_i}{\sqrt{\sigma_{jj}\sigma_{ii}}} + \frac{\theta_i^2}{\sigma_{ii}}\right).$$

Integrate out θ using

$$\int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left[A\,\theta^2 - 2B\,\theta\right]\right) d\theta = \sqrt{\frac{2\pi}{A}}\,\exp\left(\frac{B^2}{2A}\right).$$

Therefore the marginal in θ_j given θ_i is proportional to

$$\exp\left(-\frac{1}{2}\left[C - \frac{B^2}{A}\right]\right).$$

A direct simplification yields

$$C - \frac{B^2}{A} = \frac{(\theta_j - \theta_i)^2}{\sigma_{ii} + \sigma_{jj} - 2\rho_{ij}\sqrt{\sigma_{ii}\sigma_{jj}}}.$$

Hence

$$\theta_j \mid \theta_i \sim \mathcal{N}\left(\theta_i,\ \sigma_{ii} + \sigma_{jj} - 2\rho_{ij}\sqrt{\sigma_{ii}\sigma_{jj}}\right),$$

so the conditional variance is $\sigma_{ii} + \sigma_{jj} - 2\rho_{ij}\sqrt{\sigma_{ii}\sigma_{jj}}$, or equivalently $\sigma_{ii} + \sigma_{jj} - 2\sigma_{ij}$.

□

Proof of Proposition 3.1

Proof. The next proof will be divided into two parts. First, we will show that, for rationals, truth-telling is the unique equilibrium. Second, we will show that, for any specification of the function g, truth-telling is also an equilibrium for naives.

Rationals. User i, a rational user, chooses a message $m_i \in \mathbb{R}$ to maximize her within-the-platform expected utility given her private signal θ_i and the algorithm $\mathscr{F}$; that is,

$$\mathbb{E}[u_i \mid \theta_i, \mathscr{F}] = \lambda\Big(\alpha\,\mathbb{E}[e_i \mid \theta_i, \mathscr{F}] - \beta(\theta_i - m_i)^2 - (1-\beta)\,\mathbb{E}\Big[\sum_{j\in\mathscr{F}_i^{e_i}} (m_i - m_j(\theta_j))^2 \;\Big|\; \theta_i, \mathscr{F}\Big]\Big).$$

Because e_i is chosen later in the game, unlike for naive users we need not analyze the effect of m_i on expected engagement. With the algorithm and the engagement level fixed, the choice of m_i does not affect learning.

Thus the optimal m_i maximizes

$$-\beta(\theta_i - m_i)^2 - (1-\beta)\Big(e_i m_i^2 + \mathbb{E}\Big[\sum_{j\in\mathscr{F}_i^{e_i}} m_j(\theta_j)^2 \;\Big|\; \theta_i, \mathscr{F}\Big] - 2m_i\,\mathbb{E}\Big[\sum_{j\in\mathscr{F}_i^{e_i}} m_j(\theta_j) \;\Big|\; \theta_i, \mathscr{F}\Big]\Big).$$

The first-order condition with respect to m_i is

$$(\beta + (1-\beta)e_i)\, m_i = \beta\,\theta_i + (1-\beta)\,\mathbb{E}\Big[\sum_{j\in\mathscr{F}_i^{e_i}} m_j(\theta_j) \;\Big|\; \theta_i, \mathscr{F}\Big]. \tag{10}$$

This already shows that truth-telling is an equilibrium; if every other user chooses truthtelling, user i best replies with truthtelling. We now prove uniqueness for rational users. For any user ℓ, define

$$D_\ell := \beta + (1-\beta)\, e_\ell.$$

Because the first-order condition holds for every user, substitute each neighbor's best reply into user i's best reply and iterate this substitution for neighbors of neighbors up to depth m:

$$\begin{aligned}
m_i &= \frac{\beta}{D_i}\theta_i + \frac{1-\beta}{D_i}\,\mathbb{E}_i\Big[\sum_{j\in\mathscr{F}_i^{e_i}}\Big(\frac{\beta}{D_j}\theta_j + \frac{1-\beta}{D_j}\,\mathbb{E}_j\Big[\sum_{l\in\mathscr{F}_j^{e_j}} m_l(\theta_l)\Big]\Big)\Big] \\
&= \frac{\beta}{D_i}\theta_i + \frac{1-\beta}{D_i}\,\mathbb{E}_i\Big[\sum_{j\in\mathscr{F}_i^{e_i}}\frac{\beta}{D_j}\theta_j\Big] + \frac{1-\beta}{D_i}\,\mathbb{E}_i\Big[\sum_{j\in\mathscr{F}_i^{e_i}}\frac{1-\beta}{D_j}\,\mathbb{E}_j\Big[\sum_{l\in\mathscr{F}_j^{e_j}}\frac{\beta}{D_l}\theta_l\Big]\Big] \\
&\quad + \frac{1-\beta}{D_i}\,\mathbb{E}_i\Big[\sum_{j\in\mathscr{F}_i^{e_i}}\frac{1-\beta}{D_j}\,\mathbb{E}_j\Big[\sum_{l\in\mathscr{F}_j^{e_j}}\frac{1-\beta}{D_l}\,\mathbb{E}_l\Big[\sum_{q\in\mathscr{F}_l^{e_l}} m_q(\theta_q)\Big]\Big]\Big] \;+\; \cdots \\
&= \frac{\beta}{D_i}\theta_i + \sum_{r=1}^{m-1}\mathbb{E}_i\Big[\sum_{j_1\in\mathscr{F}_i^{e_i}}\frac{1-\beta}{D_i}\sum_{j_2\in\mathscr{F}_{j_1}^{e_{j_1}}}\frac{1-\beta}{D_{j_1}}\cdots\sum_{j_r\in\mathscr{F}_{j_{r-1}}^{e_{j_{r-1}}}}\frac{\beta}{D_{j_r}}\,\theta_{j_r}\Big] \\
&\quad + \mathbb{E}_i\Big[\sum_{j_1\in\mathscr{F}_i^{e_i}}\frac{1-\beta}{D_i}\sum_{j_2\in\mathscr{F}_{j_1}^{e_{j_1}}}\frac{1-\beta}{D_{j_1}}\cdots\sum_{j_m\in\mathscr{F}_{j_{m-1}}^{e_{j_{m-1}}}}\frac{1-\beta}{D_{j_{m-1}}}\,\mathbb{E}_{j_m}\Big[\sum_{p\in\mathscr{F}_{j_m}^{e_{j_m}}} m_p(\theta_p)\Big]\Big].
\end{aligned} \tag{11}$$

By improper priors and the law of iterated expectations,

$$\mathbb{E}\Big[\sum_{j\in\mathscr{F}_i^{e_i}}\theta_j \;\Big|\; \theta_i, \mathscr{F}\Big] = e_i\,\theta_i.$$

Applying this identity to the first inner term in (11) replaces each θ_j with θ_i inside the conditional expectation. Repeating the same step at each subsequent level shows that the depth-r contribution is a nonnegative multiple of θ_i.

Lets first show that the third term converges to zero. For this term, each additional

nesting contributes a factor

$$\frac{1-\beta}{D_\ell} = \frac{1-\beta}{\beta + (1-\beta)e_\ell},$$

and the adjacent sum has at most e_ℓ terms. Hence the depth-ℓ layer scales the upstream magnitude by

$$\frac{(1-\beta)e_\ell}{\beta + (1-\beta)e_\ell} < 1.$$

Let

$$\gamma_{\max} := \max_\ell \frac{(1-\beta)e_\ell}{\beta + (1-\beta)e_\ell} < 1 \quad \text{(finite since } e_\ell \le n\text{)}.$$

From the quadratic loss, admissible strategies satisfy $\mathbb{E}_i[m_j(\theta_j)^2] < \infty$; if reading a message produces an expected disutility larger than the utility that the user can receive in isolation ($e_i = 1$), then the user will not read that message. By Cauchy–Schwarz there is a finite constant

$$K := \sup_{j \in \mathcal{U}} \mathbb{E}_i\Big[\, |m_j(\theta_j)| \,\Big] < \infty.$$

Conditional expectations do not increase L^1 norms, so the absolute value of the nested remainder after m layers is bounded by $\gamma_{\max}^m K \to 0$ as $m \to \infty$. Since the third term vanishes, (11) reduces to the fixed-point form

$$m_i = \frac{\beta}{D_i}\theta_i + \frac{1-\beta}{D_i}\mathbb{E}\left[\sum_{j \in \mathscr{F}_i^{e_i}} m_j \mid \theta_i \mathscr{F}\right]$$

Define the iteration

$$m_i^{(r+1)} = \frac{\beta}{D_i}\theta_i + \frac{1-\beta}{D_i}\mathbb{E}\Big[\sum_{j \in \mathscr{F}_i^{e_i}} m_j^{(r)} \;\Big|\; \theta_i, \mathscr{F}\Big], \qquad m^{(0)} \equiv 0.$$

The map $T_i(m) := \frac{\beta}{D_i}\theta_i + \frac{1-\beta}{D_i}\mathbb{E}[\sum_j m_j \mid \theta_i, \mathscr{F}]$ is a contraction in the sup norm with modulus

$$\frac{(1-\beta)e_i}{\beta + (1-\beta)e_i} = 1 - \frac{\beta}{D_i} < 1,$$

so $m^{(r)} \to m^*$ as $r \to \infty$, and m^* is the unique fixed point. To identify m^*, plug any fixed point into (10) and subtract the identity $(\beta + (1-\beta)e_i)\theta_i = \beta\theta_i + (1-\beta)\mathbb{E}_i[\sum_j \theta_j \mid \theta_i, \mathscr{F}] = \beta\theta_i + (1-\beta)e_i\theta_i$ (by improper priors):

$$(\beta + (1-\beta)e_i)\big(m_i^* - \theta_i\big) = (1-\beta)\,\mathbb{E}\Big[\sum_{j \in \mathscr{F}_i^{e_i}} \big(m_j^* - \theta_j\big) \;\Big|\; \theta_i, \mathscr{F}\Big].$$

Taking absolute values and then the maximum over i gives

$$\|m_x^* - \theta_x\|_\infty \le \max_i \frac{(1-\beta)e_i}{\beta + (1-\beta)e_i}\,\|m_x^* - \theta_x\|_\infty.$$

The factor on the right is strictly less than 1, hence $\|m_x^* - \theta_x\|_\infty = 0$ and

$$m_i^* = \theta_i.$$

Therefore, truthful reporting is the unique equilibrium for rational users.

Naive users The proof for naïve users proceeds in three steps. First, we show that the message that maximizes engagement also maximizes learning. Second, we establish that the message maximizing within–platform utility coincides with the one that maximizes expected engagement. Third, we demonstrate that truthful reporting maximizes within–platform utility. Step 2 provides the bridge between the two optimisation problems: it links the utility–maximisation argument in Step 3 to engagement (and, via Step 1, to learning), ensuring that truthtelling simultaneously maximises all three objectives.

First, notice that a naive user that only cares about learning wants to choose a message that maximizes engagement For naive users, engagement and learning are monotonically related. Given any algorithm and message profile, if a user reads weakly more posts (i.e. $e_i' \geq e_i$), her learning weakly improves (posterior variance weakly decreases). Hence, for naive users, choosing m_i to maximize engagement also maximizes learning under any feed. This follows directly from Bayesian updating: conditional on the feed, a larger number of signals implies a weakly lower posterior variance.

Next, we show that the message that maximizes the expected within–platform utility,

$$\mathbb{E}\left[u_i(m_i; m_{-i}, e_i, \mathcal{F}_i, \theta_i) \mid \mathcal{F}_i, \theta_i\right],$$

when $m_j = \theta_j$, also maximizes the expected engagement. By definition (finite support), the expected engagement is[16]

$$\mathbb{E}\left[e_i \mid m_i, \theta_i, \mathcal{F}_i\right] = \sum_{k=1}^{n} k \Pr(e_i = k \mid m_i, \theta_i, \mathcal{F}_i). \tag{12}$$

With the continuation function $g(\cdot)$, for $1 \leq k \leq n-1$,

$$\Pr(e_i = k \mid m_i, \theta_i, \mathcal{F}_i) = \Big(1 - g(u_i(m_i, k))\Big) \prod_{r=1}^{k-1} g(u_i(m_i, r)),$$

and the probability of reaching the cap n is

$$\Pr(e_i = n \mid m_i, \theta_i, \mathcal{F}_i) = \prod_{r=1}^{n-1} g(u_i(m_i, r)).$$

Therefore,

16 We write $u_i(m_i, k)$ for the utility when user i sends m_i and reads up to k posts of the algorithm.

Here $\mathbb{E}_{\theta_{-i}}[\cdot]$ integrates over the unknown messages (via θ_{-i}) while holding the engagement realization fixed, in contrast to $\mathbb{E}_i[\cdot]$, which averages over both θ_{-i} and e_i. Differentiating expected engagement with respect to m_i (chain and product rules) yields

$$\frac{\partial}{\partial m_i}\mathbb{E}\left[e_i \mid m_i, \theta_i, \mathcal{F}_i\right] = \mathbb{E}_{\theta_{-i}}\left[\sum_{k=1}^{n-1} k\left(\left(1-g_k\right)\prod_{r=1}^{k-1} g_r \sum_{s=1}^{k-1}\frac{g_s'}{g_s}\frac{\partial u_s}{\partial m_i} - g_k'\frac{\partial u_k}{\partial m_i}\prod_{r=1}^{k-1} g_r\right) + n\prod_{r=1}^{n-1} g_r \sum_{s=1}^{n-1}\frac{g_s'}{g_s}\frac{\partial u_s}{\partial m_i}\right], \tag{13}$$

where, for brevity, $g_r \equiv g(u_i(m_i, r))$, $g_r' \equiv g'(u_i(m_i, r))$, and $u_s \equiv u_i(m_i, s)$.

Assume that all other users report truthfully, that is, $m_j = \theta_j$ for all $j \in \mathcal{U}\backslash\{i\}$. We will show that truthful reporting is a best reply for user i. Under this assumption, the utility for each engagement level k is

$$u_i(m_i, k; \theta_j, \mathcal{F}) = -\beta(m_i - \theta_i)^2 - (1-\beta)\sum_{j\in\mathcal{F}^k}(m_i - \theta_j)^2.$$

The first derivative with respect to m_i is

$$\frac{\partial u_i(m_i, k; \theta_j, \mathcal{F})}{\partial m_i} = -2\beta(m_i - \theta_i) - 2(1-\beta)\sum_{j\in\mathcal{F}^k}(m_i - \theta_j).$$

Define $z_j = m_i - \theta_j$. Conditional on θ_i, the improper–prior assumption implies $\mathbb{E}[\theta_j \mid \theta_i] = \theta_i$, so $\mathbb{E}[z_j \mid \theta_i] = m_i - \theta_i$, which equals zero under truthful reporting. *All expressions below are evaluated at* $m_i = \theta_i$, so the sincerity term $-\beta(m_i - \theta_i)^2$ vanishes identically, and u_i, g, and their derivatives depend only on the squared deviations z_j^2. Hence the function $f(z_j) := \frac{\partial u_i(m_i, k; \theta_j, \mathcal{F})}{\partial m_i}$ is odd in z_j, while the other terms are even.

Hence, if we define

$$h(z_j) := u_i(m_i = \theta_i, k; \theta_j, \mathcal{F}),$$

we have $h(-z_j) = h(z_j)$, i.e. h is an even function of z_j. The same property holds for $g(h(z_j))$ and $g'(h(z_j))$.

Since the product of even functions is even, all terms in the first-order condition of expected engagement (13) that do not involve $f(z_j)$ are even functions. Moreover, the product of an even and an odd function is odd. Therefore, if we rewrite equation (13) as a function of the z_j, each summand is an odd function of the z_j.

Because the expectation (integral over $\mathbb{R}$) of any symmetric odd function is zero, the first-order condition of expected engagement is zero when $m_j = \theta_j$ and $m_i = \theta_i$. Hence, truthful reporting is a best reply to truthful reporting. This reasoning establishes that, evaluated at $m_i = \theta_i$, the *total* derivative of expected engagement with respect to the message m_i is zero. This means that truthful reporting satisfies the first–order condition for maximizing expected engagement. However, this does not yet guarantee that the expected utility is maximized, since the latter expectation is taken over both e_i and θ_j.

To conclude, we show that, for any given feed, if all other users report truthfully, user i maximizes her expected within–platform utility by reporting truthfully herself, that is, $m_i = \theta_i$.

Fix θ_i and assume $m_j = \theta_j$ for all $j \neq i$. The within–platform expected utility is

$$\mathbb{E}\Big[u_i(m_i, \theta_i; m_{-i}) \mid \theta_i, \mathcal{F}_i\Big] = \alpha\, \mathbb{E}\Big[e_i \mid \theta_i, \mathcal{F}_i\Big] - \beta\, \mathbb{E}\Big[(\theta_i - m_i)^2 \mid \theta_i\Big] - (1-\beta)\, \mathbb{E}\Big[\sum_{j \in \mathcal{F}_i^{e_i}} (m_i - \theta_j)^2 \;\Big|\; \theta_i, \mathcal{F}_i\Big].$$

The second term is simply $-\beta(m_i - \theta_i)^2$. For the third term, apply the quadratic decomposition

$$(m_i - \theta_j)^2 = (m_i - \theta_i)^2 + (\theta_j - \theta_i)^2 + 2(m_i - \theta_i)(\theta_i - \theta_j).$$

Taking conditional expectations given θ_i and using $\mathbb{E}[\theta_j \mid \theta_i] = \theta_i$ (truthful reports and conditional symmetry), the cross term vanishes and, for any engagement size k,

$$\mathbb{E}\Big[\sum_{j \in \mathcal{F}_i^k} (m_i - \theta_j)^2 \;\Big|\; \theta_i, \mathcal{F}_i\Big] = k\,(m_i - \theta_i)^2 + \sum_{j \in \mathcal{F}_i^k} \mathbb{E}\Big[(\theta_j - \theta_i)^2 \mid \theta_i\Big].$$

Hence, irrespective of how e_i is distributed,

$$\mathbb{E}\Big[\sum_{j \in \mathcal{F}_i^{e_i}} (m_i - \theta_j)^2 \;\Big|\; \theta_i, \mathcal{F}_i\Big] = \mathbb{E}[e_i \mid \theta_i, \mathcal{F}_i]\,(m_i - \theta_i)^2 + \mathbb{E}\Big[\sum_{j \in \mathcal{F}_i^{e_i}} (\theta_j - \theta_i)^2 \;\Big|\; \theta_i, \mathcal{F}_i\Big].$$

Substituting back, we obtain

$$\begin{aligned}\mathbb{E}\Big[u_i(m_i, \theta_i; m_{-i}) \mid \theta_i, \mathcal{F}_i\Big] = \alpha\mathbb{E}[e_i \mid \theta_i, \mathcal{F}_i] - \Big[\beta + (1-\beta)\mathbb{E}[e_i \mid \theta_i, \mathcal{F}_i]\Big](m_i - \theta_i)^2 \\ - (1-\beta)\mathbb{E}\Big[\sum_{j \in \mathcal{F}_i^{e_i}} (\theta_j - \theta_i)^2 \;\Big|\; \theta_i, \mathcal{F}_i\Big].\end{aligned}$$

Define the expected utility as a function of m_i fixing the other players strategy, $f(m_i)$. Then, its first derivative equals

$$\begin{aligned}f'(m_i) = \alpha\frac{\partial}{\partial m_i}\mathbb{E}[e_i \mid \theta_i, \mathcal{F}_i] - 2\Big[\beta + (1-\beta)\mathbb{E}[e_i \mid \theta_i, \mathcal{F}_i]\Big](m_i - \theta_i) - \\ (1-\beta)\frac{\partial}{\partial m_i}\mathbb{E}[e_i \mid \theta_i, \mathcal{F}_i](m_i - \theta_i)^2.\end{aligned}$$

The first term is zero at $m_i = \theta_i$, because we already proved that truthful reporting maximizes the expected engagement. The other two terms are trivially zero too. Hence, we only need to show that the function is concave at $m_i = \theta_i$ to prove that it is a maximum and, consequently, a best reply.

$$f''(m_i) = \alpha \frac{\partial^2}{\partial m_i^2} \mathbb{E}[e_i \mid \theta_i, \mathcal{F}_i] - 2\Big[\beta + (1-\beta)\mathbb{E}[e_i \mid \theta_i, \mathcal{F}_i]\Big]$$
$$-4(1-\beta)\frac{\partial}{\partial m_i}\mathbb{E}[e_i \mid \theta_i, \mathcal{F}_i](m_i - \theta_i)$$
$$-(1-\beta)\frac{\partial^2}{\partial m_i^2}\mathbb{E}[e_i \mid \theta_i, \mathcal{F}_i](m_i - \theta_i)^2$$

Again, as truthful reporting is an engagement maximizing strategy, the first term is non-positive, while the last two are equal to zero when $m_i = \theta_i$. As the second term is strictly negative, this proves that $f''(m_i) < 0$ and then truthful reporting is a maximum. □

Proposition A.2 (Truth-telling is the only g-robust equilibrium)**.** *Suppose users are naive and the continuation rule $g : \mathbb{R} \to (0,1)$ is C^1 and (weakly) increasing in the inside–platform utility. If a message profile $m = (m_i)_i$ is a Nash equilibrium for every such g, then $m_i = \theta_i$ for all i.*

Proof. Take $g(u) = \frac{1}{2}$ for every $u \in \mathbb{R}$. Then,

$$\frac{\partial \mathbb{E}(e_i \mid \theta_i, \mathscr{F})}{\partial m_i} = 0.$$

Thus the optimal m_i maximizes

$$-\beta(\theta_i - m_i)^2 - (1-\beta)\Big(e_i m_i^2 + \mathbb{E}\Big[\sum_{j \in \mathscr{F}_i^{e_i}} m_j(\theta_j)^2 \;\Big|\; \theta_i, \mathscr{F}\Big] - 2m_i\,\mathbb{E}\Big[\sum_{j \in \mathscr{F}_i^{e_i}} m_j(\theta_j) \;\Big|\; \theta_i, \mathscr{F}\Big]\Big).$$

Using a similar reasoning to the proof of Proposition 3.1, we can prove that truthful reporting is the unique equilibrium. □

Proof of Proposition 3.3

Proof. We show this result in two steps. First, we prove that maximizing the probability of user i staying one more period means showing her in her feed the message of a user (that, of course, has not appeared yet) who minimizes the expected conformity loss between them. Second, we show that an algorithm that reversely ranks with respect to the expected conformity loss to user i is precisely the one that maximizes expected engagement.

The probability that, under algorithm $\mathscr{F}$, user i stays for one more period after reading k posts is given by $g(u_i(k, m_i, m_{-i}, \mathscr{F}, \theta_i))$. Let us refer to such probability as $g(u_i(k, \mathscr{F}))$ to easen notation. To maximize such probability, the platform chooses the next user to appear in i's feed according to

$$\mathscr{F}_i(k) = \underset{j \in \mathcal{U} \setminus \mathscr{F}_i^{k-1}}{\operatorname{argmax}} \{\mathbb{E}_p[g(u_i(k, \mathscr{F}))]\}.$$

As g is increasing on u_i and expectations preserve orders, maximizing $\mathbb{E}_p[g(u_i(k, \mathscr{F}))]$ is equivalent to maximizing $\mathbb{E}_p[u_i(k, \mathscr{F})]$ and, because of truthful reporting, the only term

in user i's within-the-platform utility u_i that the platform can affect is conformity. Hence, the platform effectively chooses the k-th user in i's feed according to

$$\begin{aligned}\mathscr{F}_i(k) &= \underset{j \in \mathcal{U} \setminus \mathscr{F}_i^{k-1}}{\operatorname{argmax}} \left\{ -\mathbb{E}_p \left[\sum_{l \in \mathscr{F}_i^{k-1}} (\theta_i - \theta_l)^2 + (\theta_i - \theta_j)^2 \right] \right\} \\ &= \underset{j \in \mathcal{U} \setminus \mathscr{F}_i^{k-1}}{\operatorname{argmax}} \left\{ -\mathbb{E}_p \left[(\theta_i - \theta_j)^2 \right] \right\}. \end{aligned} \tag{14}$$

Maximizing the probability of user i staying for one more period is equivalent to minimizing the conformity cost of such subsequent period.

Now, let us show that an algorithm built by choosing the next user according to Equation (14) maximizes expected engagement. Given $\mathscr{F}$, the probability of staying at least until period k is $\Pi_{j=1}^{k-1} g(u_i(k, \mathscr{F}))$, and the probability of staying precisely until period k is

$$(1 - g(u_i(k, \mathscr{F}))) \prod_{j=1}^{k-1} g(u_i(j, \mathscr{F})).$$

Now, let us take two algorithms, namely $\mathscr{F}$ and $\mathscr{F}'$, such that the complete feed they show to user i is identical except from the fact that two users are interchanged, i.e., there exist a pair of users t and t' such that

$$\mathscr{F}_i(t) = \mathscr{F}'_i(t') \text{ and } \mathscr{F}_i(t') = \mathscr{F}'_i(t).$$

Moreover, we assume without loss of generality that $-\mathbb{E}_p\left[(\theta_i - \theta_t)^2\right] > -\mathbb{E}_p\left[(\theta_i - \theta'_t)^2\right]$, i.e., that $\mathscr{F}_i$ shows before the user who penalizes conformity the least among the two in the pair. Also without loss of generality, we can reorder users so that $t = 1$ and $t' = 2$ and, then, $g(u_i(1, \mathscr{F})) > g(u_i(1, \mathscr{F}'))$. The goal is to show that $\mathscr{F}_i$ yields higher expected engagement, where the formal expression for expected engagement is precisely

$$\mathbb{E}_p[e_i|\mathscr{F}] = \sum_{r=1}^{n} \left[r\, \mathbb{E}_p \left[(1 - g(u_i(r, \mathscr{F}))) \prod_{k=1}^{r-1} g(u_i(k, \mathscr{F})) \right] \right].$$

By construction, $\mathbb{E}_p[g(u_i(r, \mathscr{F}))] = \mathbb{E}_p[g(u_i(r, \mathscr{F}'))]$ for all $r \geq 2$. Finally, as $g(u_i(1, \mathscr{F})) > g(u_i(1, \mathscr{F}'))$,

$$\begin{aligned}\mathbb{E}_p[e_i|\mathscr{F}] &= \sum_{r=1}^{n} \left[r\, \mathbb{E}_p \left[(1 - g(u_i(r, \mathscr{F}))) \prod_{k=1}^{r-1} g(u_i(k, \mathscr{F})) \right] \right] \\ &\geq \sum_{r=1}^{n} \left[r\, \mathbb{E}_p \left[(1 - g(u_i(r, \mathscr{F}'))) \prod_{k=1}^{r-1} g(u_i(k, \mathscr{F}')) \right] \right] = \mathbb{E}_p[e_i|\mathscr{F}'], \end{aligned}$$

where the inequality can be shown to be true by induction. Finally, consider any algorithm $\mathscr{F}$. We have shown that taking any two users in a feed it induces and reordering them reversely following their loss in conformity improves such feed. If we repeat this procedure

until no further improvement is possible, we obtain the platform-optimal algorithm $\mathscr{P}$. This argument finishes the proof. □

Lemma A.3. *The posterior distribution of θ conditional on $\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}$ is given by*

$$\theta|\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}} \sim \mathcal{N}\left(\frac{\mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}}{\mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\mathbb{1}^t}, \frac{1}{\mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\mathbb{1}^t}\right),$$

where $\mathbb{1}$ is a n-vector of ones, $\boldsymbol{\Sigma}_{\mathscr{F}_i^{e_i}}$ is the restriction of $\boldsymbol{\Sigma}$ to the users in $\mathscr{F}_i^{e_i}$, and $\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}$ is the vector of private signals of the users in $\mathscr{F}_i^{e_i}$.

Proof. Let us assume, for simplicity, that the signals user i observes in her personalized feed $\mathscr{F}_i^{e_i}$ are $\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}} = \{\theta_1, ..., \theta_{e_i}\}$. We know that $(\theta_1 \dots \theta_{e_i}) \sim \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{\mathscr{F}_i^{e_i}})$ because of the properties of the multinormal distribution. Now, the posterior distribution of θ conditional on $\theta_{\mathscr{F}_i^{e_i}}$ is proportional to the likelihood function:

$$\begin{aligned} g(\theta|\theta_{\mathscr{F}_i^{e_i}}) &\propto \left(2\pi \det(\boldsymbol{\Sigma}_{\mathscr{F}_i^{e_i}})\right)^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\mathscr{F}_i^{e_i}})^t \boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\mathscr{F}_i^{e_i}})\right] \\ &= \left(2\pi \det(\boldsymbol{\Sigma}_{\mathscr{F}_i^{e_i}})\right)^{-1/2} \exp\left[-\frac{1}{2}\left(\theta^2 \mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\mathbb{1}^t - 2\theta\mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}} + \boldsymbol{\theta}^t_{\mathscr{F}_i^{e_i}}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}\right)\right]. \end{aligned}$$

Multiplying by the constant $\sqrt{\mathbb{1}\boldsymbol{\Sigma}_{\mathscr{F}_i^{e_i}}\mathbb{1}^t}\sqrt{\det(\boldsymbol{\Sigma}_{\mathscr{F}_i^{e_i}})}$, we obtain:

$$\begin{aligned} g(\theta|\theta_{\mathscr{F}_i^{e_i}}) &= \sqrt{\frac{\mathbb{1}\boldsymbol{\Sigma}_{\mathscr{F}_i^{e_i}}\mathbb{1}^t}{2\pi}} \exp\left[-\frac{1}{2}\left(\theta^2 \mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\mathbb{1}^t - 2\theta\mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}} + \frac{(\boldsymbol{\theta}^t_{\mathscr{F}_i^{e_i}}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\mathbb{1})^2}{\mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\mathbb{1}^t}\right)\right] \\ &= \sqrt{\frac{\mathbb{1}\boldsymbol{\Sigma}_{\mathscr{F}_i^{e_i}}\mathbb{1}^t}{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\theta - \frac{\mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}}{\mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\mathbb{1}^t}}{\sqrt{\frac{1}{\mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\mathbb{1}^t}}}\right)^2\right]. \end{aligned}$$

This is the distribution function of a normal random variable with mean $\frac{\mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}}{\mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\mathbb{1}^t}$ and variance $\frac{1}{\mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\mathbb{1}^t}$. Thus,

$$\theta|\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}} \sim \mathcal{N}\left(\frac{\mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}}{\mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\mathbb{1}^t}, \frac{1}{\mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{F}_i^{e_i}}\mathbb{1}^t}\right)$$

as we wanted to show. □

Proof of Proposition 3.7

Proof. Before proving this Proposition, we will state and prove two lemmas.

Lemma A.4 (Local swap and the best prefix length)**.** *Fix a feed F and a position k such that F^{k-1} has composition (d, r) and the next two items are (R, D). Let F' be obtained from F by swapping them to (D, R). If*

$$U(d+1, r) \leq U(d, r+1),$$

then $e^*(F') \geq e^*(F)$.

Proof. For all $m \leq k-1$, $U(F^m) = U((F')^m)$. At $m = k$, the prefixes differ and the first inequality gives $U((F')^k) \leq U(F^k)$. At $m = k+1$, both orders reach composition $(d+1, r+1)$, so utilities coincide again; for $m \geq k+2$ compositions match. We only need to study the effects on engagement of he positions k and $k+1$. Define $M = U(F, e^*(F))$. By assumption

$$U(d+1, r) \leq U(d, r+1),$$

and then,

$$U(d+1, r) \leq M.$$

Hence $e^*(F') \geq e^*(F)$. □

Lemma A.5 (Explicit form of the swap condition)**.** *Under (5), the local comparison $U(d+1, r) \leq U(d, r+1)$ is equivalent to*

$$\frac{\lambda(1-\beta)}{1-\lambda} \leq \frac{-\sigma^2(1-x)(d-r)\Big(1+x(d+r)\Big)}{\Big(4xdr - dx + d + 3rx + r - x + 1\Big)\Big(4xdr + 3dx + d - rx + r - x + 1\Big)}. \tag{15}$$

Moreover, for $x \in (0, 1)$ the denominator on the right-hand side is positive, so the right-hand side has sign $-\operatorname{sgn}(d-r)$ and equals 0 at $d = r$.

The previous lemma shows that the condition defined by Lemma A.4 will occur for small values of λ (when the user wants to learn) or for large values of β (when conformity matters little).

Now, start from any feed F. Inspect $F^{e^*(F)}$ and whenever you find an adjacent pair (R, D) at composition (d, r) satisfying (15), swap it to (D, R). After finitely many steps you obtain $\widetilde{F}$ with $e^*(\widetilde{F}) \geq e^*(F)$. Each admissible local swap weakly increases e^* by Lemma A.4. By Lemma A.5, if the condition holds at (d, r) for an (R, D) pair, it continues to hold when that pair shifts one step to the right. The procedure terminates because each swap strictly reduces the number of (R, D) inversions before the first index where (15) fails. For any subset of the feed not crossing a swap, d is unchanged; at a swapped position we replace R by D, so d does not decrease. Maximizing e^* over feeds yields the concluding property. □

Proof of Proposition 4.1

Proof. For any algorithm $\mathscr{F}$, the platform's expectation over user i's engagement is a finite real number even if platform size grows asymptotically large. If platform size is n, expected engagement is given by

$$\mathbb{E}_p[e_i] = \mathbb{E}_p\left[\sum_{r=1}^{n}\left[r(1-g(u_i(r,\mathscr{F})))\prod_{k=1}^{r-1}g(u_i(k,\mathscr{F}))\right]\right].$$

Note now that we have assumed that there is some pair $0<\underline{g}<\overline{g}<1$ such that for all $x\in\mathbb{R}$, $0<\underline{g}<g(x)<\overline{g}<1$. Thus,[17]

$$\begin{aligned}\mathbb{E}_p[e_i] &= \mathbb{E}_p\left[\sum_{r=1}^{n}\left[r(1-g(u_i(r,\mathscr{F})))\prod_{k=1}^{r-1}g(u_i(k,\mathscr{F}))\right]\right]\\ &= \mathbb{E}_p\left[\sum_{r=1}^{n}\left[r\prod_{k=1}^{r-1}g(u_i(k,\mathscr{F}))\right]-\sum_{r=1}^{n}\left[rg(u_i(r,\mathscr{F}))\prod_{k=1}^{r-1}g(u_i(k,\mathscr{F}))\right]\right]\\ &\leq \mathbb{E}_p\left[\sum_{r=1}^{n}r\overline{g}^{r-1}-\sum_{r=1}^{n}r\underline{g}^{r}\right]=\sum_{r=1}^{n}r\overline{g}^{r-1}-\sum_{r=1}^{n}r\underline{g}^{r}.\end{aligned}$$

As this is true for all $n\in\mathbb{N}$, we can take limits and state that, when platform size grows asymptotically large, user i's expected engagement is finite:

$$\lim_{n\to\infty}\mathbb{E}_p[e_i]\leq\lim_{n\to\infty}\left[\sum_{r=1}^{n}r\overline{g}^{r-1}-\sum_{r=1}^{n}r\underline{g}^{r}\right]=\frac{1}{(1-\overline{g})^2}-\frac{\underline{g}}{(1-\underline{g})^2}. \tag{16}$$

Finally, let us define

$$k_{pi}:=\min\left\{\tilde{k}\in\mathbb{N}\text{ such that }\tilde{k}\geq\frac{1}{(1-\overline{g})^2}-\frac{\underline{g}}{(1-\underline{g})^2}\right\}$$

and note that $k_{pi}>1$ because $\frac{1}{(1-\overline{g})^2}-\frac{\underline{g}}{(1-\underline{g})^2}>1$ if and only if $(2\overline{g}-\underline{g}-\overline{g}^2)+(2\overline{g}\underline{g}^2-2\overline{g}\underline{g})+(\underline{g}\overline{g}^2-\overline{g}^2\underline{g}^2)>0$, which holds precisely because $0<\underline{g}<\overline{g}<1$ We can repeat this proof for k_u changing the expectation of the platform for the expectation of the user to show the existence of k_u.

Finally, we can take the maximum of all the k_{pi} and define it as k_p. Notice that this is a natural number because it is bounded from above. To see this it is enough to take the maximum of all the $\bar{g}$ and all the $\underline{g}$ and use equation (16). □

Proof of Proposition 4.2

Proof. To prove this statement, we first show that as the platform grows, it becomes arbitrarily likely to find k users whose opinions are almost perfectly correlated with that of user i. Let ρ_{ij} denote the correlation between users i and j. Since each new user's correlation with i is drawn from a process with full support on $[-1,1]$, the probability of

17 For notational convenience, we set the continuation probability after the last post to zero, $g(u_i(n,\mathscr{F}))=0$, so that the generic expression $\Pr(e_i=r)=(1-g(u_i(r,\mathscr{F})))\prod_{k<r}g(u_i(k,\mathscr{F}))$ also captures the terminal mass at $r=n$.

drawing a user with correlation above $1-\varepsilon$ is strictly positive for any $\varepsilon > 0$. Hence, by the law of large numbers, the number of users satisfying $\rho_{ij} > 1-\varepsilon$ grows linearly with n. It follows that for every pair $\varepsilon, \gamma > 0$, there exists $\bar{n} \in \mathbb{N}$ such that for all $n > \bar{n}$, the probability that user i has at least k neighbors $j_1, \ldots, j_k$ with $\rho_{j_r,i} > 1-\varepsilon$ for all $r \in \{1, \ldots, k\}$ exceeds $1-\gamma$.

Applying the Cauchy–Schwarz inequality to the correlations between user i and any two of her neighbors, say j_r and j_l, we obtain

$$\rho_{j_r,j_l} \;\geq\; \rho_{j_r,i}\rho_{j_l,i} - \sqrt{(1-\rho_{j_r,i}^2)(1-\rho_{j_l,i}^2)}.$$

Conditioning on the event that both users satisfy $\rho_{j_r,i} > 1-\varepsilon$ and $\rho_{j_l,i} > 1-\varepsilon$, the right-hand side is bounded below by $1-4\varepsilon+2\varepsilon^2$. Since each of these two events occurs with probability at least $1-\gamma$, it follows that

$$\mathbb{P}\Big[\rho_{j_r,j_l} \geq 1-4\varepsilon+2\varepsilon^2\Big] \;\geq\; (1-\gamma)^2 \quad \forall\, j_r, j_l.$$

Let $X_n := \sum_{j=1}^{n} 1\{\rho_{j,i} > 1-\varepsilon\}$. Under independence across j and $\mathbb{P}(\rho_{j,i} > 1-\varepsilon) = p > 0$, we have $\mathbb{P}(X_n \geq e_i(n)) \to 1$ as $n \to \infty$. On the event $\{X_n \geq e_i(n)\}$, the feed can be chosen entirely from users with $\rho_{j,i} > 1-\varepsilon$, so by the Cauchy–Schwarz bound every pair in the feed satisfies $\rho_{r,l} \geq 1-\delta$ with $\delta = 4\varepsilon - 2\varepsilon^2$. Hence, for all large n,

$$\mathbb{P}\Big[A(n) \leq \mathbf{\Sigma}_{\mathscr{P}_i^{e_i(n)}}\Big] \;\geq\; 1-\gamma,$$

and in fact this probability tends to 1 as $n \to \infty$.

Now, for each platform size n, engagement levels may vary, but it is always bounded by some k as already shown. The corresponding platform optimal feed is denoted by $\mathscr{P}_i^{e_i(n)}$. Let $\delta = 4\varepsilon - 2\varepsilon^2$. For every $\delta > 0$ and $\gamma > 0$, there exists some $\tilde{n} \in \mathbb{N}$ such that, for all $n > \tilde{n}$, the set of k users defined above satisfies

$$\mathbb{P}[\rho_{j_r,j_l} > 1-\delta] > 1-\gamma \quad \text{for all } j_r, j_l \in \{j_1, \ldots, j_k\},$$

which follows by choosing ε sufficiently small. For any given engagement level $e_i(n) \leq k$, the $e_i(n)$ users displayed in user i's feed are selected from this set. Define the associated $e_i(n) \times e_i(n)$ matrix

$$A(n) := \sigma^2 \begin{pmatrix} 1 & 1-\delta & \cdots & 1-\delta \\ 1-\delta & 1 & \cdots & 1-\delta \\ \vdots & \vdots & \ddots & \vdots \\ 1-\delta & 1-\delta & \cdots & 1 \end{pmatrix},$$

which represents the covariance matrix of size $e_i(n)$ signals with equal pairwise correlation $1-\delta$. Since, with probability at least $1-\gamma$, every pair of users in the feed has correlation

above $1-\delta$, it follows that

$$\mathbb{P}\Big[A(n) \le \mathbf{\Sigma}_{\mathscr{P}_i^{e_i(n)}}\Big] \ge 1-\gamma,$$

where $A(n) \le \mathbf{\Sigma}_{\mathscr{P}_i^{e_i(n)}}$ denotes elementwise inequality. Although correlations across different pairs are not independent, this bound suffices to establish that, with probability arbitrarily close to one for large n, the actual covariance matrix $\mathbf{\Sigma}_{\mathscr{P}_i^{e_i(n)}}$ dominates $A(n)$ elementwise. Now, we need an auxiliary result:

Lemma A.6 (Comparison under near–equicorrelation)**.** *Let $e \in \mathbb{N}$ and $\sigma^2 > 0$. For any $\delta \in (0,1)$, define the $e \times e$ equicorrelated covariance matrix*

$$A \;=\; \sigma^2\Big[(1-\delta)\,\mathbb{1}\mathbb{1}^\top + \delta I_e\Big],$$

whose off–diagonal correlations equal $1-\delta$. Let Σ be any symmetric positive definite matrix with the same diagonal σ^2 and off–diagonal elements satisfying $\Sigma_{rs} \ge \sigma^2(1-\delta)$ for all $r \ne s$. Write $\Sigma = A + E$, where E is symmetric with zero diagonal and nonnegative off–diagonal entries.

If the norm of the matrix E, $\|E\|_{\mathrm{op}} < \sigma^2\delta$, then both matrices are invertible and

$$\mathbb{1}^\top\Sigma^{-1}\mathbb{1} - \mathbb{1}^\top A^{-1}\mathbb{1} \;\le\; \frac{e}{\sigma^4\delta^2}\,\frac{\|E\|_{\mathrm{op}}}{1-\|E\|_{\mathrm{op}}/(\sigma^2\delta)}. \tag{17}$$

In particular, if $\|E\|_{\mathrm{op}} = o(1)$ (for instance, when all pairwise correlations in Σ converge to $1-\delta$), then

$$\mathbb{1}^\top\Sigma^{-1}\mathbb{1} \;=\; \mathbb{1}^\top A^{-1}\mathbb{1} + o(1), \quad \text{where } o(1) \to 0 \text{ as } n \to \infty.$$

Proof. Since $\|A^{-1}\|_{\mathrm{op}}\|E\|_{\mathrm{op}} < 1$, the Neumann series expansion of the inverse is valid:

$$\Sigma^{-1} = (A+E)^{-1} = A^{-1} - A^{-1}EA^{-1} + A^{-1}EA^{-1}EA^{-1} - \cdots.$$

Let $u = \mathbb{1}$. Subtracting and bounding term by term gives

$$|u^\top(\Sigma^{-1} - A^{-1})u| \;\le\; \sum_{k\ge1}\|A^{-1}\|_{\mathrm{op}}^{k+1}\|E\|_{\mathrm{op}}^{k}\|u\|_2^2 = \frac{\|A^{-1}\|_{\mathrm{op}}^2\|E\|_{\mathrm{op}}}{1-\|A^{-1}\|_{\mathrm{op}}\|E\|_{\mathrm{op}}}\|u\|_2^2.$$

For $A = \sigma^2[(1-\delta)\,uu^\top + \delta I_e]$, the eigenvalues are $\lambda_1 = \sigma^2(\delta + (1-\delta)e)$ along u and $\lambda_2 = \cdots = \lambda_e = \sigma^2\delta$ on its orthogonal complement, so $\|A^{-1}\|_{\mathrm{op}} = (\sigma^2\delta)^{-1}$. Substituting this and $\|u\|_2^2 = e$ yields (17). Finally, because E has zero diagonal and $0 \le E_{rs} \le \sigma^2\delta$, its operator norm satisfies $\|E\|_{\mathrm{op}} \le e\sigma^2\delta$. Thus the smallness condition can always be enforced by taking δ sufficiently small or restricting to the high–correlation event. □

Therefore, on the high–correlation event of probability at least $1-\gamma$ (for all n large

enough),

$$A(n) \le \mathbf{\Sigma}_{\mathscr{P}_i^{e_i(n)}} \quad \text{and} \quad \mathbb{1}^\top \mathbf{\Sigma}^{-1}_{\mathscr{P}_i^{e_i(n)}} \mathbb{1} \;\le\; \mathbb{1}^\top A(n)^{-1} \mathbb{1} \;+\; o(1),$$

where $o(1) \to 0$ as $n \to \infty$ (for fixed δ). Taking reciprocals (which reverses inequalities for positive terms) yields

$$\frac{1}{\mathbb{1}^\top A(n)^{-1}\mathbb{1} + o(1)} \;\le\; \frac{1}{\mathbb{1}^\top \mathbf{\Sigma}^{-1}_{\mathscr{P}_i^{e_i(n)}} \mathbb{1}} \;=\; \mathrm{Var}\Big[\theta \,\Big|\, \boldsymbol{\theta}_{\mathscr{P}_i^{e_i(n)}}\Big] \;\le\; \sigma^2.$$

Moreover, since $\mathbb{1}^\top A(n)^{-1}\mathbb{1} = \dfrac{e_i(n)}{\sigma^2\{\,1 + (e_i(n)-1)(1-\delta)\,\}}$, we obtain

$$\frac{\sigma^2\{\,1 + (e_i(n)-1)(1-\delta)\,\}}{e_i(n)} - o(1) \;\le\; \mathrm{Var}\Big[\theta \,\Big|\, \boldsymbol{\theta}_{\mathscr{P}_i^{e_i(n)}}\Big] \;\le\; \sigma^2 \qquad \text{with probability at least } 1-\gamma.$$

Letting n converge to infinity (for fixed δ) the lower bound converges to $\sigma^2(1-\delta)$; then letting δ converge to zero gives

$$\operatorname*{plim}_{n\to\infty} \mathrm{Var}\Big[\theta \,\Big|\, \boldsymbol{\theta}_{\mathscr{P}_i^{e_i(n)}}\Big] = \sigma^2.$$

□

Proof of Proposition 5.1

Proof. We compare the large–n expected utilities of $\mathscr{P}$ and $\mathscr{R}$ for a naïve user i.

We first study the utility under the platform optimal algorithm. By the closest–peers result (Proposition 4.2), as n increases the feed under $\mathscr{P}$ selects users with $\rho_{ij} \to 1$ while e_i remains finite, so the conformity term vanishes:

$$\lim_{n\to\infty} \mathbb{E}\left[\sum_{j \in \mathscr{P}_i^{e_i}} (\theta_i - \theta_j)^2 \,\middle|\, \mathscr{P}\right] = 0.$$

Moreover, the posterior variance under $\mathscr{P}$ converges to the prior, so the learning term equals $(1-\lambda)\sigma^2$ in the limit. Hence

$$\lim_{n\to\infty} \mathbb{E}\Big[U_i(e_i, m_i, m_{-i}, a_i, \mathscr{P}, \theta_i, \theta) \,\Big|\, \mathscr{P}\Big] = \lambda\,\alpha\,\mathbb{E}[e_i \mid \mathscr{P}] - (1-\lambda)\sigma^2.$$

We now study the utility under the reverse chronological algorithm. Users are not selected by closeness, so the conformity term need not vanish. Using $\mathbb{E}\big[(\theta_i - \theta_j)^2\big] = 2\sigma^2(1-\rho_{ij})$ (and noting $j = i$ contributes zero), we obtain

$$\mathbb{E}\left[\sum_{j \in \mathscr{R}_i^{e_i}} (\theta_i - \theta_j)^2 \,\middle|\, \mathscr{R}\right] = 2\sigma^2\,\mathbb{E}\left[\sum_{j \in \mathscr{R}_i^{e_i} \setminus \{i\}} (1 - \rho_{ij}) \,\middle|\, \mathscr{R}\right].$$

If, under $\mathscr{R}$, $\mathbb{E}[\rho_{ij}] = 0$, then

$$\mathbb{E}\left[\sum_{j\in\mathscr{R}_i^{e_i}} (\theta_i - \theta_j)^2 \,\middle|\, \mathscr{R}\right] = 2\sigma^2\,\mathbb{E}[\,e_i - 1 \mid \mathscr{R}].$$

For the learning term under $\mathscr{R}$, the posterior–variance component equals $(1-\lambda)\sigma^2\,\mathbb{E}\left[\frac{1}{e_i} \mid \mathscr{R}\right]$ (Equation (9)). Therefore,

$$\lim_{n\to\infty} \mathbb{E}\Big[U_i(e_i, m_i, m_{-i}, a_i, \mathscr{R}, \theta_i, \theta) \,\Big|\, \mathscr{R}\Big] =$$
$$\lambda\Big(\alpha\,\mathbb{E}[e_i \mid \mathscr{R}] - 2(1-\beta)\,\sigma^2\,\mathbb{E}[\,e_i - 1 \mid \mathscr{R}]\Big) - (1-\lambda)\sigma^2\,\mathbb{E}\Big[\tfrac{1}{e_i} \mid \mathscr{R}\Big].$$

To finish, let's compare both terms. User i weakly prefers $\mathscr{P}$ to $\mathscr{R}$ in the large–n limit iff

$$\lambda\,\alpha\,\mathbb{E}[e_i \mid \mathscr{P}] - (1-\lambda)\sigma^2 \;\geq\; \lambda\Big(\alpha\,\mathbb{E}[e_i \mid \mathscr{R}] - 2(1-\beta)\,\sigma^2\,\mathbb{E}[\,e_i - 1 \mid \mathscr{R}]\Big) - (1-\lambda)\sigma^2\,\mathbb{E}\Big[\tfrac{1}{e_i} \mid \mathscr{R}\Big].$$

Rearranging yields

$$\lambda \;\geq\; \frac{\sigma^2\Big(1 - \mathbb{E}\Big[\frac{1}{e_i} \mid \mathscr{R}\Big]\Big)}{\alpha(\mathbb{E}[e_i \mid \mathscr{P}] - \mathbb{E}[e_i \mid \mathscr{R}]) + 2(1-\beta)\,\sigma^2\,\mathbb{E}[\,e_i - 1 \mid \mathscr{R}] + \sigma^2\Big(1 - \mathbb{E}\Big[\frac{1}{e_i} \mid \mathscr{R}\Big]\Big)}.$$

Since $\mathscr{P}$ maximizes engagement, $\mathbb{E}[e_i \mid \mathscr{P}] \geq \mathbb{E}[e_i \mid \mathscr{R}]$, and because $e_i \geq 1$, we have $\mathbb{E}\left[\frac{1}{e_i} \mid \mathscr{R}\right] \leq 1$. Taking the maximum over $i \in \mathcal{U}$ yields the stated threshold. □

Proof of Proposition 5.2

Proof. Let $U_i(k)$ denote user i's expected utility after reading the first k items of her feed $\mathscr{R}_i^k$, where the order of the remaining items is uniformly random. At depth k, the continuation decision compares the expected marginal gain from reading one more item with the expected marginal cost. Reading one additional item increases the within–platform engagement component by $\lambda\,\alpha$.

Disagreement term. Let $\mathscr{R}_i^k$ denote the set of $n-k-1$ accounts the user has not yet read. Because the next post comes from a random user $j \in \mathscr{R}_i^k$, each such user appears with probability $1/(n-k-1)$. Hence the expected disagreement cost of the next post is the uniform average of the conditional variances of the remaining signals:

$$\mathbb{E}_i\Big[(\theta_j - \theta_i)^2 \mid \mathscr{R}_i^k\Big] = \frac{1}{n-k-1} \sum_{j\in\mathscr{R}_i^k} \mathbb{E}_i\Big[(\theta_j - \theta_i)^2 \mid \mathscr{R}_i^k\Big].$$

Learning term. The learning component improves by the expected reduction in posterior variance when moving from k to $k+1$ signals, i.e.

$$(1-\lambda)\Big(\mathbb{E}_i[\mathrm{Var}(\theta \mid \mathscr{R}_i^k)] - \mathbb{E}_i[\mathrm{Var}(\theta \mid \mathscr{R}_i^{k+1})]\Big).$$

Marginal comparison. Combining these components, the expected change in utility from reading one more post is

$$U_i(k{+}1) - U_i(k) =$$
$$\lambda\,\alpha - \frac{\lambda(1-\beta)}{n{-}k{-}1}\sum_{j\in\mathscr{R}_i^k}\mathbb{E}_i\Big[(\theta_j-\theta_i)^2 \mid \mathscr{R}_i^k\Big] + (1-\lambda)\Big(\mathbb{E}_i[\mathrm{Var}(\theta \mid \mathscr{R}_i^k)] - \mathbb{E}_i[\mathrm{Var}(\theta \mid \mathscr{R}_i^{k+1})]\Big).$$

It is therefore optimal to continue scrolling iff $U_i(k{+}1) - U_i(k) > 0$, which yields exactly the inequality in the statement. If the inequality fails (weakly), the user stops. □

Proof of Proposition 5.3

Proof. The argument has two steps. We first compare conformity losses, then learning. Signals are jointly normal with common variance σ^2, and

$$\mathbb{E}\Big[(\theta_i-\theta_j)^2\Big] = 2\sigma^2(1-\rho_{ij}).$$

First, we will study conformity under $\mathscr{B}$. Fix $k := \min\{\,\tilde{k}\in\mathbb{N}:\ \tilde{k} \geq (1-\bar{g})^{-2} - \underline{g}(1-\underline{g})^{-2}\,\}$ as in the proof of Proposition 4.2. For every $\varepsilon,\nu>0$ there exists $\bar{n}$ such that for all $n>\bar{n}$ we can find a k-set of users $\{1,\dots,k\}$ with

$$\mathbb{P}\Big[\rho_{ij} \geq 1-\varepsilon\Big] \geq 1-\nu \quad \text{for each } j=1,\dots,k.$$

Likewise, for every $\delta,\varphi>0$ there exists $\tilde{n}$ such that for all $n>\tilde{n}$ there is a user l with

$$\mathbb{P}\Big[\rho_{il} \leq -1+\delta\Big] \geq 1-\varphi.$$

Fix any engagement e_i and take $n \geq \max\{\bar{n},\tilde{n}\}$. Define the $\mathscr{B}$ prefix by

$$\mathscr{B}_i^{e_i} = \{\,l, 1,\dots,e_i-1\,\},$$

where the indices in $\{1,\dots,e_i-1\}$ are chosen from the k-pool (so $e_i-1\leq k$). Then, as $n\to\infty$,

$$\operatorname{plim}\rho_{ij} = 1 \quad \text{for each } j\in\{1,\dots,e_i-1\}, \qquad \operatorname{plim}\rho_{il} = -1.$$

Using the law of iterated expectations and decomposing $\mathscr{B}$ into the "closest block" plus the out-group user l,

$$\mathbb{E}\left[\sum_{j\in\mathscr{B}_i^{e_i}}(\theta_i-\theta_j)^2\,\middle|\,\mathscr{B}\right] = \mathbb{E}\left[\sum_{j\in\mathscr{P}_i^{e_i-1}}(\theta_i-\theta_j)^2\,\middle|\,\mathscr{P}\right] + \mathbb{E}\Big[(\theta_i-\theta_l)^2\Big]$$
$$= 2\sigma^2\,\mathbb{E}\left[\sum_{j\in\mathscr{P}_i^{e_i-1}}\Big(1-\rho_{ij}\Big)\,\middle|\,\mathscr{P}\right] + 2\sigma^2\,\mathbb{E}[1-\rho_{il}].$$

By bounded convergence, the first expectation tends to 0 (since $\rho_{ij} \to 1$ for all j in the closest block), while the second tends to $4\sigma^2$ (since $\rho_{il} \to -1$). Hence, as $n \to \infty$,

$$\mathbb{E}\left[\sum_{j \in \mathscr{B}_i^{e_i}} (\theta_i - \theta_j)^2 \,\middle|\, \mathscr{B}\right] = \underbrace{o(1)}_{\text{closest block}} + \underbrace{4\sigma^2}_{\text{out-group } l}.$$

Because Proposition 4.2 implies that the conformity loss under $\mathscr{P}$ vanishes, the two algorithms differ asymptotically by a constant $4\sigma^2$.

Now we study learning under $\mathscr{B}$. Under $\mathscr{P}$, Proposition 4.2 (Prop. 4.2) shows that learning vanishes asymptotically:

$$\mathrm{Var}(\theta \,|\, \{\theta_j : j \in \mathscr{P}_i^{e_i}\}) \xrightarrow[n\to\infty]{} \sigma^2.$$

Intuitively, the feed becomes perfectly homogeneous, so observing additional similar signals conveys no new information.

Under $\mathscr{B}$, by contrast, user i also observes one maximally dissimilar signal θ_l with $\rho_{il} \to -1$. In a two-signal Gaussian system (θ_i, θ_l) with correlation ρ_{il} and variance σ^2, the posterior variance of θ given both signals equals

$$\mathrm{Var}(\theta \mid \theta_i, \theta_l) = \frac{\sigma^2(1+\rho_{il})}{2}.$$

As $\rho_{il} \to -1$, the denominator $1 + \rho_{il} \to 0$, so the precision diverges and

$$\mathrm{Var}(\theta \mid \theta_i, \theta_l) \xrightarrow[n\to\infty]{} 0.$$

Since adding further (highly correlated) signals cannot increase posterior variance,

$$\mathrm{Var}(\theta \,|\, \{\theta_j : j \in \mathscr{B}_i^{e_i}\}) \le \mathrm{Var}(\theta \mid \theta_i, \theta_l) \xrightarrow[n\to\infty]{} 0.$$

Thus, learning under $\mathscr{B}$ becomes asymptotically perfect.

To finish, let's compare utilities and find the λ that makes the user indifferent. The only asymptotically nonvanishing differences between $\mathscr{P}$ and $\mathscr{B}$ are: (i) a constant conformity penalty of $4\sigma^2$ under $\mathscr{B}$, (ii) the learning gain under $\mathscr{B}$, and (iii) the difference in expected engagement. Hence, $\mathscr{B}$ yields higher expected utility if and only if

$$\lambda\Big(\alpha\big(\mathbb{E}[e_i \mid \mathscr{P}] - \mathbb{E}[e_i \mid \mathscr{B}]\big) + 4\sigma^2\Big) \;\le\; (1-\lambda)\,\sigma^2,$$

that is,

$$\lambda \;\le\; \frac{\sigma^2}{\alpha\big(\mathbb{E}[e_i \mid \mathscr{P}] - \mathbb{E}[e_i \mid \mathscr{B}]\big) + 5\sigma^2}.$$

□

Proof of Proposition 6.1

Proof. Define

$$\Delta e_i := \mathbb{E}\Big[e_i(n+1) - e_i(n) \,\Big|\, \theta_i, \mathscr{P}(n)\Big],$$

$$\Delta C_i := \mathbb{E}\left[\sum_{j \in \mathscr{P}_i^{e_i(n+1)}(n+1)} (\theta_i - \theta_j)^2 - \sum_{j \in \mathscr{P}_i^{e_i(n)}(n)} (\theta_i - \theta_j)^2 \,\middle|\, \theta_i, \mathscr{P}(n)\right],$$

$$\Delta V_i := \mathbb{E}\Big[\mathrm{Var}\Big(\theta \,\Big|\, \boldsymbol{\theta}_{\mathscr{P}(n+1)}\Big) - \mathrm{Var}\Big(\theta \,\Big|\, \boldsymbol{\theta}_{\mathscr{P}(n)}\Big) \,\Big|\, \theta_i, \mathscr{P}(n)\Big].$$

From the utility definition,

$$\Delta U_i = \lambda\Big(\alpha\, \Delta e_i - (1-\beta)\, \Delta C_i\Big) \;-\; (1-\lambda)\, \Delta V_i.$$

Whenever $\alpha\, \Delta e_i - (1-\beta)\, \Delta C_i + \Delta V_i \neq 0$, the condition $\Delta U_i \geq 0$ is equivalent to

$$\lambda \;\geq\; \frac{\Delta V_i}{\alpha\, \Delta e_i - (1-\beta)\, \Delta C_i + \Delta V_i}.$$

If the denominator is zero, then no threshold is needed. We now document the signs under the platform-optimal algorithm.

Engagement. Notice that the expected engagement always increases when platform size grows. So $\Delta e_i \geq 0$.

Conformity (sum). Going from n to $n+1$ under $\mathscr{P}$ either leaves the feed unchanged or replaces the least–correlated member k (with correlation ρ_{ik}) by an entrant $(n+1)$ with $\rho_{i(n+1)} \geq \rho_{ik}$. Using $\mathbb{E}[(\theta_i - \theta_j)^2] = 2\sigma^2(1-\rho_{ij})$ and the fact that all other users are common to both feeds,

$$\mathbb{E}_i\left[\sum_{j \in \mathscr{P}_i^e} (\theta_i - \theta_j)^2\right] \text{ weakly decreases, hence } \quad \Delta C_i \leq 0,$$

with strict < 0 if the replacement is strict. Therefore $-\lambda(1-\beta)\, \Delta C_i \geq 0$.

Learning (variance). It is convenient to decompose the change into: (i) closest replacement at fixed e_i, and (ii) any increase in e_i. For (i), replacing a signal by a weakly more informative one (larger $|\rho|$) weakly reduces posterior variance in the Gaussian setting; for (ii), when e_i rises, the added posts can only help if they are informative, but in general the net effect depends on the informativeness and correlation structure of the added signals. Thus ΔV_i can, in principle, have either sign.

Putting these pieces together:

$$\alpha\, \Delta e_i - (1-\beta)\, \Delta C_i \;\geq\; 0 \quad \text{and} \quad \Delta V_i \;\text{ is potentially sign–ambiguous.}$$

Hence, inside the platform utility is always increasing while the effect on learning is ambiguous. The inequality above gives the exact λ-threshold that guarantees $\Delta U_i \geq 0$. If one restricts attention to the Gaussian case with closest replacement and (when

applicable) at least one informative added post, then $\Delta V_i \leq 0$ and the right–hand side is ≤ 0; in that case $\Delta U_i \geq 0$ for all $\lambda \in [0,1]$, with strict improvement whenever at least one of the weak inequalities is strict. □

Proof of Proposition 6.2

Proof. Fix user i who chooses engagement e_i. By truth-telling, the sincerity term vanishes at $m_i = \theta_i$ and is unaffected by the pool size. For n large enough so that the optimum is interior ($e_i^* < n$),

$$\mathbb{E}[u_i \mid \mathscr{F}] = \lambda\Big[\alpha e_i - (1-\beta) \sum_{j \in \mathscr{F}_i^{e_i}} \mathbb{E}\Big((\theta_i - \theta_j)^2 \mid \mathscr{F}\Big)\Big].$$

Under the standing assumptions (same variance and Gaussian signals with the usual linear conditional mean), the expected conformity cost is

$$\mathbb{E}\Big((\theta_i - \theta_j)^2 \mid \mathscr{F}\Big) = 2\sigma^2\Big(1 - \rho_{ij}\Big),$$

so

$$\mathbb{E}[u_i \mid \mathscr{F}] = \lambda\Big[\alpha e_i - 2(1-\beta)\sigma^2 \sum_{j \in \mathscr{F}_i^{e_i}} \Big(1-\rho_{ij}\Big)\Big] = \lambda\Big[\Big(\alpha - 2(1-\beta)\sigma^2\Big)e_i + 2(1-\beta)\sigma^2 \sum_{j \in \mathscr{F}_i^{e_i}} \rho_{ij}\Big].$$

Consider adding one more user j to the feed. The marginal change in expected inside the platfor utility is

$$\Delta u_j = \lambda\Big[\alpha - 2(1-\beta)\sigma^2\Big(1-\rho_{ij}\Big)\Big] = \lambda\Big[\alpha - 2(1-\beta)\sigma^2 + 2(1-\beta)\sigma^2\rho_{ij}\Big].$$

Hence $\Delta u_j > 0$ iff

$$\rho_{ij} \;>\; \rho^\star \qquad \text{where} \qquad \rho^\star := 1 - \frac{\alpha}{2(1-\beta)\sigma^2}.$$

Note that if $\alpha \geq 4(1-\beta)\sigma^2$ then $\Delta u_j > 0$ for every $j \in [-1,1]$; otherwise, only sufficiently correlated users (those with $\rho_{ij} > \rho^\star$) are marginally valuable.

By the standing full-support assumption on pairwise covariances (equivalently on $\rho_{ij} \in [-1,1]$), for any fixed $\rho^\star < 1$ the probability of encountering users with $\rho_{ij} > \rho^\star$ is positive and increases with platform size n. Consequently, the expected number of users with positive marginal contribution grows with n, and the platform can form feeds that include such users. Because the platform maximizes engagement rather than learning, it need not always select the most correlated users; therefore, equilibrium feeds need not converge to perfectly correlated signals, and learning need not vanish even in large platforms. In fact, learning will improve as platform size increases, which implies that an increase in n will improve both inside the platform utility and learning. □

B Example

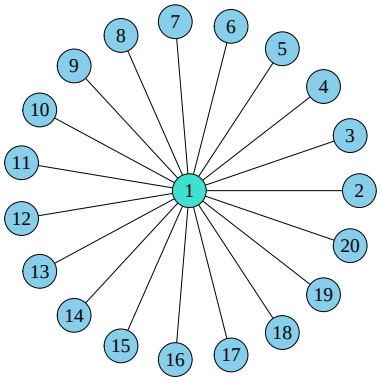

Figure 1: Platform size $n = 20$.

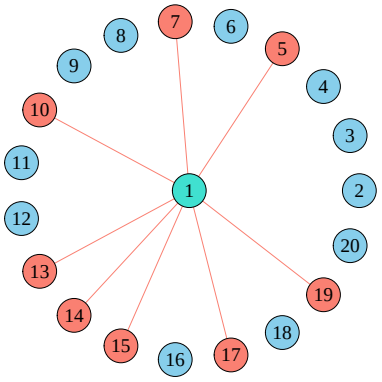

Figure 2: Platform-optimal feed.

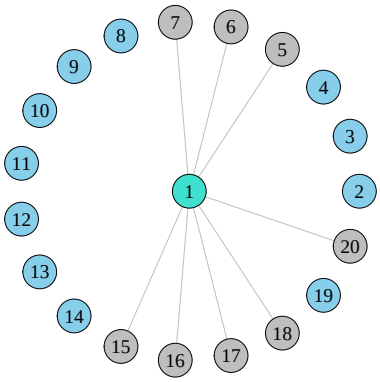

Figure 3: Reverse-chronological feed.

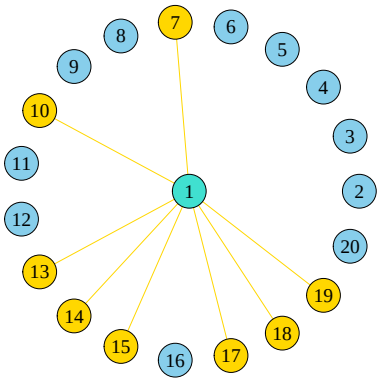

Figure 4: User-optimal feed.

Here we present the feeds user 1 would observe in a platform of size $n = 20$ (Figure 1) with similarity matrix $\mathbf{\Sigma}$ as displayed below. We fix parameters to $\alpha = 0.001$, $\lambda = 0.5$ and $\beta = 0.2$.

Platform-optimal feed. Displayed in Figure 2, it order users as 7, 10, 14, 13, 15, 19, 5, 17, 3, 9, 6, 16, 11, 8, 4, 12, 2, 20, 18, producing an expected engagement (from the platform's point of view) of 8.14 that we approximate to 8.

User-optimal feed. Displayed in Figure 4, it is generated by an algorithm that maximizes user 1's expected utility. The order is 18, 7, 10, 14, 13, 15, 19, 17, 5, 3, 6, 11, 9, 16, 8, 4, 12, 2, 20 and induces expected engagement of 7.84, that we round

to 8. This algorithm swaps user 5 and user 18, which is the least correlated one to user 1. This is precisely what the breaking-echo-chambers algorithm would do, and then we observe here how similar both of them are.

Reverse-chronological feed. Displayed in Figure 3, it is a random feed for user 1. The order is given by 7, 5, 20, 15, 6, 16, 17, 18, 19, 10, 3, 11, 14, 8, 4, 2, 13, 9, 12. Engagement is high in this realization (namely, 7.81) just because user 7 is at the top of the feed. However, the utility provided by this algorithm is substantially lower than that of the other algorithms.

$$
\boldsymbol{\Sigma} = \begin{pmatrix}
5 & -0.288 & 0.099 & -0.092 & 0.255 & 0.037 & 1.047 & -0.085 & 0.051 & 0.775 & -0.023 & -0.261 & 0.651 & 0.689 & 0.386 & 0.009 & 0.118 & -1.478 & 0.327 & -0.483 \\
-0.288 & 5 & -0.007 & -0.028 & -0.021 & -0.012 & 0.072 & -0.022 & -0.033 & 0.048 & -0.005 & -0.05 & 0.042 & 0.045 & 0.021 & -0.033 & 0.006 & -0.183 & 0.001 & -0.086 \\
0.099 & -0.007 & 5 & 0.026 & 0.123 & 0.035 & 0.186 & 0.017 & 0.082 & 0.149 & 0.003 & 0.018 & 0.122 & 0.126 & 0.08 & 0.07 & 0.025 & -0.099 & 0.1 & 0.02 \\
-0.092 & -0.028 & 0.026 & 5 & 0.071 & 0.017 & 0.165 & -0.001 & 0.037 & 0.126 & -0.001 & -0.015 & 0.105 & 0.11 & 0.065 & 0.028 & 0.02 & -0.167 & 0.068 & -0.034 \\
0.255 & -0.021 & 0.123 & 0.071 & 5 & 0.097 & 0.521 & 0.046 & 0.224 & 0.415 & 0.008 & 0.047 & 0.339 & 0.353 & 0.223 & 0.19 & 0.07 & -0.288 & 0.277 & 0.05 \\
0.037 & -0.012 & 0.035 & 0.017 & 0.097 & 5 & 0.163 & 0.011 & 0.061 & 0.128 & 0.002 & 0.007 & 0.105 & 0.11 & 0.068 & 0.052 & 0.021 & -0.109 & 0.082 & 0.003 \\
1.047 & 0.072 & 0.186 & 0.165 & 0.521 & 0.163 & 5 & 0.117 & 0.389 & 0.46 & 0.023 & 0.185 & 0.37 & 0.38 & 0.259 & 0.344 & 0.082 & 0.042 & 0.381 & 0.277 \\
-0.085 & -0.022 & 0.017 & -0.001 & 0.046 & 0.011 & 0.117 & 5 & 0.022 & 0.089 & -0.001 & -0.014 & 0.074 & 0.078 & 0.045 & 0.016 & 0.014 & -0.127 & 0.046 & -0.03 \\
0.051 & -0.033 & 0.082 & 0.037 & 0.224 & 0.061 & 0.389 & 0.022 & 5 & 0.306 & 0.003 & 0.01 & 0.251 & 0.262 & 0.162 & 0.116 & 0.051 & -0.277 & 0.191 & -0.003 \\
0.775 & 0.048 & 0.149 & 0.126 & 0.415 & 0.128 & 0.46 & 0.089 & 0.306 & 5 & 0.017 & 0.137 & 0.308 & 0.318 & 0.214 & 0.269 & 0.068 & -0.01 & 0.308 & 0.202 \\
-0.023 & -0.005 & 0.003 & -0.001 & 0.008 & 0.002 & 0.023 & -0.001 & 0.003 & 0.017 & 5 & -0.004 & 0.014 & 0.015 & 0.009 & 0.002 & 0.003 & -0.028 & 0.008 & -0.008 \\
-0.261 & -0.05 & 0.018 & -0.015 & 0.047 & 0.007 & 0.185 & -0.014 & 0.01 & 0.137 & -0.004 & 5 & 0.115 & 0.122 & 0.069 & 0.003 & 0.021 & -0.259 & 0.059 & -0.084 \\
0.651 & 0.042 & 0.122 & 0.105 & 0.339 & 0.105 & 0.37 & 0.074 & 0.251 & 0.308 & 0.014 & 0.115 & 5 & 0.256 & 0.173 & 0.222 & 0.055 & 0.004 & 0.251 & 0.171 \\
0.689 & 0.045 & 0.126 & 0.11 & 0.353 & 0.11 & 0.38 & 0.078 & 0.262 & 0.318 & 0.015 & 0.122 & 0.256 & 5 & 0.178 & 0.232 & 0.057 & 0.013 & 0.26 & 0.181 \\
0.386 & 0.021 & 0.08 & 0.065 & 0.223 & 0.068 & 0.259 & 0.045 & 0.162 & 0.214 & 0.009 & 0.069 & 0.173 & 0.178 & 5 & 0.142 & 0.038 & -0.028 & 0.168 & 0.099 \\
0.009 & -0.033 & 0.07 & 0.028 & 0.19 & 0.052 & 0.344 & 0.016 & 0.116 & 0.269 & 0.002 & 0.003 & 0.222 & 0.232 & 0.142 & 5 & 0.044 & -0.261 & 0.165 & -0.013 \\
0.118 & 0.006 & 0.025 & 0.02 & 0.07 & 0.021 & 0.082 & 0.014 & 0.051 & 0.068 & 0.003 & 0.021 & 0.055 & 0.057 & 0.038 & 0.044 & 5 & -0.011 & 0.053 & 0.03 \\
-1.478 & -0.183 & -0.099 & -0.167 & -0.288 & -0.109 & 0.042 & -0.127 & -0.277 & -0.01 & -0.028 & -0.259 & 0.004 & 0.013 & -0.028 & -0.261 & -0.011 & 5 & -0.148 & -0.428 \\
0.327 & 0.001 & 0.1 & 0.068 & 0.277 & 0.082 & 0.381 & 0.046 & 0.191 & 0.308 & 0.008 & 0.059 & 0.251 & 0.26 & 0.168 & 0.165 & 0.053 & -0.148 & 5 & 0.077 \\
-0.483 & -0.086 & 0.02 & -0.034 & 0.05 & 0.003 & 0.277 & -0.03 & -0.003 & 0.202 & -0.008 & -0.084 & 0.171 & 0.181 & 0.099 & -0.013 & 0.03 & -0.428 & 0.077 & 5
\end{pmatrix}
$$