Discussion Paper Series – CRC TR 224

Discussion Paper No. 580
Project B 05

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

Miguel Risco[1]
Manuel Lleonart-Anguix[2]

December 2024
(First version: July 2024)

[1]University of Bonn, Email: risco@uni-bonn.de
[2]The Autonomous University of Barcelona, Email: manuel.lleonart@bse.eu

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Collaborative Research Center Transregio 224 - www.crctr224.de
Rheinische Friedrich-Wilhelms-Universität Bonn - Universität Mannheim

Feed for good? On the effects of personalization algorithms in social media platforms

Miguel Risco[*] Manuel Lleonart-Anguix[†]

December 6, 2024

Abstract

This paper builds a theoretical model of communication and learning on a social media platform, and describes the algorithm an engagement-maximizing platform implements in equilibrium. This algorithm overexploits similarities between users, locking them in echo chambers. Moreover, learning vanishes as platform size grows large. As this is far from ideal, we explore alternatives. The reverse-chronological algorithm that social platforms reincorporated after the DSA was enacted turns out to be insufficient, so we construct the "breaking-echo-chambers" algorithm, which improves learning by promoting opposite viewpoints. Finally, we advocate for horizontal interoperability as a regulatory measure to align platform incentives with social welfare. By eliminating platform-specific network effects, interoperability incentivizes platforms to adopt algorithms that maximize user well-being.

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

JEL Codes: D43, D85, L15, L86.

[*]IGIER, Università Bocconi. miguel.riscobermejo@unibocconi.it

[†]Universitat Autònoma de Barcelona. manuel.lleonart@bse.eu

1 Introduction

On May 25, 2024, a video on TikTok went viral, exposing a striking disparity in how Instagram comments were displayed to two users viewing the same post: Eli, the uploader, and her boyfriend.[1] The post, which was public, depicted a girl recording herself every half hour as she waited for her boyfriend, who was late for their meeting after a golf game. Under the *most relevant* comments section, Eli saw remarks reflecting her frustration, such as "oh, this is so rude!" and "it is a disregard for her time". In contrast, her boyfriend, who opened the post at the same time, encountered drastically different comments: "or you could get your own hobby instead of waiting around for him", "he meant 3 a.m., he is ahead of schedule", and "God forbid he has a good time". Eli remarked in the video, "the only difference in how we interact with Instagram is that he is a man, and I am a woman". She attempted to locate the comments her boyfriend had seen but found they were absent from her feed.

The video, which amassed nearly three million views, added to the growing public debate about the effects of personalized social media feeds on users' beliefs and perspectives. While entertainment may be the primary reason users log into platforms like TikTok or Instagram, these platforms have become pivotal sources of news and opinion formation. Indeed, more than half of U.S. adults now receive news from social media, with the share growing significantly in recent years: between 2020 and 2024, the proportion of adults regularly accessing news on TikTok rose from 3% to 17%, and on Instagram, from 11% to 20%.[2] This underscores the crucial role of social media platforms as distributors of information, shaping the way individuals learn and form beliefs. Concerns over these platforms' societal impacts are far from new. Since the 2016 U.S. presidential election, public scrutiny of algorithmic personalization has intensified, particularly regarding its potential to foster polarization, spread misinformation, and amplify echo chambers and hate speech (Silverman, 2016; Allcott and Gentzkow, 2017). Empirical research has since provided compelling evidence of these harms (Levy, 2021; Allcott et al., 2022; Braghieri et al., 2022; Bursztyn et al., 2023) and how social media platforms trap users in their echo chambers (Levy, 2021). In particular, Bursztyn et al. (2023) show that users would be willing to pay to have others, including themselves, deactivating their TikTok and Instagram accounts. Internal documents from Meta, revealed in the *The Facebook Files* investigation (Horwitz et al., 2021), confirm that platforms are aware of these harmful effects, particularly on vulnerable groups such as teenage girls. Despite these issues, it is essential to acknowledge the significant benefits of social media, including access to information, opportunities for social connection, and professional networking (Allcott et

1 The video is public and can be accessed at https://www.tiktok.com/@elieli0000/video/7373012517016079649?lang=es.

2 Data from the *Pew Research Center*, available at https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/.

al., 2020; Armona, 2023).

The *feed* is a customized scroll of friends' content and news stories that appears on most social media platforms. Until around 2015, it was reverse-chronological.[3] Now, a proprietary algorithm controls what appears on the screen, based on user behavior on the platform. Since platforms' revenues come from advertising, maximizing profits is precisely maximizing engagement, which may not align with promoting informative communication. If, as Eli's video shows, a biased set of comments will maximize the probability you stay on the platform longer, this is what you will receive.[4] Personalized algorithms account for the increase in engagement and addictive behavior in social media platforms, regardless of the field (Guess et al., 2023).

The approval of the Digital Services Act (DSA) by the European Commission in 2022 marks one of the first significant regulatory efforts to address issues related to algorithmic personalization. Specifically, the DSA mandates that very large online platforms "provide at least one option for each of their recommender systems which is not based on profiling",[5] offering users an alternative to personalized content. In response, many social media platforms have reinstated reverse-chronological feeds as the alternative. However, the mere availability of a reverse-chronological feed does not appear to be alleviating any of the urgent media-related problems society faces. Moreover, personalization need not be detrimental to social welfare; it could be used for an improvement, as the following quote illustrates (Lauer, 2021): "If Facebook employed a business model focused on efficiently providing accurate information and diverse news, rather than addicting users to highly engaging content within an echo chamber, the algorithmic outcomes would be very different". To achieve this, however, we would need to find a way to align the platform's incentives with social well-being, so that naturally, the optimal algorithm for the platform would also be optimal for the users.

With all this in mind, we claim that there is a need for theoretical research that guides the optimal regulatory approach, understanding the incentives of platforms in designing their optimal algorithm and how they would respond to regulation. Hence, it is crucial to understand the strategic interplay between an engagement-maximizing platform and users who value not just the instantaneous joy (coming from scrolling down and posting their thoughts) that induces them to stay online longer, but also the reward from learning.

3 Social media platforms began transitioning from reverse-chronological feeds to personalized feeds at different times. Facebook started implementing personalized feeds in 2009, while Twitter (now X) and Instagram transitioned between 2015 and 2016. Younger platforms, like TikTok, have provided curated content since their launch.

4 See https://blog.x.com/engineering/en_us/topics/open-source/2023/twitter-recommendation-algorithm, where X sketches the functioning of its algorithm and explains how it relies on *RealGraph*, a predictor of user interaction described in Kamath et al. (2014).

5 Instagram, Facebook, X, TikTok, YouTube and others are classified as very large online platforms as they serve more than 45 million users.

This is precisely what this paper intends to achieve, having in mind the relation that users establish with platforms as Instagram, TikTok or X.

We build a theoretical model where users post messages and learn through a feed designed by the engagement-maximizing platform. We assume that users derive instantaneous utility from engaging in communication with peers about some underlying topic, and we call this utility stream *within-the-platform utility*. It has three channels: satisfaction is brought by reading posts, expressing one's own views (in the sense of being loyal to own innate opinions; *sincerity*), and conforming with the rest (in the sense of matching the opinions that others have shared; *conformity*).[6] This model choice is taken from Galeotti et al. (2021), where agents prefer taking actions closer to those of their neighbors and to their own ideal points. One of their motivating examples fits our model: the message "may be declaring political opinions or values in a setting where it is costly to disagree with friends, but also costly to distort one's true position from the ideal point of sincere opinion". The strength of these incentives depends on model parameters. In particular, we encompass situations in which conformity is almost negligible. The second utility stream comes from gathering valuable information on the platform to improve a decision, termed *action utility*. The effectiveness of the Covid-19 vaccine, which triggered significant public debate,[7] is our leading example, as each part of the previously described utility function can be easily identified. People, driven by both a desire for sincerity and conformity, used social media to express their views about the benefits and risks of vaccination. Note that individuals sought to communicate their personal viewpoints because vaccination was a pivotal societal concern, but at the same time expressing dissenting opinions proved to be socially taxing. Additionally, gathering information was crucial to deciding whether to get vaccinated or not.

In our model, engagement is defined as the number of posts a user reads. It is equivalent to the time spent scrolling down before logging out. Crucially, we assume that users do not decide how much to engage, but that their engagement is controlled by a stochastic process driven by the instantaneos joy of consuming content. After reading each post, a user continues scrolling down with some probability depending on the within-the-platform utility instantaneously derived. Otherwise, she logs out. Scrolling is, then, seen as an addictive behavior the user does not control: it is a rather automatic process corresponding with the intrinsic happiness derived within the platform.[8] However, the explicit decision

6 Conformity is a driving-force in social media behavior (Mosleh et al., 2021). It is defined as the act of matching attitudes, beliefs and behaviors to group norms (Cialdini and Goldstein, 2004). Here we treat conformity as a behavioral bias included at the outset, but it has been widely found as a product of rational models. See Bernheim (1994) for a theory of conformity and Chamley (2004) for an overview.

7 Loomba et al. (2021) find that the acceptance of the Covid-19 vaccine in US and UK declined an average of 6 percentage points due to misinformation.

8 This assumption could be also interpreted following the *Dual Process Theory* as in Benhabib and

to post a message is regarded as a rational move in which the user consciously acts to maximize her utility. This aligns with the evidence presented by Allcott et al. (2022), which shows that people are well aware of habit formation yet consume social media as if they are inattentive to it. Thus, users post messages and then observe those that appear in their feeds until they log out. Afterwards, they take an action based on the information gathered. Feeds are the product of an algorithm designed by the platform which, as explained earlier, has no incentives to promote learning: engagement purely depends on within-the-platform utility. The platform, which does not read messages, designs the algorithm leveraging its information on users' similarities in views. We assume the platform knows perfectly how similar users are, and utilizes this information to maximize total engagement. We think of similarities being derived from past interactions and users' personal data by using sophisticated machine learning techniques.[9]

Beyond the model itself, this paper makes three key contributions. The first one is to identify the platform-optimal algorithm and study its properties. As expected, the platform-optimal algorithm is driven by the desire to maximize expected conformity, because it is the main force behind engagement. However, the fact that each user knows her own signal but the platform does not creates an information friction and the platform overexploits similarity in the feeds. The user would have preferred to have more diverse views, but is locked in an echo chamber, precisely as Eli shows in her video. This excess of similarity in the feeds becomes more pronounced as the platform size grows. Then, the feed becomes flooded with close *copies* of a user and consequently learning vanishes, contrasting with classical results where large societies learn better (Golub and Jackson, 2010).

The second contribution consists in studying alternatives to the platform-optimal algorithm. We start with the reverse-chronological algorithm, which platforms reinstated following the enactment of the DSA, and show that it is generally not sufficient to be considered a suitable alternative. This finding supports a quote from Aridor et al. (2024): "Clearly, going back to reverse-chronological ordered feeds is not viable, as platforms derive profit and users derive utility from algorithmically curated content". Hence, we analyze a variation of the platfom-optimal algorithm that maximizes learning when platform size is large. This is the breaking-echo-chambers algorithm, which, inspired in mechanisms of *crowd-sourced fact-checking* adds a user with opposite views at the top of each feed induced by the platform-optimal algorithm. It improves learning but slightly decreases

Bisin (2005). For an overview on Dual Process Theory, see Grayot (2020). The fact that engagement depends only on within-the-platform utility can be seen under the light of present bias: the user weights disproportionally low the benefits of learning when reading posts.

9 Facebook's FBLearner Flow, a machine learning platform, is able to predict user behavior through the use of personal information collected within the platform. See Biddle (2018) for a news piece on it. The early paper Kosinski et al. (2013) already showed that less sophisticated techniques could predict a wide range of personal attributes by just using data on "likes".

conformity and consequently engagement. While it still outperforms the platform-optimal algorithm for many types of users, it is plausible that real-world individuals may disregard information from a completely opposing source, complicating its practical implementation.

Regulating the market structure is not only a more practical policy recommendation than intervening directly at the algorithmic level, but the utilitarian optimal algorithm also emerges as the only one that maximizes social welfare. This observation leads to the third and final contribution of this paper: a discussion of whether mandating horizontal interoperability would suffice to incentivize social media platforms to adopt utilitarian optimal algorithms through competition. While horizontal interoperability has been a topic in policy debates in recent years as a means to enhance market contestability, we approach the issue from a different perspective. We propose horizontal interoperability as an ideal mechanism to align platform incentives with social welfare, thereby fostering the implementation of healthier algorithms.

Without horizontal interoperability, the network effects that social media platforms feature (i.e., the fact that the more users join a platform, the more valuable its service becomes) create high barriers to entry and induce winner-takes-all (or most) market dynamics. Market tips and a large incumbent dominates. Horizontal interoperability compels platforms to connect, so that users from different platforms can interact.[10] A user's feed would then be an ordered list of the posts coming from all her friends, regardless of which platform they are registered on, designed by the specific algorithm of the platform the user joined. Crucially, network effects will be shared, and platforms will have to compete along the non-interoperable dimension, i.e., they will have to compete in the utility that algorithms offer to users. Each user will choose the platform whose algorithm offers the highest expected utility, disregarding platform size. And this algorithm is, of course, the utilitarian optimal algorithm. Then, competing platforms would be *forced* to implement this algorithm; otherwise, they risk losing their user base. This is precisely why we argue that, in the particular case of social media platforms, horizontal interoperability would not only be *necessary* for raising social welfare (as it is argued to be in any digital market with strong network effects (Kades and Scott Morton, 2020)), but also *sufficient*. However, although the Digital Markets Act introduces horizontal interoperability for providers of number-independent interpersonal communication services (NI-ICS), such as WhatsApp, Telegram, or Signal, through Article 7, social media platforms are not yet subject to any regulation in this regard.

10 For a general intuition of the benefits of interoperability in digital markets, we quote Kades and Scott Morton (2020): "Interoperability eliminates or lowers the entry barrier, which is the anticompetitive advantage the platform has maintained and exploited. Users will not switch to a new social network until their friends and families have switched. [...] Interoperability causes network effects to occur at the market level—where they are available to nascent and potential competitors—instead of the firm level where they only advantage the incumbent."

The rest of the paper is organized as follows. After the literature review, each section corresponds to each of the contributions described above: Section 2 develops the model, Section 3 finds the platform-optimal algorithm and characterizes it, Section 4 analyzes alternative algorithms, and Section 5 discusses horizontal interoperability. Finally, Section 6 concludes.

1.1 Related literature

To the best of our knowledge, this paper provides one of the first theoretical models examining personalization algorithms in social media platforms. It is crucial for the social media literature to understand the effects an strategic platform seeking to maximize engagement can have on social learning and social welfare by ordering the posts a user will potentially read. This theoretical paper complements recent empirical research by Guess et al. (2023), which explores the effects of Facebook's and Instagram's algorithms. In particular, their findings are consistent with our theoretical predictions: transitioning users to a reverse-chronological feed reduces both platform usage and engagement, while decreasing the prevalence of ideologically similar content and mitigating echo chamber effects. In our model, echo chambers arise as a deliberate outcome of the platform's strategy, which overexploits user preferences for similarity. This result aligns with the hypothesis advanced by Pariser (2011) and debated by Levy (2021), who contest the prevalence of "filter bubbles".

Our work contributes mainly to the literature on the impact of profit-maximizing platforms on social welfare, particularly in relation to social learning. Acemoglu et al. (2023) is the closest to us: in a model where agents decide whether or not to pass on (mis)information, a social media platform creates more homophilic patterns to increase engagement. Our work is different from theirs in two important dimensions. First, while their model is about sequential share of (mis)information and agents care about their reputation, we include social learning in the presence of simultaneous communication where each additional message read weakly increases knowledge. Second, their platform favors homophily to increase article sharing, as like-minded users are less likely to scrutinize misinformation. In contrast, we explain the excessive formation of echo chambers through information frictions, even when platform and user preferences are aligned. Similar theoretical findings about the harmful effects of profit-driven platforms on user well-being are seen in Beknazar-Yuzbashev et al. (2024), where, under the assumption of harmful content being complementary to time spent, the platform finds it optimal to expose users to it, and in Mueller-Frank et al. (2022), where the platform manipulates information flow to increase revenue even if this ultimately decreases social welfare.

We also provide theoretical grounding for claims that recommendation systems connect users to content that reinforces their existing beliefs to drive engagement, a phenomenon previously discussed by scholars such as Sunstein (2017) and empirically verified by Holtz et al. (2020). Additional research on the economic motivations for platforms to distort content can be found in Reuter and Zitzewitz (2006), Ellman and Germano (2009), Abreu and Jeon (2019), and Kranton and McAdams (2022).

Lastly, this paper engages with the growing literature on interventions and regulations in social media platforms. For instance, Jackson et al. (2022) analyze how limiting the breadth and/or depth of social networks can enhance message accuracy, while Guriev et al. (2023) investigate measures to combat misinformation on platforms like Twitter. Our proposed breaking-echo-chambers algorithm aligns in spirit with these direct interventions, but, as most of them, ultimately remains imperfect. On the market regulation side, scholars such as Kades and Scott Morton (2020) underscore the importance of horizontal interoperability as a crucial condition for fostering contestability in digital markets characterized by strong network effects. Another stream of research (Bourreau and Krämer, 2022; Bourreau et al., 2023; Dhakar and Yan, 2024) examines horizontal interoperability in the context of number-independent interpersonal communication services under the Digital Markets Act (DMA), remaining agnostic about the policy's ultimate impact, noting that the elimination of multihoming might inadvertently harm competition. However, we argue that the social media platforms market operates under distinct dynamics, where the non-interoperability of algorithms forces platforms to compete on them to capture users. Finally, for a comprehensive review of the social media literature, we refer to Aridor et al. (2024).

2 A model of communication and learning through personalized feeds

We present the baseline model of the paper, inspired by social media platforms such as X, Instagram, and TikTok. While users primarily log in to these platforms for instantaneous entertainment, they also serve as spaces for learning and information exchange, where users gather news and insights through communication with others. This section begins with an overview, proceeds with the formal details, and concludes with a discussion of the assumptions underlying the model.

There is an underlying state of the world that users aim to discover in view of a subsequent action. Joining a social media platform offers users the benefit of accessing information, as fellow users share messages related to that state of the world. However, beyond mere information retrieval, users also derive utility from engaging in non-informative interactions within the platform. Expressing personal opinions and reading others' posts brings satisfaction, yet encountering disagreement imposes a burden. We define user engagement as the measure of messages read, representing the time spent on the platform until the user discontinues browsing and exits.

Users' utility comprises two components: the *within-the-platform utility*, influenced by engagement, conformity, and sincerity, and the *action utility*, which depends on learning, i.e., how close users can get to the state of the world after communicating on the platform. The platform's revenues, in turn, are contingent upon user engagement. Hence, the platform designs an algorithm seeking to maximize such engagement by leveraging information on similarities between users' worldviews. This algorithm curates a personalized feed for each user, determining the order in which messages appear on the scrolling screen.

In our baseline model, we assume a monopolistic platform with all users already on board. Once a user logs in, she decides on which message to post. Engagement, however, is not a choice variable but follows an addictive process: after reading each message, with some probability depending on the amount of within-the-platform utility experienced so far, the user continues scrolling down, while she logs out otherwise.

Now, let us describe the model in detail. There is a set, $\mathcal{U}$, of n users aboard a social media platform. We assume that every user is a friend of all others, and hence her neighborhood is the whole user base. Users receive information on the state of the world θ in the form of a private signal $\theta_i \in \mathbb{R}$. Conditional on θ, signals $\{\theta_1, ..., \theta_n\}$ are jointly normal and their structure is given by

$$\begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix} \sim \mathcal{N}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}\right),$$

where $\boldsymbol{\theta} = (\theta, ..., \theta)$ and $\boldsymbol{\Sigma} = (\sigma_{ij})$ is an $n \times n$ symmetric and positive definite matrix. The signal θ_i is interpreted as the information the user has about the state of the world prior to her entry on the social platform. It might be based on inherent personal characteristics as well as on information collected privately. As information sources, as well as ideology, might be similar, different users' private information might be correlated. This is captured by the matrix $\boldsymbol{\Sigma}$.

Users know their private signals, the distribution of all signals, the covariance matrix $\boldsymbol{\Sigma}$, and the distribution of the state of the world, for which we crucially assume improper priors. Thus, conditional on θ_i, the posterior distributions of θ_j and θ are normal and centered on θ_i, namely $\theta_j|\theta_i \sim \mathcal{N}\left(\theta_i, \sigma_{jj} - \frac{\sigma_{ij}^2}{\sigma_{ii}}\right)$ for all $j \in N$ and $\theta|\theta_i \sim \mathcal{N}\left(\theta_i, \sigma_{ii}\right)$.

Once logged in the platform, each user i posts a message $m_i \in \mathbb{R}$ and then observes $e_i \in \mathbb{N}$ messages that appear in her personalized feed, which is provided by the platform. The number $e_i \leq n$ represents her engagement, and platform profits depend precisely on the sum of all users' engagement, $\sum_{i=1}^{n} e_i$. In order to maximize user engagement, the platform designs an algorithm consisting of an assigment that, given a pair of users i, j, tells which position user j occupies in user i's feed. Given engagement e_i, user i's feed is the set of users from whom user i will observe messages. Formally, an algorithm $\mathscr{F}$ is a collection $(\mathscr{F}_i)_{i\in\mathcal{U}}$ where $\mathscr{F}_i \in \text{Bij}\left(\{1,...,n-1\},\mathcal{U}\backslash\{i\}\right)$. Given $r \leq n$, $\mathscr{F}_i(r)$ is the r-th user in i's ranking induced by $\mathscr{F}$, so i's feed for engagement e_i is precisely

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

Users derive utility from two streams: within-the-platform utility and action utility. Their within-the-platform utility has three components: (i) a positive linear payoff coming from reading messages; (ii) *sincerity*: agents dislike deviating from their own signals,[11] and (iii) *conformity*: disagreeing with others' opinions is taxing. Formally, user i's realized within-the-platform utility is

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

where $\alpha > 0$ and the parameter $\beta \in (0,1)$ represents how much sincerity is weighted with respect to conformity. Note that, in particular, we encompass situations in which conformity is almost negligible. Within-the-platform utility is not the only source of utility for users, as they are also concerned about taking an action $a_i \in \mathbb{R}$ that matches the state of the world. Total realized utility is the weighted average of within-the-platform utility and action utility, which we define as the squared difference of the action from the state of the world:

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

where $\lambda \in (0,1)$ weights the relative importance of within-the-platform and action utilities. Summarizing, user i observes θ_i, chooses a message m_i and, after learning messages $\{m_j\}_{j\in\mathscr{F}_i^{e_i}}$, chooses an action a_i to maximize the conditional expectation of U_i.

Along the lines of digital addiction theory, we assume that the user does not control her scrolling time but, after reading k posts, 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. With probability $1 - g(u_i(k-1, m_i, m_{-i}, \mathscr{F}_i, \theta_i))$, the user discontinues scrolling down and exits. Hence, user i features engagement e_i with probability $(1 - g(u_i(e_i, m_i, m_{-i}, \mathscr{F}, \theta_i))) \prod_{r=1}^{e_i-1} g(u_i(r, m_i, m_{-i}, \mathscr{F}_i, \theta_i))$. In particular, we assume that there exists some pair $0 < \gamma < \beta < 1$ such that $\forall\ x \in \mathbb{R}$, $\gamma < g(x) < \beta$, i.e., no feed guarantees either continuation or abandonement. Because of the addictive nature of engagement, we assume that the user treats its time in the platform and something

11 Due to improper priors, sincerity would yield the same results if, instead of being punished for deviating with her message m_i from θ_i, the user were penalized for deviating from θ.

exogenous but expected to be finite. Formally, we assume that every user i takes e_i as an independent random variable with some CDF such that $\mathbb{E}_i[e_i] \leq \tilde{e}$ for some $\tilde{e} \in \mathbb{R}$, where $\mathbb{E}_i$ stands for user i's expectations.

The platform knows how engagement works and the function g, the distributions, and the matrix $\mathbf{\Sigma}$, but not θ nor $\{\theta_i\}_{i=1}^n$. It builds the algorithm $\mathscr{F}$ based on $\mathbf{\Sigma}$ to maximize $\sum_{i=1}^n \mathbb{E}_p[e_i]$ (where $\mathbb{E}_p$ stands for the platform's expectations), the sum of the expected engagement of all users. In summary, the game of *communication and learning through personalized feeds* described above consists of the following sequence of events:

1. The platform chooses an algorithm $\mathscr{F}$ and (publicly) commits to it.
2. Each user observes her private signal θ_i.
3. Each user i posts a message $m_i \in \mathbb{R}$.
4. Each user i observes e_i messages in her feed $\mathscr{F}_i^{e_i}$ and chooses an action a_i.
5. The state of the world is revealed and payoffs are realized.

We devote the last part of this section to a discussion on some of the assumptions that build the model.

Complete network. This model could be extended to any network given by some undirected graph $\mathcal{G}$. In this scenario, each user i would belong to a neighborhood $\mathcal{U}_i$ and hence $e_i \leq |\mathcal{U}_i|$. While the results presented below remain valid in such general context, we choose to work with a complete network for the sake of simplicity of notation and exposition.

Monopolistic platform, all users on board. We assume that there is only one platform, and all users are already on board. Hence, the platform does not need to care about capturing them, but only about maximizing their engagement. This is, of course, a simplifying assumption, but the main social media platforms (Facebook, Instagram, TikTok or X) are monopolists of their fields:[12] even though they can be broadly described as social media platforms that enable public posting and private communication, they differ in their core functionality. Each site dominates a specific field: photography (Instagram), short videos (TikTok), reciprocal communication with friends (Facebook), and micro-blogging (X). In most cases, there is no realistic alternative for the average user but to stay out, and then, as Bursztyn et al. (2023) show, fear of missing out makes users join even when they would prefer the platform not to exist.

Improper priors. Users' prior distribution is uniform along $\mathbb{R}$. Intuitively, this means that no user understands whether her signal is extreme. Indeed, every user believes her

12 Regarding monopoly structures in the social media platform market, 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.

opinion is central (Ross et al., 1977; Greene, 2004). This assumption is made for the sake of model tractability. Under normal priors, we can only determine the users' optimal linear messaging strategies, but we cannot derive an explicit expression for the platform-optimal algorithm.[13]

Engagement. Following the literature on digital addiction (Allcott et al., 2020, 2022; Aridor, 2024), we exclude engagement as a choice variable and instead adopt a simplified framework in which digital addiction emerges as a by-product of habit formation and self-control problems. The user continues scrolling based on the instantaneous utility experienced within the platform up to that point. This setup captures addictive behavior in a reduced form while reflecting key characteristics: the probability of engaging for k periods is always higher than for k' periods if $k < k'$, greater utility from reading a message increases the likelihood of continued engagement, and engagement is independent of the utility derived from actions. The latter feature can be intuitively understood as an extreme form of present bias: when scrolling, the user heavily discounts the long-term rewards of learning (Guriev et al., 2023).

Platform's profits as a function of total engagement. Social media platforms are typically free to access, generating revenue from advertisers who pay for product placements. These payments are directly tied to user engagement: the greater the engagement, and therefore the exposure to advertisements, the higher the payment advertisers are willing to make. To simplify, this model assumes the platform's objective is to maximize total engagement. Consequently, the platform's profit function is expressed as $\Pi_p(\mathscr{F}, \mathbf{\Sigma}) = \sum_{i=1}^{n} \mathbb{E}_p[e_i]$.

3 Platform-optimal algorithm

This section characterizes the platform's equilibrium algorithm and its implications. We demonstrate that users truthfully report their private signals irrespective of the feed-curation algorithm, and the platform, in response, designs a feed excessively driven by similarity. While this approach maximizes engagement, it exacerbates echo chambers and undermines user learning as platform size increases, eventually reducing learning to negligible levels.

The equilibrium concept in this game is Bayesian Nash Equilibrium (BNE). The platform chooses an algorithm $\mathscr{F}$, while each user i chooses a message m_i to maximize

$$\lambda \left(\alpha \mathbb{E}_i[e_i|\theta_i, \mathscr{F}] - \beta(\theta_i - m_i)^2 - (1-\beta)\mathbb{E}_i \left[\sum_{j \in \mathscr{F}_i^{e_i}} \frac{(m_i - m_j(\theta_j))^2}{e_i} |\theta_i, \mathscr{F} \right] \right),$$

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

and an action a_i (after learning the messages in her feed) to maximize

$$-(1-\lambda)\mathbb{E}_i\left[(a_i-\theta)^2|\{m_j\}_{j\in\mathscr{F}_i^{e_i}}\right].$$

In this framework, for any algorithm the platform picks, users disclose their private signals in their messages.

Proposition 3.1. *Given any algorithm $\mathscr{F}$, every user plays truthtelling in equilibrium, i.e., $m_i^* = \theta_i$ for all $i \in \mathcal{U}$.*

Proof. See Appendix A. □

Due to the use of improper priors, any other user's signal is, in expectation, equal to user i's signal, i.e., $\mathbb{E}_i[\theta_j \mid \theta_i, \mathscr{F}] = \theta_i$. Consequently, the platform cannot influence first-order moments through the feed it designs and, hence, user i believes that, in expected terms, every other user will play θ_i. Deviating from $m_i = \theta_i$ is then not profitable if everyone else is also playing truthtelling.

From the result above it directly follows that the platform can only affect user i's within-the-platform utility through the choice of her own feed $\mathscr{F}_i^{e_i}$, but not through any other user's feed. Hence, as user i's engagement depends on within-the-platform utility and not on action utility, maximizing $\sum_{i=1}^n \mathbb{E}_p[e_i]$ is for the platform equivalent to maximizing $\mathbb{E}_p[e_i]$ for each $i \in \mathcal{U}$ separately.

Corollary 3.2. *It is equivalent for the platform to maximize total user engagement and maximizing each user's individual engagement separately.*

We now outline the design of the platform-optimal algorithm $\mathscr{P}$, while the remaining formal details are left to Appendix A as part of the proof of Proposition 3.3 below. From the point of view of the platform, and because of truthtelling, user i's within-the-platform utility simplifies to $\mathbb{E}_p[u_i(r-1,\theta_i,\theta_{-i},\mathscr{F},\theta_i)] = \mathbb{E}_p[(\alpha(r-1)-(1-\beta)\frac{1}{r-1}\sum_{j\in\mathscr{F}_i^{r-1}}(\theta_i-\theta_j)^2]$ when she has read $r-1$ messages. The expected probability of staying after reading the r-th message is then $\mathbb{E}_p[g(u_i(r,\theta_i,\theta_{-i},\mathscr{F},\theta_i))]$. To maximize this probability, the platform chooses a user j to be included next in the feed among those who have not been chosen yet, i.e., $j \in \mathcal{U}\backslash\mathscr{F}_i^{r-1}$. As g is increasing in u_i, maximizing g is equivalent to maximizing u_i. Moreover, recall that conformity is the only term in which the platform can affect user i's within-the-platform utility at this stage. Hence, j is chosen according to

$$j = \underset{l\in\mathcal{U}\backslash\mathscr{F}_i^{r-1}}{\operatorname{argmax}}\{-\mathbb{E}_p[(\theta_i-\theta_l)^2]\},$$

and j is the user whose message has not yet been shown and minimizes the loss coming from conformity. Thus, the platform-optimal algorithm $\mathscr{P}$ is precisely the one which, when applied to user i, ranks other users in reverse order regarding their loss in conformity with her. In other words, for any $r \leq n$, the feed $\mathscr{P}_i^r$ shows the messages of the r users who conform the most with user i. This happens, crucially, from the perspective of the platform, which is unaware of the particular realizations of the users' private signals. In short: the algorithm $\mathscr{P}$ applied to user i induces a feed given by

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

For an explicit example of how the platform designs $\mathscr{P}$ leveraging $\mathbf{\Sigma}$, please refer to Appendix B.

Proposition 3.3. *In equilibrium, the platform chooses the algorithm $\mathscr{P}$ as specified in Equation (3). In other words, the algorithm that maximizes user engagement is the one that, for each user i, designs a feed in which others appear in reverse order regarding the expected loss in conformity with user i they induce.*

Proof. We refer to Appendix A for the remaining formal details. □

The platform's inability to observe private signals fundamentally drives our next result. On the one hand, the platform minimizes conformity loss, effectively maximizing $-\mathbb{E}_p[(\theta_i-\theta_j)^2]$ through the choice of j. But $-\mathbb{E}_p[(\theta_i-\theta_j)^2] = -\sigma_{ii}-\sigma_{jj}+2\sigma_{ij}$, so

$$j = \underset{j\in\mathcal{U}/\mathscr{F}_i^{r-1}}{\operatorname{argmax}}\{-\sigma_{jj}+2\sigma_{ij}\}.$$

On the other hand, user i, with knowledge of her own signal, would prefer $-\mathbb{E}_i[(\theta_i-\theta_j)^2|\theta_i] = -\sigma_{jj}+\frac{\sigma_{ij}^2}{\sigma_{ii}}$ to be maximized and, hence, to be matched next with some $j\in\mathcal{U}$ either very similar or very opposite to her. As the platform is less informed, it only selects very similar users to user i and, on top of that, fixes the weight of similarity to 2, when the user would rather it depend on $\frac{1}{\sigma_{ii}}$. All this drives the user to an *excessive* similarity bubble (an echo chamber), while she would prefer to observe a more diverse feed.

Proposition 3.4. *The platform overexploits similarity between users when designing its optimal algorithm.*

Proof. As indicated above, the platform selects the next user j in the feed $\mathscr{F}_i^k$ according to $j = \underset{j\in\mathcal{U}/\mathscr{F}_i^{r-1}}{\operatorname{argmax}}\{-\sigma_{jj}+2\sigma_{ij}\}$, while user i would prefer j to be selected according to

$$j = \underset{j\in\mathcal{U}/\mathscr{F}_i^{r-1}}{\operatorname{argmax}}\left\{-\sigma_{jj}+\frac{\sigma_{ij}^2}{\sigma_{ii}}\right\}.$$

□

When variances are homogeneous, meaning that each user has an equally precise posterior of θ, the feed simplifies to a pure reverse ranking based solely on similarities. In this case, users are displayed in decreasing order of their correlation with the feed viewer.

Corollary 3.5. *If variances are homogeneous, i.e., $\sigma_{ii} = \sigma_{jj} = \sigma^2$ for all $i, j \in \mathcal{U}$, the platform-optimal algorithm $\mathscr{P}$ ranks uniquely in terms of similarity.*

Proof. If variances are homogeneous, $-\mathbb{E}_p[(\theta_i - \theta_j)^2] = -2\sigma^2 + 2\sigma_{ij}$, and users are ranked following a weakly decreasing order regarding their covariance to user i (if there were ties, they would be broken randomly). Hence, $\mathscr{P}_i^1$ is the first user in the ranking, $\mathscr{P}_i^2$ includes also the second, and so on. □

Note that user i's expected conformity becomes (from her point of view) $-\mathbb{E}_i[(\theta_i - \theta_j)^2|\theta_i] = -\sigma^2 + \frac{\sigma_{ij}^2}{\sigma^2}$, while for the platform it becomes $-\mathbb{E}_p[(\theta_i - \theta_j)^2] = -2\sigma^2 + 2\sigma_{ij}$. This reveals a disparity consistent with Proposition 3.4: users are served a less diverse feed than they would prefer, even when learning is disregarded.

Next, we characterize user i's optimal action. Although optimal messages are not affected by the feed, optimal actions are. The next result provides a formal expression for user i's optimal action a_i^*.

Proposition 3.6. *For any algorithm $\mathscr{F}$, user i's optimal action after reading e_i messages is*

$$a_i^* = \frac{\mathbb{1}\boldsymbol{\Sigma}_{\mathscr{F}_i^{e_i}}^{-1}\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}^t}{\mathbb{1}\boldsymbol{\Sigma}_{\mathscr{F}_i^{e_i}}^{-1}\mathbb{1}^t},$$

where $\boldsymbol{\Sigma}_{\mathscr{F}_i^{e_i}}$ is the restriction of $\boldsymbol{\Sigma}$ to the users in $\mathscr{F}_i^{e_i}$ and $\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}^t$ is the vector of private signals of the users in $\mathscr{F}_i^{e_i}$.

Proof. User i's optimal action maximizes $\mathbb{E}_i[(a_i - \theta)^2|\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}]$. Hence, a simple first order condition yields to the optimal action being

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

by Lemma A.1. □

To understand the impact of the platform-optimal algorithm on social welfare, we must examine its effect on *learning*, which refers to how information gathered on the platform improves decision-making. Before doing so, we make an additional assumption for tractability purposes. We assume that users' variances are homogeneous, i.e., that $\sigma_{ii} = \sigma_{jj}$ for all users $i, j \in \mathcal{U}$. Thus, we are in the case described by Corollary 3.5. To clearly indicate that we are now working under homogeneous variances, we will denote the platform-optimal algorithm as $\mathscr{C}$, referring to the "closest" algorithm, as now the platform simply matches users with those who are most similar, or closest, to them.

Now, let us analyze how personalization algorithms affects learning. The scenario we study is precisely that of a large platform size, reflecting the substantial growth in social media usage in recent years.[14] Learning is defined as the increase in expected action utility resulting from reading messages. When a user picks her optimal action, its expected value is the posterior variance of θ conditional on the messages in the feed:

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

Lemma A.1 allows us to provide an explicit expression for the posterior variance:

$$\mathrm{Var}\left[\theta|\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}\right]=\frac{1}{\mathbb{1}\boldsymbol{\Sigma}_{\mathscr{F}_i^{e_i}}^{-1}\mathbb{1}^t},\tag{4}$$

which, in turn, enables us to calculate the improvement in learning, i.e., in decision-making after reading a feed, for any algorithm and any platform characterized by $\mathcal{U}$ and $\boldsymbol{\Sigma}$. Note that, in this model, the posterior variance is always weakly lower than σ^2 for each user i, meaning that learning is always weakly enhanced after reading a feed.

We now consider a scenario where the platform expands by adding new users to a given user base $\mathcal{U}$, assuming that the covariances between new and existing users are drawn from a continuous and symmetric distribution centered at zero and supported on $[-\sigma^2,\sigma^2]$. The resulting covariance matrix $\boldsymbol{\Sigma}$ is assumed to be symmetric and positive definite. As platform size increases, the mechanism $\mathscr{C}$ for feed selection becomes more apparent. To maximize expected user engagement, the platform selects the feed composed of the most similar neighbors, minimizing conformity loss. Notably, the expected feed size remains finite regardless of platform size because the platform expects user engagement to be bounded: as $g(.)\in(0,1)$, no feed guarantees continuation and the probability of the user reading an infinite amount of posts converges to zero.[15] This observation is formalized in Lemma 3.7.

Lemma 3.7. *There exists a well-defined $k\in\mathbb{N}$ such that $\mathbb{E}_p[e_i]\leq k$ for all $i\in\mathcal{U}$.*

Proof. See Appendix A. □

As a consequence of this result, the longest expected feed user i would receive is contained within $\mathscr{C}_i^k$ from the platform's perspective or $\mathscr{C}_i^{\tilde{e}}$ from the user's perspective. Considering how the platform designs its optimal algorithm, the finiteness of the feed implies that as the platform size grows and more neighbors are added, user i's feed under the closest algorithm will increasingly include individuals with higher correlations. In

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/.

15 Remember that, by assumption, each user believes their own expected engagement to be finite and bounded by $\tilde{e}$.

other words, user i will observe a feed of closely similar individuals that become even more similar as n grows. This has two significant consequences: first, user i's learning asymptotically vanishes, as formally shown in Proposition 3.8. Second, the user becomes confined to an echo chamber composed of these highly similar individuals. Notably, this outcome does not align with the user's preferences (even without considering learning) since she would favor matches that include both highly similar and diverse individuals. The latter, in particular, would significantly enhance learning.

Proposition 3.8. *Under the closest algorithm $\mathscr{C}$, user i's learning becomes negligible as $n \to \infty$:*

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

Proof. See Appendix A. □

This finding is in stark contrast to classic learning models where the wisdom of the crowd enhances learning as the population grows. The main difference is that, here, the platform's strategic role in feed selection undermines learning, making it vanish. Indeed, as soon as more similar users are available for the platform, they join user i's feed and, at some point, her feed is flooded with only positively correlated neighbors and hence diversity gradually disappears. Interestingly, when n grows very large, every user i has at least k (where k is taken from Lemma 3.7) other users almost equal to her, and all those users are very likely to share the same feed. Indeed, in the proof of Proposition 3.8 we show that for every pair of users j and l verifying $\rho_{ij} > 1-\varepsilon$ and $\rho_{il} > 1-\varepsilon$, then $\rho_{jl} > 1-4\varepsilon+2\varepsilon^2$. This means that if j and l are in i's feed, they will also be in each others' feed except from the case in which there are e_j or e_l other users even closer to them. These ideas fit completely with the notion of echo chambers and filter bubbles widely discussed in the literature (Pariser, 2011).

Corollary 3.9. *Given $\boldsymbol{\Sigma}$, the probability of finding some $\tilde{n} \in \mathbb{N}$ such that, for all $n > \tilde{n}$, every user in $\mathscr{C}_i^k$ is positively correlated with user i converges to one.*

Proof. Given $\boldsymbol{\Sigma}$, there are some m users positively correlated with user i. If $m < k$, then we have to look for the $k-m$ remaining users in the extension of $\boldsymbol{\Sigma}$. The probability of not getting $k-m$ positively correlated new users for platform size $\tilde{n}$ converges to zero. □

The platform-optimal algorithm not only overexploits similarity and creates echo chambers but also harms learning in large populations. These issues are significant in public debate, raising concerns about the impact of social media platforms on social welfare. The approval of the DSA by the European Commission addresses these concerns. In particular, the DSA forces platforms to include a non-strategic algorithm which is not based on profiling. As a consequence, social media platforms have chosen to reinstate the reverse-chronological algorithm. The next section is devoted to an analysis of alternative algorithms, including the already mentioned reverse-chronological algorithm, the user-optimal algorithm, and the breaking-echo-chambers algorithm.

4 The reverse-chronological algorithm and other alternatives

The reverse-chronological algorithm, which will be denoted by $\mathscr{R}$, displays friends' posts in the (reverse) order they were written. Before the implementation of personalization algorithms, every social media platform relied on this simple method of presenting the feed, which is not strategic at all. In this model, we understand the reverse-chronological algorithm as a *random* algorithm in which a post will be at the top of the feed with probability $\frac{1}{n-1}$. Consequently, this algorithm does not create echo chambers but provides users with diverse views. As one could expect, when the platform size grows large, the reverse-chronological algorithm outperforms the platform-optimal algorithm in terms of learning, although it performs worse in terms of conformity. The effect on overall utility depends on how users weight sincerity, conformity, and learning. However, for small populations, it is not even the case that individuals consistently learn better under the reverse-chronological algorithm, and within-the-platform utility is, of course, higher under the platform-optimal algorithm. Therefore, we are far from being able to state unambiguously that this algorithm is a reliable alternative for the platform-optimal algorithm, which motivates our search for a better substitute. Let us go through these ideas in detail.

Remember that, under the closest algorithm, $\mathscr{C}$, learning vanishes when platform size grows large (Proposition 3.8). However, the random nature of $\mathscr{R}$ yields better learning asymptotically:

$$\operatorname*{plim}_{n\to\infty} \mathrm{Var}[\theta|\boldsymbol{\theta}_{\mathscr{R}_i^{e_i}}] = \frac{\sigma^2}{e_i}.$$

This is, of course, not surprising. Hence, there is a trade-off between learning and conformity when a user compares her expected utility under both algorithms.

Proposition 4.1. *Given* $\boldsymbol{\Sigma}$, *the closest algorithm outperforms the reverse-chronological algorithm in large populations if and only if*

$$\lambda > \max_{i\in\mathcal{U}} \left\{ \frac{1-\mathbb{E}_i\left[\frac{1}{e_i}\right]}{1-\mathbb{E}_i\left[\frac{1}{e_i}\right] + (1-\beta)\mathbb{E}_i\left[\sum_{j\in\mathscr{R}_i^{e_i}} \frac{(1-\rho_{ij}^2)}{e_i} |\theta_i, \mathscr{F}\right]} \right\}.$$

Moreover, the closest algorithm is always worse than the reverse-chronological algorithm in terms of learning.

Proof. See Appendix A. □

Note that we can rewrite the condition as

$$\lambda > \max_{i\in\mathcal{U}} \left\{ \frac{1}{1 + \frac{(1-\beta)\mathbb{E}_i\left[\sum_{j\in\mathscr{R}_i^{e_i}} \frac{(1-\rho_{ij}^2)}{e_i} | \theta_i, \mathscr{F}\right]}{1-\mathbb{E}_i\left[\frac{1}{e_i}\right]}} \right\},$$

where the term $\mathbb{E}_i\left[\sum_{j\in\mathscr{R}_i^{e_i}} \frac{(1-\rho_{ij}^2)}{e_i} | \theta_i, \mathscr{F}\right]$ is non-negative, and, hence, there is always some $\lambda \in (0,1)$ satisfying the condition. Indeed, for most configurations of the game, the closest algorithm outperforms the reverse-chronological algorithm for almost every λ, as illustrated in Figure 1, where we have used the configuration from the example in Appendix B.

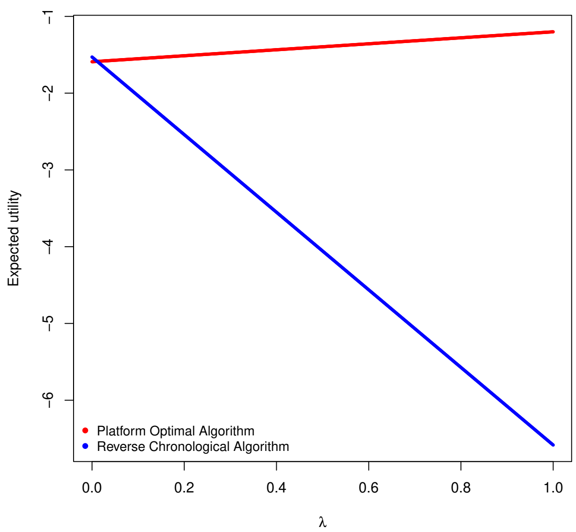

Figure 1: Expected utility under $\mathscr{C}$ and $\mathscr{R}$ for $\lambda \in (0,1)$.

To build intuition for the subsequent results, let us analyze futher how $\mathrm{Var}\left[\theta \mid \boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}\right]$ depends on the covariance structure $\boldsymbol{\Sigma}$. Recall the explicit expression for the variance derived in Equation (4), which relies on the *precision matrix* $\boldsymbol{\Sigma}^{-1}$. This matrix encodes the partial correlations between signals. Specifically, the precision matrix describes conditional dependencies: the partial correlation between θ_i and θ_j, controlling for all other signals, is given by

$$-\frac{x_{ij}}{\sqrt{x_{ii}x_{jj}}},$$

where x_{ij} is the (i,j)-th entry of $\boldsymbol{\Sigma}^{-1}$. This highlights that learning depends not only on direct correlations but also on conditional relationships across the feed. In other words: the effect of placing user j in user i's feed for user i's learning depends not just on how correlated i and j are, but also on how correlated is j with the rest of users in i's feed.

To further understand how learning responds to changes in the the relation between user i and some other user j in her feed, i.e., σ_{ij}, observe that the derivative of i's posterior variance with respect to σ_{ij} depends on how both i and j relate to each other user l in i's feed:

$$\frac{\partial}{\partial \sigma_{ij}} \mathrm{Var}\left[\theta \mid \boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}\right] \propto \sum_{l \in \mathscr{F}_i^{e_i}} x_{il} x_{jl}.$$

Thus, bringing to user i's feed some user j that is more correlated to her improves learning if the interactions between i, j and each other user are aligned, but could reduce learning if they are misaligned. The fact that taking a more correlated user to the feed might decrease learning is illustrated in the next example, in which platform size is small. For large platform sizes, however, finding highly correlated users with i that are misaligned with the rest of the feed is complicated essentially because feeds become echo chambers. This phenomenon causes learning to be non-monotonic under the closest algorithm, as we can observe in Figure 2, where we plot realizations of learning under $\mathscr{R}$ and $\mathscr{C}$ as n increases for a population growing from $n = 30$ to $n = 5000$, constant engagement $k = 30$ and parameters $\lambda = 0.5$ and $\beta = 0.2$.

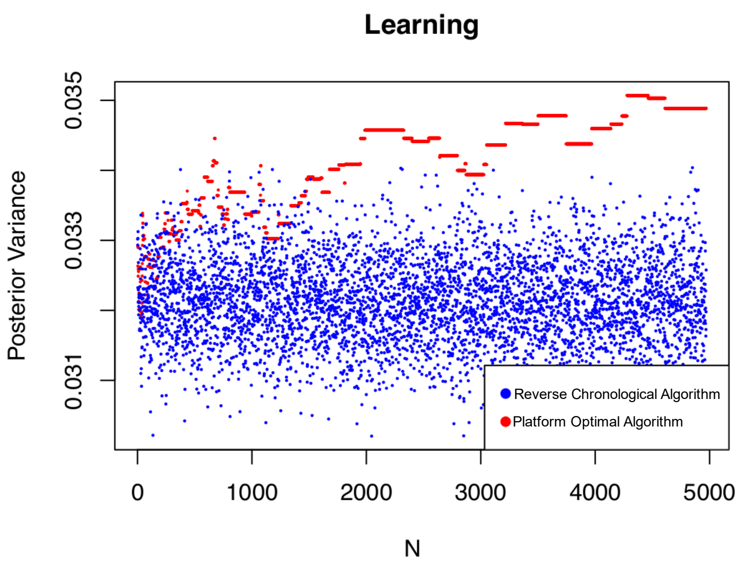

Figure 2: User's posterior variance as population grows; engagement is fixed to $k = 30$.

Consider now a tiny network composed of four individuals ($n = 4$), and assume, for this exercise, $e_i = 3$ and that the distribution of signals, conditional on θ, is as follows:

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

The closest algorithm induces, for user 1, a feed given by $\mathscr{C}_1^{e_i} = \{1, 2, 3\}$. Assume that a particular realization of the reverse-chronological algorithm induces the feed $\mathscr{R}_1^{e_i} = \{1, 3, 4\}$. Posterior variances are $\text{Var}[\theta|\{\theta_1, \theta_2, \theta_3\}] = 0.58$ for the closest algorithm and $\text{Var}[\theta|\{\theta_1, \theta_3, \theta_4\}] = 0.68$ for the reverse-chronological algorithm. Surprisingly, $\mathscr{C}$ yields better learning. The covariance to user 1 is not the unique driving force of learning; the conditional correlations between other users in the feed also play a role as we have indicated above.

From the discussion above, we conclude that the reverse-chronological algorithm might be a potential alternative to enhance learning, but it does not seem realistic to expect its return to social media platforms to be effective: most users are better off under the closest algorithm, even if the platform size is large. While one of the DSA's goals is to address the harms caused by personalization algorithms (a significant one being the echo chambers they create, as described in Section 3), reinstating the reverse-chronological feed does not seem to be a meaningful solution. This is understandable: personalization algorithms were designed with the objective of maximizing engagement, and they have certainly succeeded (Guess et al., 2023), as the platform-optimal algorithm is a sophisticated tool built to provide users with immediate gratification.

Another alternative worth exploring is modifying the platform-optimal algorithm. Recently, platforms like X and Facebook have adjusted their algorithms by adding new tools, such as crowd-sourced fact-checking or directly incorporating sponsored messages.[16] Next, we study how a modified platform-optimal algorithm, which we call the breaking-echo-chambers algorithm, $\mathscr{B}$, would function in our model. It simply involves adding a user with opposing views to the closest feed. Formally, for every $i \in \mathcal{U}$, $\mathscr{B}_i(k) = \mathscr{C}_i(k-1)$ for all $2 \leq k \leq n$, and $\mathscr{B}_i(1)$ is the user in the platform with the highest negative correlation to user i.

The result below shows that, asymptotically, $\mathscr{B}$ allows the user to correctly learn the state of the world, maximizing learning at no cost in conformity. Remember that user i's expected conformity is $\sigma^2 - \frac{\sigma_{ij}^2}{\sigma^2}$, so that she is indifferent between a covariance of $\sigma_{ij} = -\sigma^2$ or $\sigma_{ij} = \sigma^2$. Then, asymptotically, the breaking-echo-chambers algorithm incurs in no penalty in conformity, and both conformity and learning are simultaneously maximized. When platform size grows large, the breaking-echo-chambers algorithm converges to a utilitarian optimal algorithm. In contrast, the finite case is ambiguous: if there is a *very* opposite user available to be selected, neither conformity nor engagement will be that harmed and learning will be significantly improved. However, it might be the case that no such user is available, and then conformity and engagement are compromised: even though there might be an improvement in learning, the platform-optimal algorithm keeps providing higher utility.

16 See Martel et al. (2024) for an overview on crowd ratings as a measure to identify misinformation.

Proposition 4.2. *When platform size grows large, the breaking-echo-chambers algorithm outperforms the closest algorithm and converges to a utilitarian optimal algorithm.*

Proof. See Appendix A. □

Social media platforms have already implemented algorithm modifications that promote content aimed at improving user information. For example, in 2021, Twitter (now X) launched Birdwatch, a feature that gained widespread use in 2023, allowing contributors to provide context for posts. Since in our model the platform does not know the messages of the users, we cannot create a similar feature where the platform aggregates user messages to obtain (and potentially share) an estimator for θ. However, this is effectively what the breaking-echo-chambers algorithm accomplishes when platform size is large enough. Note that the outcome would be the same in both cases: each user reads (and learns about) the state of the world while simultaneously deriving significant instantaneous utility from interacting with similar friends.

In any case, implementing such an algorithm has some obvious drawbacks: it requires some complicated regulatory enforcement (the platform has no incentives to modify its optimal algorithm), and its long-term viability in the real world remains questionable. Even if opposite content is enforced, maybe through sponsored public service announcements with regular frequency or by directly incorporating dissimilar views into the feed, any user may simply choose to disregard such artificially added content and, perhaps naively, opt not to engage with it.

So far, we have explored two alternatives to the platform-optimal algorithm: the reverse-chronological algorithm, which is the current alternative to the platform-optimal algorithm for users in most social media platforms, and an artificial improvement to the platform-optimal algorithm: the breaking-echo-chambers algorithm. However, neither of these alternatives is fully satisfactory, as either their performance or their viability (or both) is questionable. There is, however, one alternative we have not yet examined: the utilitarian optimal algorithm. This algorithm maximizes user well-being, and, by Corollary 3.2, this is equivalent to maximizing the expected utility of each user. Nonetheless, we cannot provide a closed-form expression for this algorithm. We do, however, offer an example in Appendix B. We will denote the user-optimal algorithm as $\mathscr{U}$, and the next section is dedicated to exploring how, under the imposition of horizontal interoperability in the social media market, platforms would implement it.

5 The need for horizontal interoperability

Thus far, we have assumed a monopolist platform serving a pool of n users already on board. In this scenario, the platform is not concerned with user acquisition but focuses solely on maximizing engagement. This reflects the current landscape of social media platforms, where large incumbents like Facebook or X effectively operate as monopolists within specific niches. For instance, if someone seeks to join a microblogging community, they are likely to choose X. While alternatives like Mastodon or Bluesky exist, the decisive factor is that their friends are on X. Network effects, which are a defining feature of social media platforms where a platform's value increases with user participation, strongly protect these incumbents. Consequently, platforms have significant incentives to grow their user base, offering greater network benefits than their competitors. This creates high barriers to entry for new challengers, who must provide a vastly superior service to overcome the incumbents' advantages. For example, this dynamic was evident when Facebook displaced MySpace as the leading social networking site in 2009, though arguably, such an event has not occurred again since.

Let us formally define network effects regarding algorithms: we say that an algorithm $\mathscr{F}$ features network effects for user i if user i prefers platform size n to grow to $n+1$. Now, note that algorithms heavily rely on platform size, so network effects are particularly relevant.[17] The larger the network, the more possibilities for optimizing feeds and, eventually, the higher the expected utility for users. This is evident for the user-optimal algorithm: since platform and user incentives are aligned, a larger pool from which the platform can curate a feed translates to higher expected utility. However, the situation is more nuanced for the closest algorithm, as two opposing forces come into play when the platform size increases. On one hand, within-the-platform utility increases due to better matching possibilities. On the other hand, learning might decrease (we know that learning does not behave monotonically for small size increases, but that it asymptotically vanishes). Intuitively, the strategic role of the platform means the first force should dominate for users that care enough about within-the-platform utility: the feed is chosen to maximize it, with the effects on learning being a secondary consequence.

Proposition 5.1. *There exists some $\bar{\rho} < 0$ such that if $\rho_{ij} \geq \bar{\rho}$ for all $j \in \mathscr{C}_i^k$, the closest algorithm features network effects for user i if and only if $\lambda > \lambda_i^n$, where*

$$\lambda_i^n := -\frac{\Delta \operatorname{Var}_i^n}{\Delta \operatorname{Var}_i^n - \sigma^2 \mathbb{E}_i \left[\frac{1}{e_i}\right] \left(\rho_{ir}^2 - \mathbb{E}_i \left[\rho_{i(n+1)}^2 \mid \theta_i, \mathscr{C}_i(n)\right]\right)},$$

and $\Delta \operatorname{Var}_i^n = \mathbb{E}_i \left[\operatorname{Var}[\theta \mid \boldsymbol{\theta}_{\mathscr{C}(n+1)}] - \operatorname{Var}[\theta \mid \boldsymbol{\theta}_{\mathscr{C}(n)}] \mid \theta_i, \mathscr{C}(n)\right]$ denotes the expected change in posterior variance for user i when platform size grows to $n+1$, $\rho_{ir} = \min_{j \in \mathscr{C}_i^k} \{\rho_{ij}\}$ and (abusing notation) $n+1$ denotes the user who joins the platform to make it grow.

Proof. See Appendix A. □

This result implies that when users are matched with positively correlated individuals (or at least not too negatively correlated ones, as $\bar{\rho} < 0$), the closest algorithm exhibits

17 Many other services, such as privacy protection tools, accessibility or design do not depend on platform size.

network effects for users who do not disproportionally weight action utility with respect to within-the-platform utility. By Corollary 3.9, this holds true for sufficiently large platforms. Thus, large social media platforms that implement the closest algorithm inherently exhibit network effects. As a result, when a large incumbent adopts this algorithm, it attracts users and causes the market to tip in its favor. There are no incentives for either the platform or users to deviate, effectively locking users into the ecosystem. This underscores the importance of horizontal interoperability, which the European Commission defines as "the ability of information systems to exchange data and enable information sharing". Horizontal interoperability has been argued to be a necessary condition for fostering contestability in digital markets dominated by strong network effects (Kades and Scott Morton, 2020). With horizontal interoperability, network effects are no longer exclusive to incumbents but instead shared with new entrants, transforming them into a public good. This shifts competition dynamics from platforms competing *for the market* to competing *within the market* (Belleflamme and Peitz, 2020). Some industries, such as mobile phone and email services, already feature interoperability. For example, a Yahoo user can seamlessly send an email to a Gmail user, and *vice versa*.

The Digital Markets Act (DMA), enacted by the European Commission in November 2022, is the first regulatory attempt to introduce horizontal interoperability in digital markets. In its Article 7, it mandates horizontal interoperability between messaging services provided by gatekeepers. However, some experts remain skeptical about the overall effect of horizontal interoperability in digital markets and, in particular, in the messaging services market (Bourreau and Kraemer, 2023; Bourreau et al., 2023; Dhakar and Yan, 2024). Nevertheless, there is a critical distinction between social media platforms and the rest of the digital markets: personalization algorithms. These algorithms represent a non-interoperable feature where, in the absence of platform-specific network effects, platforms would need to compete to attract users. This is why we argue that horizontal interoperability acts as a *silver bullet* in the social media platforms market.

Let us elaborate on how horizontal interoperability would work on social media platforms. It would enable a user from platform A to view posts from friends on platform B, and *vice versa*. In practical terms, user i on platform A would receive a feed curated by platform A's algorithm, which could include posts from friends on platform B. Similarly, user j on platform B would see a feed where all her friends' posts appear, regardless of the platform they are registered on, arranged according to platform B's algorithm.[18] In this model, the only way for platforms to differentiate themselves is by offering users a better algorithm, i.e., one that maximizes their expected utility. Following a simple à

18 Although we consider a complete network in this paper, the fact that results apply to general networks might be interesting for this section. With horizontal interoperability, users can maintain their network of connections regardless of the platform each friend uses. This is analogous to the mobile phone industry, where the focus is on whether a friend owns a mobile phone, not which company provides the service.

la Bertrand argument, platforms would be compelled to optimize for user well-being (in other words, they would have to implement the user-optimal algorithm) because failing to do so would result in users migrating to competitors, as network effects would no longer lock them in. While implementing horizontal interoperability on social media platforms might face technical challenges and strong resistance from the largest platforms, we argue that ensuring user well-being in Europe requires extending the DMA to include horizontal interoperability for social media platforms.

6 Conclusion

We have developed a theoretical model of communication and learning through personalized feeds and demonstrated that engagement-maximizing platforms tend to overexploit user similarities when designing these feeds. As platform size increases, feeds become dominated by like-minded users, leading to the formation of echo chambers and severely impaired learning. As Pariser (2011) warned, algorithmic filtering can result in intellectual isolation and social fragmentation by exposing users primarily to like-minded individuals. Our model provides a theoretical foundation for this observation and suggests that such outcomes represent the trade-off inherent in platforms managing information exchanges to maximize engagement, as Guess et al. (2023) show empirically. Institutional efforts to improve this situation have relied on the reverse-chronological algorithm, but our analysis suggests that it may not be sufficient. Users enjoy receiving recommended content, and while a random selection might enhance learning, the associated disutility may outweigh the benefits. Although the breaking-echo-chambers algorithm, whose effects resemble those of crowd-source fact-checking, might be a promising alternative when the platform size is large, its practical implementation faces challenges. We then propose the implementation of horizontal interoperability. This approach leverages the non-interoperable nature of algorithms, ensuring platforms are forced to compete in the absence of network effects. Such competition would naturally incentivize platforms to adopt healthier algorithms that balance users' desire for conformity with significant improvements in learning.

Our modeling choices provide several avenues for extending this work. First, we assumed improper priors for tractability, while normal priors are more standard in the literature. Under normal priors, users gain a clearer understanding of their relative ideological positions. In this framework, truthful reporting is no longer an equilibrium; instead, users post a convex combination of their private signal and the prior. Crucially, the algorithm would exert a significant influence on each user's message. We hypothesize that this could intensify echo chambers through a dual mechanism, without altering our conclusions conceptually. First, users closer to like-minded individuals would find their messages increasingly aligned with both their private signals and group consensus. Second, the existing tendency toward conformity would amplify this effect. However, deriving explicit expressions for algorithms under this framework remains intractable.

Second, our model assumes that user engagement stems from a desire for both sincerity and conformity. While plausible, this assumption overlooks an important behavioral aspect: platforms also exploit user interaction driven by exposure to opposing content. Evidence suggests that interaction likelihood follows a U-shaped curve, peaking when users encounter either very similar or highly dissimilar content. A promising extension of our work would involve modeling this U-shaped interaction curve to understand how learning unfolds in the presence of opposing viewpoints. In such a framework, user engagement might follow the U-shaped dynamic, while user utility would continue to depend on conformity and learning. We speculate that this extension would yield similar distinctions between platform-optimal and utilitarian algorithms, ultimately leading to comparable conclusions with the ones we provide.

Finally, a promising research avenue involves developing a comprehensive theoretical model to study the implementation of horizontal interoperability in digital markets, particularly for social media platforms. Such a model could provide insights into how interoperability affects innovation and whether challenges such as security and privacy concerns, the loss of multi-homing, or the discoverability problem are significant enough to hinder its effectiveness.

Acknowledgments. 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, Fernando Payro, Jordi Caballé, Inés Macho-Stadler, Iván Rendo and Diego Fica 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, the ENTER Jamboree, the BSE Jamboree, 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.

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, G, "Measuring substitution patterns in the attention economy: An experimental approach," *RAND Journal of Economics*, 2024.

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.

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.

Benhabib, Jess and Alberto Bisin, "Modeling internal commitment mechanisms and self-control: A neuroeconomics approach to consumption–saving decisions," *Games and Economic Behavior*, 2005, *52* (2), 460–492.

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

Biddle, Sam, "Facebook uses artificial intelligence to predict your future actions for advertisers, says confidential document," *The Intercept*, 2018, *13* (04), 2018.

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.

Galeotti, Andrea, Benjamin Golub, Sanjeev Goyal, and Rithvik Rao, "Discord and harmony in networks," *available at https://arxiv.org/pdf/2102.13309*, 2021.

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

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.

Grayot, James D., "Dual process theories in behavioral economics and neuroeconomics: A critical review," *Review of Philosophy and Psychology*, 2020, *11* (1), 105–136.

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.

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.

Kosinski, Michal, David Stillwell, and Thore Graepel, "Private traits and attributes are predictable from digital records of human behavior," *Proceedings of the National Academy of Sciences*, 2013, *110* (15), 5802–5805.

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.

Lauer, David, "Facebook's ethical failures are not accidental; they are part of the business model," *AI and Ethics*, 2021, *1* (4), 395–403.

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.

Martel, Cameron, Jennifer Allen, Gordon Pennycook, and David G Rand, "Crowds can effectively identify misinformation at scale," *Perspectives on Psychological Science*, 2024, *19* (2), 477–488.

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

Proof of Proposition 3.1

Proof. User i chooses a message $m_i \in \mathbb{R}$ to maximize her within-the-platform expected utility, knowing her private signal θ_i and the algorithm $\mathscr{F}$. I.e., user i picks m_i to maximize

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

Note that e_i is for the user an exogenous random variable that does not depend on m_i. Thus, such optimal message m_i is chosen to maximize

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

The first order condition with respect to m_i yields

$$m_i = \beta\theta_i + (1-\beta)\mathbb{E}_i \left[\sum_{j \in \mathscr{F}_i^{e_i}} \frac{m_j(\theta_j)}{e_i} |\theta_i, \mathscr{F} \right]. \tag{5}$$

As this holds for all $j \in \mathcal{U}$, we substitute in this expression $m_j(\theta_j) = \beta\theta_j + (1-\beta)\mathbb{E}_j \left[\sum_{l \in \mathscr{F}_j^{e_j}} \frac{m_l(\theta_l)}{e_j} |\theta_j, \mathscr{F} \right]$ for all $j \in \mathscr{F}_i^{e_i}$, and then we repeat the procedure for every $l \in \mathscr{F}_j^{e_j}$ and for all $j \in \mathscr{F}_i^{e_i}$ and so on and so forth up to m times.

$$\begin{aligned} m_i &= \beta\theta_i + (1-\beta)\mathbb{E}_i \left[\sum_{j \in \mathscr{F}_i^{e_i}} \frac{m_j(\theta_j)}{e_i} |\theta_i, \mathscr{F} \right]. \\ &= \beta\theta_i + (1-\beta)\mathbb{E}_i \left[\sum_{j \in \mathscr{F}_i^{e_i}} \frac{1}{e_i} \left(\beta\theta_j + (1-\beta)\mathbb{E}_j \left[\sum_{l \in \mathscr{F}_j^{e_j}} \frac{m_l(\theta_l)}{e_j} |\theta_j, \mathscr{F} \right] \right) \right] \end{aligned} \tag{6}$$

Note, at this stage, that because of improper priors and by using the Law of Iterated Expectations, we obtain $\mathbb{E}_i \left[\frac{1}{e_i} \sum_{j \in \mathscr{F}_i^{e_i}} \theta_j |\theta_i, \mathscr{F} \right] = \theta_i$. In detail:

$$\begin{aligned} \mathbb{E}_i \left[\sum_{j \in \mathscr{F}_i^{e_i}} \frac{\theta_j}{e_i} |\theta_i, \mathscr{F} \right] &= \mathbb{E} \left[\mathbb{E}_i \left[\sum_{j \in \mathscr{F}_i^{e_i}} \frac{\theta_j}{e_i} |\theta_i, \mathscr{F}, e_i \right] |\theta_i, \mathscr{F} \right] = \mathbb{E} \left[\sum_{j \in \mathscr{F}_i^{e_i}} \frac{\mathbb{E}_i \left[\theta_j|\theta_i, \mathscr{F}, e_i\right]}{e_i} |\theta_i, \mathscr{F} \right] \\ &= \mathbb{E} \left[\frac{e_i \mathbb{E}_i \left[\theta_j|\theta_i, \mathscr{F}, e_i\right]}{e_i} |\theta_i, \mathscr{F} \right] = \mathbb{E}_i[\theta_j|\theta_i, \mathscr{F}]. \end{aligned}$$

Applying this property to the first term in the brackets of Equation (6) and continuing with the expansion, we obtain:

$$m_i = \beta\theta_i + (1-\beta)\beta\theta_i + (1-\beta)^2\mathbb{E}_i\left[\sum_{j\in\mathscr{F}_i^{e_i}}\mathbb{E}_j\left[\sum_{l\in\mathscr{F}_j^{e_j}}\frac{m_l(\theta_l)}{e_ie_j}|\theta_j,\mathscr{F}\right]|\theta_i,\mathscr{F}\right] = \dots$$

$$= \beta\theta_i\sum_{r=0}^{m-1}(1-\beta)^r$$

$$+ (1-\beta)^m\mathbb{E}_i\left[\sum_{j\in\mathscr{F}_i^{e_i}}\mathbb{E}_j\left[\sum_{l\in\mathscr{F}_j^{e_j}}\dots\mathbb{E}_m\left[\sum_{p\in\mathscr{F}_m^{e_m}}\frac{m_p(\theta_p)}{e_ie_j\dots e_m}|\theta_m,\mathscr{F}\right]\dots|\theta_j,\mathscr{F}\right]|\theta_i,\mathscr{F}\right]. \quad (7)$$

This expression holds for all $m\in\mathbb{N}$, so we can take limits when $m\to\infty$. On the one hand, $\lim_{m\to\infty}\sum_{r=0}^{m}(1-\beta)^r = \frac{1}{\beta}$, and, hence, the first term in Equation (7) is simply θ_i. On the other hand, the second term vanishes as $m\to\infty$. Hence, $m_i^* = \theta_i$ for all $i\in\mathcal{U}$ and we have truthtelling for any algorithm $\mathscr{F}$. □

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

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

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

$$\mathscr{F}_i(k) = \underset{j\in\mathcal{U}\setminus\mathscr{F}_i^{k-1}}{\text{argmax}}\left\{-\frac{1}{k}\mathbb{E}_p\left[\sum_{l\in\mathscr{F}_i^{k-1}}(\theta_i-\theta_l)^2 + (\theta_i-\theta_j)^2\right]\right\}$$
$$= \underset{j\in\mathcal{U}\setminus\mathscr{F}_i^{k-1}}{\text{argmax}}\left\{-\mathbb{E}_p\left[(\theta_i-\theta_j)^2\right]\right\}. \quad (8)$$

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 (8) 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.1. *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. 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 $0 < \gamma < \beta < 1$ such that for all

$x \in \mathbb{R}$, $0 < \gamma < g(x) < \beta < 1$. Thus,

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

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:

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

Finally, let us define

$$k := \min\left\{\tilde{k} \in \mathbb{N} \text{ such that } \tilde{k} \geq \frac{1}{(1-\beta)^2}-\frac{\gamma}{(1-\gamma)^2}\right\}$$

and note that $k > 1$ because $\frac{1}{(1-\beta)^2}-\frac{\gamma}{(1-\gamma)^2} > 1$ if and only if $(2\beta-\gamma-\beta^2)+(2\beta\gamma^2-2\beta\gamma)+(\gamma\beta^2-\beta^2\gamma^2) > 0$, which holds precisely because $0 < \gamma < \beta < 1$. □

Proof of Proposition 3.8

Proof. Remember that ρ_{ij} refers to the correlation between user i and user j. Now, given the generating process for new users, for every $\varepsilon > 0$, there is some $\bar{n} \in \mathbb{N}$ such that if $n > \bar{n}$, there are k user i's neighbors $j_1, ..., j_k$ such that $\mathbb{P}[\rho_{j_r,i} > 1-\varepsilon] > 1-\gamma$ for all $r \in \{1, ..., k\}$ and $\gamma > 0$. On the other hand, applying the Cauchy-Schwarz inequality to the correlations between the pairs formed by user i and two other users, say j_r and j_l, we get

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

Using the ε-bounds derived above, we obtain:

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

Now, for each n, engagement might change and the feed will be $\mathscr{C}_i^{e_i(n)}$. As $e_i(n) \leq k$, users in i's feed are chosen from the set of k users defined above and denoted by $\{j_1, ..., j_k\}$. Let us now define $\delta = 4\varepsilon - 2\varepsilon^2$. For every $\delta > 0$ and $\gamma > 0$, there is some $\tilde{n} \in \mathbb{N}$ such that if $n > \tilde{n}$, the set of k users defined above verifies that if $j_r, j_l \in \{j_1, ..., j_k\}$, $\mathbb{P}[\rho_{j_r,j_l} > 1-\delta] > 1-\gamma$ (it is enough to choose ε accordingly). For each specific $e_i(n)$, such number of users will be selected. Hence, we have that for the $e_i(n) \times e_i(n)$ matrix

$\mathbf{A}(n)$ defined as

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

the probability of $\mathbf{A}(n)$ being element-wise lower or equal than the covariance matrix for the users in $\mathscr{C}_i^{e_i(n)}$, $\mathbf{\Sigma}_{\mathscr{C}_i^{e_i(n)}}$, is precisely the probability that the off-diagonal elements of $\mathbf{A}(n)$ are lower or equal than the corresponding off-diagonal elements of $\mathbf{\Sigma}_{\mathscr{C}_i^{e_i(n)}}$. Given that $\mathbf{A}(n)$ is symmetric, we have that $\mathbb{P}[\mathbf{A}(n) \leq \mathbf{\Sigma}_{\mathscr{C}_i^{e_i(n)}}] \geq (1-\gamma)^{(e_i(n)^2 - e_i(n))/2}$. Now, we need an auxiliary result:

Lemma A.2. *In this particular case, and for each $n \in \mathbb{N}$, $\boldsymbol{A}(n) \leq \mathbf{\Sigma}_{\mathscr{C}_i^{e_i(n)}}$ implies $\mathbb{1}\mathbf{\Sigma}^{-1}_{\mathscr{C}_i^{e_i(n)}}\mathbb{1}^t \leq \mathbb{1}\boldsymbol{A}^{-1}(n)\mathbb{1}^t$.*

Proof. Let us denote by b_{ij} the elements of the covariance matrix restricted to the users selected by the closest algorithm for user i when platform size equals n, $\mathbf{\Sigma}_{\mathscr{C}_i^{e_i(n)}}$:

$$\mathbf{\Sigma}_{\mathscr{C}_i^{e_i(n)}} = \begin{pmatrix} 1 & b_{12} & b_{13} & \dots & b_{1e_i} \\ b_{12} & 1 & b_{23} & \dots & b_{2e_i} \\ b_{13} & b_{23} & 1 & \dots & b_{3e_i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{1e_i} & b_{2e_i} & b_{3e_i} & \dots & 1 \end{pmatrix}.$$

Then, renaming the elements of $\mathbf{A}(n)$ as $a = 1-\delta$, we rewrite $\mathbf{A}(n)$ as

$$\mathbf{A}(n) = \sigma^2 \begin{pmatrix} 1 & a & a & \dots & a \\ a & 1 & a & \dots & a \\ a & a & 1 & \dots & a \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a & a & a & \dots & 1 \end{pmatrix},$$

with $a = 1-\delta$ such that $\mathbf{A}(n) \leq \mathbf{\Sigma}_{\mathscr{C}_i^{e_i(n)}}$ element-wise, as above. Now, we denote the elements of the inverse matrices $\mathbf{A}^{-1}(n)$ and $\mathbf{\Sigma}^{-1}_{\mathscr{C}_i^{e_i(n)}}$ as follows:

$$\mathbf{\Sigma}^{-1}_{\mathscr{C}_i^{e_i}(n)} = \begin{pmatrix} \bar{b}_{11} & \bar{b}_{12} & \bar{b}_{13} & \dots & \bar{b}_{1e_i} \\ \bar{b}_{12} & \bar{b}_{22} & \bar{b}_{23} & \dots & \bar{b}_{2e_i} \\ \bar{b}_{13} & \bar{b}_{23} & \bar{b}_{33} & \dots & \bar{b}_{3e_i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{b}_{1e_i} & \bar{b}_{2e_i} & \bar{b}_{3e_i} & \dots & \bar{b}_{e_ie_i} \end{pmatrix},$$

and

$$\mathbf{A}^{-1}(n) = \alpha \begin{pmatrix} 1 & \bar{a} & \bar{a} & \dots & \bar{a} \\ \bar{a} & 1 & \bar{a} & \dots & \bar{a} \\ \bar{a} & \bar{a} & 1 & \dots & \bar{a} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{a} & \bar{a} & \bar{a} & \dots & 1 \end{pmatrix}.$$

Now, as $\mathbf{\Sigma}_{\mathscr{C}_i^{e_i(n)}}\mathbf{\Sigma}^{-1}_{\mathscr{C}_i^{e_i(n)}} = \mathbf{Id}$, $\bar{b}_{11} + b_{12}\bar{b}_{12} + b_{13}\bar{b}_{13} + ... + b_{1e_i}\bar{b}_{1e_i} = 1$. Moreover, $\mathbf{A} \leq \mathbf{\Sigma}_{\mathscr{C}_i^{e_i(n)}}$ implies that $\bar{b}_{11} + a\sum_{j=2}^{e_i}\bar{b}_{1j} \leq 1$. On the other hand, as $\mathbf{A}(n)\mathbf{A}^{-1}(n) = \mathbf{Id}$, $\alpha(1 + a\bar{a}(e_i(n) - 1)) = 1$. Hence,

$$\bar{b}_{11} + a\sum_{j=2}^{e_i(n)}\bar{b}_{1j} \leq \alpha(1 + a\bar{a}(e_i(n) - 1)), \quad \forall\, a \in (0, 1).$$

This inequality implies that the sum of the first row elements in $\mathbf{\Sigma}_{\mathscr{C}_i^{e_i(n)}}$ is smaller than that of the first row elements of $\mathbf{A}(n)$. Repeating this simple reasoning let us conclude the lemma with the desired result: $\mathbb{1}\mathbf{\Sigma}^{-1}_{\mathscr{C}_i^{e_i(n)}}\mathbb{1}^t \leq \mathbb{1}\mathbf{A}^{-1}(n)\mathbb{1}^t$. □

Therefore, with probability $(1-\gamma)^{(e_i^2(n)-e_i(n))/2}$, it holds that

$$\mathbb{1}\mathbf{\Sigma}^{-1}_{\mathscr{C}_i^{e_i(n)}}\mathbb{1}^t \leq \mathbb{1}\mathbf{A}^{-1}(n)\mathbb{1}^t \Rightarrow \frac{1}{\mathbb{1}\mathbf{A}^{-1}(n)\mathbb{1}^t} \leq \frac{1}{\mathbb{1}\mathbf{\Sigma}^{-1}_{\mathscr{C}_i^{e_i(n)}}\mathbb{1}^t} \Rightarrow \frac{1}{\mathbb{1}\mathbf{A}^{-1}(n)\mathbb{1}^t} \leq \mathrm{Var}\left[\theta|\boldsymbol{\theta}_{\mathscr{C}_i^{e_i(n)}}\right].$$

On the other hand, we have that $\mathrm{Var}[\theta|\boldsymbol{\theta}_{\mathscr{C}_i^{e_i(n)}}] \leq \sigma^2$ by construction (note that $\mathrm{Var}[\theta|\theta_i] = \sigma^2$ and additional information weakly reduces the posterior variance). Consequently, after calculating $\mathbb{1}\mathbf{A}^{-1}(n)\mathbb{1}^t = \frac{e_i(n)}{\sigma^2(1+(e_i(n)-1)(1-\delta))}$, we finally get that, for every $\delta \in (0, 1)$, there is some underlying $\tilde{n}$ such that, if $n > \tilde{n}$, the following inequality

$$\frac{\sigma^2(1 + (e_i(n) - 1)(1 - \delta))}{e_i(n)} \leq \mathrm{Var}\left[\theta|\boldsymbol{\theta}_{\mathscr{C}_i^{e_i(n)}}\right] \leq \sigma^2$$

holds with probability $(1-\gamma)^{(e_i^2(n)-e_i(n))/2}$. Then, taking limits in the above expression we obtain that $\mathrm{plim}_{n\to\infty}\left(\mathrm{Var}\left[\theta|\boldsymbol{\theta}_{\mathscr{C}_i^{e_i(n)}}\right]\right) = \sigma^2$. □

Proof of Proposition 4.1

Proof. The result on learning follows from the direct comparison between Equation (4) and Proposition 3.8. Next, let us find the necessary and sufficient condition for all users in $\mathcal{U}$ preferring the closest algorithm. Just recall that when platform size grows large, the closest algorithm yields expected utility (in equilibrium) $\lambda\alpha\mathbb{E}_i[e_i] - (1-\lambda)\sigma^2$ because the loss in conformity converges to zero:

$$\lim_{n\to\infty} \mathbb{E}_i \left[\frac{1}{e_i} \sum_{j\in\mathscr{C}_i^{e_i}} (\theta_i - \theta_j)^2 | \theta_i, \mathscr{C} \right] = \lim_{n\to\infty} \mathbb{E}_i \left[\mathbb{E}_i \left[\frac{1}{e_i} \sum_{j\in\mathscr{C}_i^{e_i}} (\theta_i - \theta_j)^2 | e_i, \theta_i, \mathscr{C} \right] | \theta_i, \mathscr{F} \right]$$

$$= \lim_{n\to\infty} \mathbb{E}_i \left[\frac{1}{e_i} \sum_{j\in\mathscr{C}_i^{e_i}} \mathbb{E}_i \left[(\theta_i - \theta_j)^2 | e_i, \theta_i, \mathscr{C} \right] | \theta_i, \mathscr{F} \right] = \lim_{n\to\infty} \mathbb{E}_i \left[\frac{\sigma^2}{e_i} \sum_{j\in\mathscr{C}_i^{e_i}} (1 - \rho_{ij}^2) | \theta_i, \mathscr{F} \right] = 0,$$

where we have used the Law of Iterated Expectations. Hence,

$$\lim_{n\to\infty} \mathbb{E}_i \left[U_i(e_i, m_i, m_{-i}, a_i, \mathscr{C}, \theta_i, \theta) \right] = \lambda\, \alpha\, \mathbb{E}_i[e_i] - (1-\lambda)\sigma^2,$$

and the reverse-chronological algorithm yields

$$\lim_{n\to\infty} \mathbb{E}_i \left[U_i(e_i, m_i, m_{-i}, a_i, \mathscr{R}, \theta_i, \theta) \right] = \lambda \left(\alpha \mathbb{E}_i[e_i] - (1-\beta) \mathbb{E}_i \left[\frac{\sigma^2}{e_i} \sum_{j\in\mathscr{R}_i^{e_i}} (1 - \rho_{ij}^2) | \theta_i, \mathscr{F} \right] \right) - (1-\lambda) \mathbb{E}_i \left[\frac{\sigma^2}{e_i} \right].$$

Now, the expected utility derived by user i from the closest algorithm is greater or equal than that derived from the reverse-chronological algorithm, in equilibrium and for a platform size asymptotically large, if and only if

$$-(1-\lambda)\sigma^2 \geq -\lambda(1-\beta)\mathbb{E}_i \left[\frac{\sigma^2}{e_i} \sum_{j\in\mathscr{R}_i^{e_i}} (1 - \rho_{ij}^2) | \theta_i, \mathscr{F} \right] - (1-\lambda)\mathbb{E}_i \left[\frac{\sigma^2}{e_i} \right].$$

This condition is equivalent to

$$\lambda > \frac{1 - \mathbb{E}_i \left[\frac{1}{e_i} \right]}{1 - \mathbb{E}_i \left[\frac{1}{e_i} \right] + (1-\beta) \mathbb{E}_i \left[\sum_{j\in\mathscr{R}_i^{e_i}} \frac{(1-\rho_{ij}^2)}{e_i} | \theta_i, \mathscr{F} \right]},$$

so that any λ that meets this condition for every user $i \in \mathcal{U}$ guarantees that $\mathscr{C}$ is preferred over $\mathscr{R}$. □

Proof of Proposition 4.2

Proof. The following proof consists of two parts. First, we will show that for large platform sizes, conformity yields no loss under $\mathscr{B}$, exactly as it does under $\mathscr{C}$. Second, we will show that, asymptotically, learning is perfect under $\mathscr{B}$.

Let us define $k := \min \left\{ \tilde{k} \in \mathbb{N} \text{ such that } \tilde{k} \geq \frac{1}{(1-\beta)^2} - \frac{\gamma}{(1-\gamma)^2} \right\}$ as in the proof of Proposition 3.8 and follow a similar reasoning. For every $\varepsilon, \nu > 0$ there exists a $\bar{n} \in \mathbb{N}$ such that for every $n > \bar{n}$, there is a set of k users such that $\mathbb{P}[\rho_{ij} \geq 1 - \varepsilon] \geq 1 - \gamma$ for every $j = 1, ..., k$. Moreover, for every $\delta, \varphi > 0$, there exists a $\tilde{n} \in \mathbb{N}$ such that for every $n > \tilde{n}$

there is a user l such that $\mathbb{P}[\rho_{il} < \delta - 1] \geq 1 - \varphi$. Let us now take $n \geq \max\{\bar{n}, \tilde{n}\}$ and define, for any engagement e_i, $\mathscr{B}_i^{e_i} = \{l, 1, ..., e_i - 1\}$, where users in $\{1, ..., e_i - 1\}$ are taken from the pool of size k defined above. Then, we have that

$$\operatorname*{plim}_{n\to\infty} \rho_{ij} = 1 \quad \forall j \in \{1, ..., k\}, \text{ and } \operatorname*{plim}_{n\to\infty} \rho_{il} = -1,$$

and expected conformity under the breaking-echo-chambers algorithm becomes, by using the Law of Iterated Expectations,

$$\begin{aligned}
\mathbb{E}_i\left[\frac{1}{e_i}\left(\sum_{j\in\mathscr{B}_i^{e_i}}(\theta_i-\theta_j)^2)\right)|\theta_i,\mathscr{B}\right] &= \mathbb{E}_i\left[\frac{1}{e_i}\left(\sum_{j\in\mathscr{C}_i^{e_i-1}}(\theta_i-\theta_j)^2)+(\theta_i-\theta_l)^2\right)|\theta_i,\mathscr{B}\right]\\
&= \mathbb{E}_i\left[\frac{1}{e_i}\sum_{j\in\mathscr{C}_i^{e_i-1}}(\theta_i-\theta_j)^2|\theta_i,\mathscr{C}\right]+\mathbb{E}_i\left[\frac{1}{e_i}(\theta_i-\theta_l)^2\right]\\
&= \mathbb{E}_i\left[\frac{1}{e_i}\sum_{j\in\mathscr{C}_i^{e_i-1}}(1-\rho_{ij}^2)|\theta_i,\mathscr{C}\right]+\mathbb{E}_i\left[\frac{1}{e_i}(1-\rho_{il}^2)\right]\\
&= \mathbb{E}_i\left[\frac{1}{e_i}\sum_{j\in\mathscr{C}_i^{e_i-1}}(1-\rho_{ij}^2)|\theta_i,\mathscr{C}\right],
\end{aligned}$$

which converges to zero as $n \to \infty$.

Finally let us show that $\mathscr{B}$ yields perfect learning asymptotically. When user i learns user l's private signal, her precision matrix becomes

$$\mathbf{\Sigma}_{il}^{-1} = \frac{1}{\sigma^4 - \sigma_{il}^2}\begin{pmatrix}\sigma^2 & -\sigma_{ij}\\ -\sigma_{ij} & \sigma^2\end{pmatrix},$$

and, consequently, the posterior variance of θ is

$$\operatorname{Var}[\theta|\theta_i, \theta_l] = \frac{1}{\mathbb{1}\mathbf{\Sigma}_{il}^{-1}\mathbb{1}^t} = \frac{\sigma^4 - \sigma_{il}^2}{2(\sigma^2 - \sigma_{il})},$$

which coverges to zero as σ_{il} converges to σ^2 $n \to \infty$. If user i learns more signals, her posterior variance will weakly decrease. Hence, when she learns the signals in her feed $\mathscr{B}_i^{e_i}$,

$$\operatorname{Var}[\theta|\theta_i, \theta_l] \geq \operatorname{Var}[\theta|\boldsymbol{\theta}_{\mathscr{B}_i^{e_i}}] \geq 0,$$

which implies that $\lim_{n\to\infty} \operatorname{Var}[\theta|\boldsymbol{\theta}_{\mathscr{B}_i^{e_i}}] = 0$ as we wanted to show. □

Proof of Proposition 5.1

Proof. The expected utility of user i under the closest algorithm when platform size grows to $n + 1$, conditional on her information in the n-size platform, is

$$
\begin{aligned}
\mathbb{E}_i\left[U_i(e_i,\theta_i,\theta_{-i},\mathscr{C}(n+1)) \mid \theta_i,\mathscr{C}(n)\right] = \lambda\alpha\mathbb{E}_i[e_i \mid \theta_i,\mathscr{C}(n))] \\
- \lambda(1-\beta)\mathbb{E}_i\left[\sum_{j\in\mathscr{C}_i^{e_i}(n+1)} \frac{(\theta_i-\theta_j)^2}{e_i} \mid \theta_i,\mathscr{C}(n)\right] \\
- (1-\lambda)\mathbb{E}_i\left[\mathrm{Var}[\theta \mid \boldsymbol{\theta}_{\mathscr{C}(n+1)}] \mid \theta_i,\mathscr{C}(n)\right],
\end{aligned}
$$

where we ahve denoted user i's feed for platform size $n+1$ by $\mathscr{C}_i^{e_i}(n+1)$. Hence, we can calculate the difference in expected utility when platform size grows to $n+1$ as $\mathbb{E}_i\left[U_i(e_i,\theta_i,\theta_{-i},\mathscr{C}(n+1)) \mid \theta_i,\mathscr{C}(n)\right] - \mathbb{E}_i\left[U_i(e_i,\theta_i,\theta_{-i},\mathscr{C}(n)) \mid \theta_i,\mathscr{C}(n)\right]$, which simplifies to

$$
\begin{aligned}
&- \lambda(1-\beta)\mathbb{E}_i\left[\sum_{j\in\mathscr{C}_i^{e_i}(n+1)} \frac{(\theta_i-\theta_j)^2}{e_i} - \sum_{j\in\mathscr{C}_i^{e_i}(n)} \frac{(\theta_i-\theta_j)^2}{e_i} \mid \mathscr{C}(n)\right] \\
&- (1-\lambda)\mathbb{E}_i\left[\mathrm{Var}[\theta \mid \boldsymbol{\theta}_{\mathscr{C}(n+1)}] - \mathrm{Var}[\theta \mid \boldsymbol{\theta}_{\mathscr{C}(n)}] \mid \theta_i,\mathscr{C}(n)\right]. \qquad (9)
\end{aligned}
$$

Let us call $\Delta\,\mathrm{Var}_i^n := \mathbb{E}_i\left[\mathrm{Var}[\theta \mid \boldsymbol{\theta}_{\mathscr{C}(n+1)}] - \mathrm{Var}[\theta \mid \boldsymbol{\theta}_{\mathscr{C}(n)}] \mid \theta_i,\mathscr{C}(n)\right]$. Then, when $\Delta\,\mathrm{Var}_i^n \leq 0$ (i.e., learning improves with platform size), utility increases for any $\lambda > 0$ if the expected change in conformity is positive. However, when $\Delta\,\mathrm{Var}_i^n > 0$, expected utility increases for user i only if $\lambda \geq \lambda_i^n$, where

$$
\lambda_i^n := -\frac{\Delta\,\mathrm{Var}_i^n}{\Delta\,\mathrm{Var}_i^n - \mathbb{E}_i\left[\sum_{j\in\mathscr{C}_i^{e_i}(n+1)} \frac{(\theta_i-\theta_j)^2}{e_i} - \sum_{j\in\mathscr{C}_i^{e_i}(n)} \frac{(\theta_i-\theta_j)^2}{e_i} \mid \mathscr{C}(n)\right]}.
$$

To complete the proof, we show that the expected conformity increases with platform size and, then, $\lambda_i^n \in [0,1]$. First, let us call $\rho_{ik} = \min_{j\in\mathscr{C}_i(n)^k}\{\rho_{ij}\}$. Then, note that under the closest algorithm, any new user denoted $n+1$ with a correlation $\rho_{i(n+1)} < \rho_{ik}$ will not appear in the feed. In this case, both the change in expected conformity and the change in expected variance are zero. This implies that, for any λ, the user is indifferent to the size of the platform.

However, if $\rho_{i(n+1)} \geq \rho_{ik}$, such user $n+1$ would enter user i's feed and would be placed between some users t and $t+1$ satisfying $\rho_{it} \geq \rho_{i(n+1)} \geq \rho_{i(t+1)} \geq 0$. In this scenario, the difference in conformity in Equation (9) simplifies to

$$
\sum_{j\in\mathscr{C}_i^{e_i}(n+1)} \frac{(\theta_i-\theta_j)^2}{e_i} - \sum_{j\in\mathscr{C}_i^{e_i}(n)} \frac{(\theta_i-\theta_j)^2}{e_i} = \frac{(\theta_i-\theta_{n+1})^2}{e_i} - \frac{(\theta_i-\theta_k)^2}{e_i},
$$

as all other users are common in the feed. Recall that, in expectation,

$$
\begin{aligned}
\mathbb{E}_i\left[\frac{(\theta_i-\theta_{n+1})^2}{e_i} - \frac{(\theta_i-\theta_k)^2}{e_i} \mid \theta_i,\mathscr{C}_i(n)\right] &= \mathbb{E}_i\left[\frac{\sigma^2}{e_i}\left(1-\rho_{i(n+1)}^2 - 1 + \rho_{ik}^2\right)\right] \\
&= \sigma^2\mathbb{E}_i\left[\frac{1}{e_i}\right]\left(\rho_{ik}^2 - \mathbb{E}_i\left[\rho_{i(n+1)}^2 \mid \theta_i,\mathscr{C}_i(n)\right]\right).
\end{aligned}
$$

If ρ_{ik} is sufficiently large, the expected conformity increases relative to the n-user feed due to the monotonic properties of the closest algorithm. From user i's perspective, the expected conformity increases if and only if $\text{Var}\left[\rho_{i(n+1)}|\theta_i, \mathscr{C}_i(n)\right] \geq \rho_{ik}^2$. Hence, there is some $\bar{\rho} < 0$, defined as the negative square root of $\text{Var}\left[\rho_{i(n+1)}|\theta_i, \mathscr{C}_i(n)\right]$ such that user i is indifferent, in expectation, between platform sizes n and $n+1$. Thus, whenever $\rho_{ik} \geq \bar{\rho}$, the expected conformity increases with n, and if $\lambda > \lambda_i^n$, user i experiences network effects in the platform.

□

B Example

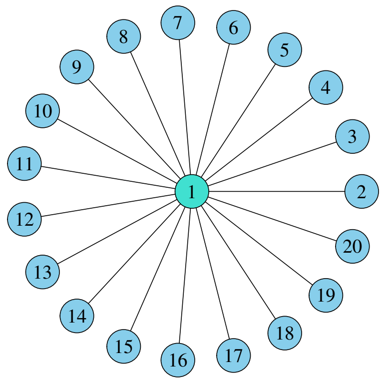

Figure 3: Platform size $n = 20$.

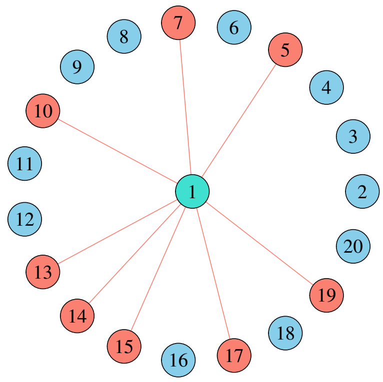

Figure 4: Platform-optimal feed.

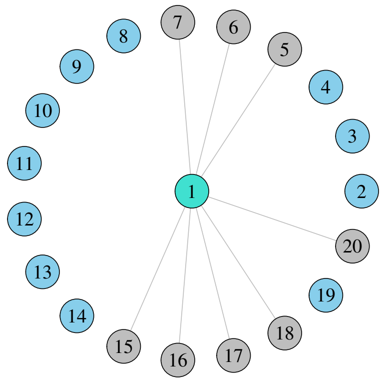

Figure 5: Reverse-chronological feed.

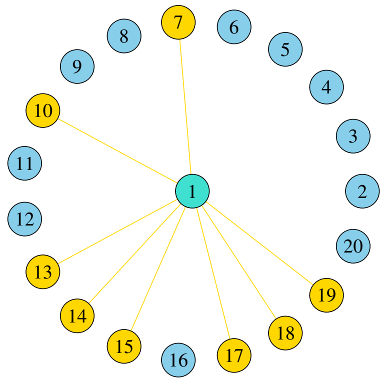

Figure 6: User-optimal feed.

Here we present the feeds user 1 would observe in a platform of size $n = 20$ (Figure 3) 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 4, 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 6, 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 5, 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.

Finally, let us examine how the user-optimal algorithm changes under the same parameter specification as λ varies. Naturally, when $\lambda = 1$, the user-optimal algorithm aligns with the platform-optimal algorithm, but the two diverge progressively as the user places greater emphasis on learning.

User's utility maximizer feed		
λ	**Feed**	**Expected Engagement**
0	1 18 20 7 10 14 13 15 19 17 3 5 11 6 8 9 16 4 2 12	7.716777
0.2	1 18 7 10 14 13 15 19 17 5 3 6 11 9 16 8 4 12 2 20	7.838447
0.4	1 7 10 14 13 15 19 17 5 3 6 11 9 16 8 4 12 2 20 18	8.135309
0.6	1 7 10 14 13 15 19 5 17 3 6 9 11 16 8 4 12 2 20 18	8.137124
0.8	1 7 10 14 13 15 19 5 17 3 9 6 16 11 8 4 12 2 20 18	8.137390

$$\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}$$