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

Miguel Risco† Manuel Lleonart-Anguix‡

November 10, 2023

Latest version available here

Abstract

In this paper, a social media platform governs the exchange of information among users with preferences for sincerity and conformity by providing personalized feeds. We show that the pursuit of engagement maximization results in the proliferation of echo chambers. A monopolistic platform implements an algorithm that disregards social learning and provides feeds that primarily consist of content from like-minded individuals. We study the consequences on learning and welfare resulting from transitioning to this algorithm from the previously employed chronological feed. While users' experience improves under the platform's optimal algorithm, social learning is worsened. Indeed, learning vanishes in large populations. However, the platform could create value by using its privileged information to design an algorithm that balances learning and engagement, maximizing users' welfare. We discuss interoperability as a possible regulatory solution that would eliminate entry barriers in platform competition caused by network effects, thereby inducing competing platforms to adopt the socially optimal algorithm.

Keywords: personalized feed, social learning, network effects, interoperability

JEL Codes: D43, D85, L15, L86.

*Risco acknowledges financial support by the German Research Foundation (DFG) through CRC TR 224 (Project B05). We thank Sven Rady, Pau Milán, Francesc Dilmé, Alexander Frug, Carl-Christian Groh, David Jiménez-Gómez, Daniel Krähmer, Philipp Strack, 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 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, UPF, TSE, UAB and the University of Valencia for their helpful comments and discussions.

†Bonn Graduate School of Economics, University of Bonn. risco@uni-bonn.de

‡IDEA (Autonomous University of Barcelona) and BSE. manuel.lleonart@bse.eu

1 Introduction

More than a third of Americans get their news from Facebook, which has 2.9 billion users worldwide (Gottfried and Shearer, 2019). Instagram, another platform owned by Meta, has 1.1 billion users. While both platforms were once considered harmless places for communication and content sharing, the current consensus suggests otherwise. They have been criticized for causing polarization and spreading misinformation, promoting echo chambers, and fueling hate speech (Silverman, 2016; Allcott and Gentzkow, 2017). The 2016 US presidential election was a significant example, as Facebook was accused of failing to combat fake news (Solon, 2016).

However, the biggest issue seems not to be just the leniency of the platforms, but also the impact of personalized content (Faelens et al., 2021). The *news feed* is a customized scroll of friends' content and news stories that appears on most social media platforms. Until around 2016, it used to be chronological (posts were displayed in the order they were written).[1] Now, a proprietary algorithm controls what appears on the screen, considering factors such as the users' friends, joined groups, liked pages, targeted advertisers, and popular stories to provide a personalized feed.[2] Since platforms' revenues come from advertising, their primary goal is to maximize engagement, which may not align with promoting informative communication. When the feed was chronological, observing just a few friends from one's neighborhood could provide a representative view of it. However, this is no longer the case, and it seems likely that under personalized feeds, agents will be biased and make incorrect extrapolations (Bandy and Diakopoulos, 2021). The following quote illustrates this tension (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". Thus, there is a need for research that guides the optimal regulatory approach, understanding the incentives of platforms in designing their optimal algorithm and how they would respond to regulation.

Two main trends have emerged when proposing policies to alleviate this situation: intervening directly on the information that is passed on and intervening on the platform structure. Information-targeting policies include censoring, fact-checking, nudging, or providing platform-generated content. Structural interventions could consist of capping how many times information can be relayed (the depth of the network) or the number of others with whom a typical user shares information (the breadth of the network), shut-

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

2 For an in-depth investigation of Facebook's algorithm and its negative effects, see Horwitz et al. (2021), which is based on an internal document review.

ting down certain communities (deplatforming), or regulating the level of homophily that the algorithm can induce. Our work follows a market structure approach, examining the incentives of an engagement-maximizing monopolist platform when providing users who have preferences for expressing their opinion truthfully and to conform with others in their personalized feed. We find that the monopolist optimally provides a feed in which the messages of those friends with the most similar views are shown, which often worsens social learning with respect to the traditional feed where users observe friends' messages in chronological order.[3] In fact, as the population grows larger, the platform's optimal algorithm leads to no learning. It simply matches users with those who perfectly conform to their views. As this algorithm is sub-optimal for user welfare, we discuss whether competition can incentivize platforms to adopt the socially optimal "user optimal" algorithm (the one that maximizes user welfare).

Social media platforms feature network effects (the more people use a platform, the more valuable the platform's services become), which result in high barriers to entry and induce a winner-takes-all (or most) market dynamic. To address this issue, we propose interoperability. Two platforms are interoperable if the users of one are able to interact with the users of the other. This is the case, for instance, in the cell phone and email industries. If an entering email company was not permitted to interconnect with Google or Yahoo, the mail account it provides would be much less useful than a Google or Yahoo mail account. With interoperability, the entry barrier would be removed and competition could discipline platforms, forcing them to implement the socially optimal algorithm.

It is worth noting that most successful platforms, such as Facebook, Instagram, TikTok, and Twitter,[4] are enormous in size and monopolies in their respective fields.[5] Although these platforms 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 (Twitter). While network effects are mainly the cause for this market tipping, personalized algorithms account for the increase in engagement and addictive behavior in such platforms, independently of their field (Guess et al., 2023). Moreover, it is rooted in the public consciousness that personalized feeds create echo chambers, foster polarization and enable the spread of misinformation, even though their opaqueness poses a challenge in substantiating such claims.

3 Most platforms, among them Facebook or Instagram, provide users with a personalized feed by default. Twitter, on the other hand, allows allows users to choose their feed's display mode directly in the timeline screen.

4 Even though Twitter has been rebranded to X, we stick in this paper to the well known old name.

5 Regarding monopoly structures in the social media platform market, a quote from the Bundeskartellamt (the German competition protection authority) in its case against Facebook (B6-22/16, "Facebook", p. 6) states: "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).

Still, there is crescent evidence on the harm platforms cause, both at the platform level, as Horwitz et al. (2021) report ("[t]ime and again, the documents show, Facebook's researchers have identified the platform's ill effects"), and at the consumer level, as Bursztyn et al. (2023) show (they find that users would be willing to pay to have others, including themselves, deactivating their TikTok and Instagram accounts). However, platforms are able to create value through their superior level of information.[6] Thus, it is essential to investigate how personalization algorithms affect social welfare, as their repercussions have emerged as a significant economic concern.

We highlight three main contributions of this paper. First, we build a model of *communication and learning through personalized feed* where a strategic engagement-maximizing platform chooses an algorithm that provides personalized feeds and users post messages. We assume that an individual (she) joins a social media platform for two main reasons. First, to engage in communication with peers about some underlying topic. This social activity generates utility through two channels: 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 neighbors have shared; *conformity*).[7] The strength of these incentives depends on model parameters. In particular, we encompass situations in which conformity is almost negligible. Second, to learn some valuable information (the state of the world) and make the best decision outside the platform.[8] The effectiveness of the Covid-19 vaccine, which triggered significant public debate,[9] is our leading example. People have used social media platforms to express their views about the benefits and risks of vaccination, driven by both a desire for sincerity and conformity, as well as the need for information. Learning was needed to make the best possible decision on getting vaccinated or not. Simultaneously, expressing dissenting opinions proved to be socially taxing, as individuals were hesitant to differ from their peers. Users sought to communicate their personal viewpoints or knowledge, recognizing that vaccination was a pivotal societal concern.

6 Quoting Scott Morton et al. (2019): "The speed, scale, and scope of the internet, and of the ever-more powerful technologies it has spawned, have been of unprecedented value to human society."

7 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.

8 The first component of the utility function is similar to the payoffs in Galeotti et al. (2021), where agents prefer taking actions closer to those of their neighbors and to their own ideal points. Utility is given by a weighted average of two loss functions representing *miscoordination* and *distance from favourite action*, and the action is not necessarily a message, as it is in our within-the-platform utility. However, one of their motivating examples perfectly fits our model: "the action 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".

9 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.

Because of advertising revenues, profit maximization leads platforms to user engagement maximization.[10] We define user engagement as being equal to the payoffs users derive from interacting on the platform. In contrast, learning is seen as a long-term reward and does not affect engagement. This implies that the platform has no innate incentives to foster learning. Deciding on how much time to spend scrolling down (for which the user only takes into account the immediate joy of interacting within the platform) entails a different mental process than that of deciding on which content to post or even which platform to register on.[11] Inside the platform, users communicate by posting messages about the topic of interest and then learn through reading the messages that appear in their personalized feed. The feed is a subset of the population from which she reads the messages. It is designed by the platform, which leverages its information on users' similarities in views. We assume the platform knows perfectly how similar users are, and utilizes this information to maximize its profits. We might think of similarities being derived from past interactions and users' personal data by using sophisticated machine learning techniques.[12]

Our second contribution is to study the properties of the platform's optimal algorithm. We show that it displays to each user the messages of their most similar neighbors. Remarkably, this result persists even when conformity has minimal weight. We refer to this algorithm as the "closest" algorithm. In turn, users report truthfully their innate opinion. Surprisingly, this algorithm does not always harm social learning compared to the traditional chronological feed. However, when the network grows large, the closest algorithm provides no learning: a user would be matched with *copies* of herself. This contrasts with classical results where large societies learn better (Golub and Jackson, 2010). Remarkably, users are generally satisfied with this outcome, as the payoff stream coming from conformity is maximized. Still, the closest algorithm is socially sub-optimal, so we analyze how to improve it in this asymptotic context and propose a measure to break echo chambers. It consists of adding a user with opposite views to each feed. This would improve learning (to the point of making it perfect) at a low cost coming from conformity. The "breaking echo chambers" algorithm outperforms both the closest and the chronological algorithms in most cases. However, it is plausible that real-world users may disregard information from a completely opposing source, making its practical

10 An ad-revenue-maximizing platform as in Mueller-Frank et al. (2022) maximizes revenues by maximizing the number of users that observe advertisements.

11 This assumption could be also interpreted following the *Dual Process Theory* as in Benhabib and Bisin (2005). Becoming engaged is a rather automatic process that corresponds with the intrinsic happiness derived within the platform, while taking explicit decisions (posting messages or eventually choosing whether to register in a platform) is a control process where the user rationally acts to achieve a goal. For an overview on Dual Process Theory, see Grayot (2020).

12 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".

enforcement complicated. Moreover, the user optimal algorithm is still socially preferred, and hence we must consider its practical implementation.

Our third contribution is to discuss whether competition would suffice for platforms to implement the user optimal algorithm. Due to network effects, this is not necessarily the case. We show that under almost all conditions, both the user optimal algorithm and the closest algorithm feature network effects, which creates entry barriers that protect large incumbents (monopolists) and deter competition. We propose interoperability to eliminate entry barriers by shutting down network effects at the firm level. Interoperability compels platforms to connect, so that users from different platforms can be neighbors. A user's feed would consist in a subset of all her friends, independently of which platform they are registered in, designed by the platform she has joined. Hence, users will choose the platform that provides the best service (in this paper, the best algorithm), disregarding how many users it has. Then, competing platforms would be *forced* to implement the user optimal algorithm; otherwise, they would risk losing their user base. The pursuit of this goal aligns with the intentions of EU regulators, as reflected in the Digital Markets Act,[13] which mandates certain large social platforms to achieve interoperability in their messaging communications in the immediate future. Quoting Kades and 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, Section 2 introduces the model. Section 3 analyzes the equilibrium of the game and Section 4 characterizes the main algorithms appearing in the paper. Section 5 deals with the robustness of the model through some extensions, while Section 6 discusses interoperability in the context of platform competition. Section 7 concludes.

1.1 Related literature

The effects of personalized feeds on social welfare have not, to the best of our knowledge, been studied from a theoretical perspective. However, a recent paper by Guess et al. (2023) examines the empirical effects of Facebook's and Instagram's feed algorithms. The study reveals that transitioning users back to chronological feeds decreases the time they spend on the platforms as well as their overall activity (i.e., engagement). Additionally, it leads to a reduction in the proportion of content derived from ideologically like-minded sources, thereby diminishing the impact of the echo-chamber effect.

13 See regulation (EU) 2022/1925 of the European Parliament and of the council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828.

In broad terms, our paper is related to two areas of literature. The first area studies the impact of revenue-maximizing platforms on social learning. This is a growing field, and we highlight two papers for their similarities to our work. Mueller-Frank et al. (2022) build a model of network communication and advertising where the platform controls the flow of information. In equilibrium, the platform may manipulate or even suppress information to increase revenue, even though this ultimately decreases social welfare. In a model where agents decide whether or not to pass on (mis)information, Acemoglu et al. (2021) studies the algorithm choice of an engagement-maximizing platform. They show that when the platform has the ability to shape the network, it will design algorithms that create more homophilic communication patterns. Thus, in line with our results, both papers find that platforms' incentives are not aligned with users' preferences and that engagement-maximizing behavior harms social welfare. Homophilic communication patterns, commonly known as echo chambers or "filter bubbles", were introduced in Pariser (2011): to increase metrics like engagement and ad revenue, recommendation systems connect users with information already similar to their current beliefs. This hypothesis is further discussed in Sunstein (2017), while Chitra and Musco (2020) experimentally check its effects on polarization and show the large impact of minor algorithm changes. Relatedly, Demange (2023) shows that platforms promote the visibility of their most influential individuals. Additional research on media platforms providing distorted content for economic reasons can be found in Reuter and Zitzewitz (2006), Ellman and Germano (2009), Abreu and Jeon (2019), and Kranton and McAdams (2022). Hu et al. (2021) shows that rational, inattentive users prefer to learn from like-minded neighbors, while Törnberg (2018) shows that echo chambers harm social welfare by increasing the spread of misinformation.

Not just the mentioned literature, but also empirical work (Sagioglou and Greitemeyer, 2014; Levy, 2021) reveals the need for further intervention or regulation on social media platforms. This topic constitutes the second strand to which our paper is closely related. Franck and Peitz (2023) provides perspective on social media platform competition, claiming that market power (mainly represented by the network effects) leads to suboptimal outcomes for society. It suggests that it may not be the platform with the best offer that dominates the market. Biglaiser et al. (2022) provides a micro-foundation for incumbent advantage. Essentially, network effects prevent users to migrate to even Pareto-superior equilibria if they receive stochastic opportunities to migrate to an entrant. This is also covered in Kades and Morton (2020), who also provide an overview on interoperability for digital markets. Popiel (2020) and Evens et al. (2020) assert that regulations to manage digital platform markets in the US and EU, respectively, are inadequate in addressing their negative effects. In response to this need, there has been a surge of recent papers examining interventions. Regarding structural interventions, Jackson et al. (2022) examines how limiting the breadth and/or depth of a social network improves message accuracy. The work of Benzell and Collis (2022) aligns with our own, as they analyze the optimal strategy of a monopolistic social media platform and evaluate the impact of taxation and regulatory policies on both platform profits and social welfare. However, in their paper, the platform chooses net revenue per user rather than shaping communication among users. The authors apply their model to Facebook and find that a successful regulatory intervention to achieve perfect competition would increase social welfare by 4.8%. Finally, Agarwal et al. (2022) provides empirical evidence of the negative consequences of deplatforming (shutting down a community on a platform), mainly due to migration effects, which supports a call for globally applicable regulations.

There is a plethora of recent empirical contributions regarding informational interventions: Habib et al. (2019), Hwang and Lee (2021) or Mudambi and Viswanathan (2022). Mostagir and Siderius (2023b) models community formation and shows that the effect of interventions is non-monotonic over time. Additionally, there is another important aspect to consider when analyzing informational policies: Mostagir and Siderius (2022) demonstrates that cognitive sophistication matters when faced with misinformation, and Mostagir and Siderius (2023a) finds that different populations (Bayesian and DeGrootians) react differently to certain interventions. While some papers, such as Mostagir and Siderius (2023a), include cases where sophisticated users are outperformed by their naive counterparts, Pennycook and Rand (2019, 2021) show that higher cognitive ability is associated with better ability to discern fake content. In our model, the results hold for both Bayesian and DeGrootian users, but the sophisticated agents always learn better. Finally, we also relate to the literature on learning in networks, for both naive and sophisticated users: DeMarzo et al. (2003), Acemoglu and Ozdaglar (2011), Jadbabaie et al. (2012), Molavi et al. (2018) or Mueller-Frank and Neri (2021).

2 Model

We first outline the model. There is an underlying state of the world denoted by $\theta \in \mathbb{R}$, and every user receives a private signal about it. Users join a social media platform to communicate through messages, deriving utility through two streams. Within-the-platform utility depends on the posted messages and the users' original opinions about θ. The larger the within-the-platform utility, the more engaged a user is. Outside-the-platform utility (or *action utility*), is given by a quadratic loss function accounting for the deviation of the user's action from θ. Such an action is based on both private signals and the messages learnt within the platform. We interpret within-the-platform utility as *media user* utility, representing the immediate payoff a consumer obtains while spending time in the platform. Total utility, on the other hand, corresponds to a sort of *citizen* utility, capturing the overall payoff an agent derives from her entire experience on the platform and the subsequent behavior.[14] The platform's revenue is given by total engagement,

14 The concepts of *media user* utility and *citizen* utility are taken from Pariser (2011).

so its primary concern is maximizing users' within-the-platform utility. To achieve this goal, the platform selects a subset of neighbors whose messages appear in each user's personalized feed by leveraging information on users' similarity.

Now, let us describe the model in detail. There is a set, N, of n users that join a social media platform described by the undirected graph $\mathcal{G}$. Each node in the graph represents a user in the platform, and two nodes are linked if they are friends on the platform. In this paper, the network structure does not play a significant role, so we will consider a complete network moving forward.[15] Every user is friend of the others and hence neighborhoods equal the whole network. Each user receives a private signal $\theta_i \in \mathbb{R}$ about θ. 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 with $\Sigma_{ii} = \sigma^2$ for every i. We denote by ρ_{ij} the correlation between θ_i and θ_j, so that $\rho_{ij} = \frac{\sigma_{ij}}{\sigma^2}$ for every i, j. The random variable θ_i is user i's private signal on θ, intepreted as her private opinion on the state of the world.[16] Users know their private signals, the distribution of all signals, the covariance matrix $\boldsymbol{\Sigma}$, and the distribution of the state of the world: we assume improper priors on θ.[17] Thus, user i's posterior distributions are $\theta_j|\theta_i \sim \mathcal{N}\left(\theta_i, \sigma^2(1-\rho_{ij})\right)$ for all $j \in N$ and $\theta|\theta_i \sim \mathcal{N}(\theta_i, \sigma^2)$.

Regarding communication, each user i simultaneously posts a message $m_i \in \mathbb{R}$. Then, she observes her personalized feed. This subset is strategically chosen by the platform through an algorithm that will be defined later. In reality, social media platforms rank friends' posts and users observe as many as their scrolling time allows them to. For the purpose of this analysis, we assume that all users spend a fixed amount of time reading

15 All our results hold if we consider a general network. The only necessary condition is that the algorithm has some neighbors to choose from when constructing the feed. However, since a general network structure does not offer any particular insight, we opt to work with a complete network for convenience.

16 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, that 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 $\boldsymbol{\Sigma}$.

17 For a discussion of improper priors, see Hartigan (1983). The distribution of the prior is assumed to be uniformly distributed along the whole real line. The technical consequence of this assumption is that the posterior distribution of the state of the world conditional on one's signal is normal. Intuitively, one could think that every user considers herself as central; no extreme user would acknowledge that she is, indeed, extreme.

messages, i.e., the number of observed messages is exogenously fixed, and that the number of available friends' posts exceeds it. Without loss of generality, we assume that every user observes the same number of messages $k \in \mathbb{N}$. The key assumption here is exogeneity: neither the users nor the platform are able to adjust the number of messages that will be read.[18] Although user behavior on social media platforms is highly diverse, existing data support these assumptions.[19]

Users derive utility after communication takes place. Their within-the-platform utility has three components: (i) a fixed and common positive value for joining the platform, v; (ii) *sincerity*: agents dislike deviating from their own signals,[20] and (iii) *conformity*: matching others' opinions is rewarded. Formally, user i's realized within-the-platform utility is

$$u_i(m_i, m_{-i}, \mathcal{S}_i, \theta_i) = v - \beta \overbrace{(\theta_i - m_i)^2}^{\text{Sincerity}} - (1-\beta) \overbrace{\sum_{j \in \mathcal{S}_i} \frac{(m_i - m_j)^2}{k}}^{\text{Conformity}}, \tag{1}$$

where $\beta \in (0, 1)$ represents how much sincerity is weighted with respect to conformity, and $\mathcal{S}_i \subset N$ denotes user i's personalized feed. We assume the cardinality of all $\{\mathcal{S}_i\}_{i=1}^n$ to be exogenously set to $k \in \mathbb{N}$, where $k \leq |N|$.[21] Within-the-platform utility is not the only source of utility for users, as they are also concerned about taking an action that matches the state of the world. Platform communication allows them to learn about θ and optimize the decision making proccess. Total realized utility is the weighted average of within-the-platform utility and action utility (the squared distance of the action from the state of the world):

$$U_i(m_i, m_{-i}, a_i, \mathcal{S}_i, \theta_i, \theta) = \lambda \left[v - \beta(\theta_i - m_i)^2 - (1-\beta) \sum_{j \in \mathcal{S}_i} \frac{(m_i - m_j)^2}{k} \right] - (1-\lambda) \overbrace{(a_i - \theta)^2}^{\text{Action utility}}, \tag{2}$$

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 $\{m_j\}_{j \in \mathcal{S}_i}$, chooses an action a_i to maximize the conditional expectation of U_i. The optimal action a_i^* is the best guess of θ conditional on the observed messages.

Next, we introduce the platform as a strategic agent of the game. It knows the distributions and $\mathbf{\Sigma}$, but not θ or $\{\theta_i\}_{i=1}^n$. For each user i it chooses the set of neighbors

---

18 If we consider $k_i \neq k_j$, all the results of the paper go through.

19 According to the statistics portal *Statista*, an average Facebook user has 335 friends and spends 35 minutes on the platform, reading between 10 and 50 posts. However, at least 300 stories are produced in her network ($k << N$). Regarding the assumption on a fixed k, *Statista* says that, on average, users tend to consistently allocate a similar amount of screen time to social media platforms.

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

21 Should we consider a general network, it is assumed that $k \leq \min_{i \in N} |N_i|$ for every neighborhood N_i.

that she will observe messages from. As explained above, such set is referred to as personalized feed and is denoted by $\mathcal{S}_i$. We define the platform's *algorithm* as the collection of feeds provided to users: $\mathcal{F} = (\mathcal{S}_1, ..., \mathcal{S}_n)$. The set of algorithms is denoted by $\mathcal{A}$.

In this paper, we assume user's engagement equals user's within-the-platform utility. The intuition hinges on the fact that the happier a user is inside the platform, the higher the probability of her spending time reading more messages or coming back in the subsequent periods. In other words: we think of users who are not forward-looking when choosing how many messages to read (i.e., how much time they spend in the platform or simply how much they engage). Thus, at this stage, they do not take long-term learning into account but myopically value only the intrinsic incentives of the platform. This is in line with the main case in Bonatti and Cisternas (2020), where consumers ignore the link between their current actions and the future consequences. We refer to Appendix B for an extension showing that if the user endogenously chooses her engagement level k, the platform-optimal algorithm does not vary. Platforms' revenues come essentially from advertising, which depends directly on total engagement. Hence, the platform objective is to maximize total engagement and its profit function is given by:

$$\Pi_p(m, \mathcal{F}) = \sum_{i=1}^{n} \mathbb{E}_p[u_i(m_i, m_{-i}, \mathcal{S}_i, \theta_i)],$$

where $\mathbb{E}_p$ denotes the expected value conditional on the platform's information.

In the main model of this paper, where there is just one monopolistic platform, all users join it automatically. Hence, it does not need to focus on attracting users but rather on maximizing their within-the-platform utility. Bursztyn et al. (2023) show that because of Fear of Missing Out (FOMO), users may join a platform even when they would prefer it not to exist (in the paper, they show that users would be willing to pay to have others, including themselves, deactivating TikTok and Instagram). Hence, when the outside option is to be alone, it is reasonable to assume that all users would join. However, in Subsection 5.1, we formally show that if society is large enough, every user joins when we introduce a previous step in which they have to decide whether to enter the monopolistic platform or to stay out (the outside option yields no penalty for conformity or sincerity but also no learning).

In summary, the game of *communication and learning through personalized feed* described above consists of the following sequence of events:

1. The platform chooses an algorithm $\mathcal{F}$ 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 the messages in her feed $\mathcal{S}_i$ and chooses an action a_i.
5. The state of the world is revealed and payoffs are realized.

3 Equilibrium

Here we describe and analyze the equilibrium of the game. The equilibrium concept is Bayesian Nash equilibrium. More specifically, we focus on Bayesian Nash equilibrium in which every user plays a linear messaging strategy (i.e., her message is a linear function of her private signal). In particular, we establish that against a profile of linear messaging strategies, it is optimal to play linear. The platform chooses an algorithm $\mathcal{F} \in \mathcal{A}$. In turn, each user decides on a message $m_i : \mathbb{R} \times \mathcal{A} \to \mathbb{R}$ (which is a function of her private signal and the algorithm) and an action $a_i : \mathbb{R}^k \times \mathcal{A} \to \mathbb{R}$ (which is a function of her private signal, the $k-1$ messages appearing in her feed, and the algorithm). Note that the algorithm is publicly disclosed before the users decide on their strategies. As we show below, in equilibrium users truthfully report their private signal for any algorithm, and the platform-optimal algorithm is the one that displays, for each user, the messages of those neighbors who feature the highest correlation with her (in other words, the most similar, or the *closest* friends). We refer to such an algorithm as the "closest" algorithm and denote it by $\mathcal{C}$.

Before formally deriving the equilibrium of the game, let us show two auxiliary results. First, we prove that when users report their types truthfully, it is equivalent for the platform to maximize profits and to maximize each user's within-the-platform utility separately (i.e., there are no inter-dependencies across feeds). This result also implies that an algorithm that maximizes each user's individual utility is precisely the algorithm that an utilitarian social planner would implement.

Lemma 3.1. *If $m_i = \theta_i$ for all $i \in N$,then*

$$\underset{(\mathcal{S}_1,\ldots,\mathcal{S}_n)}{\operatorname{argmax}} \left\{ \sum_{i=1}^{n} \mathbb{E}_p[u_i] \right\} = \left(\underset{\mathcal{S}_1 \subseteq N}{\operatorname{argmax}} \{\mathbb{E}_p[u_1]\}, \ldots, \underset{\mathcal{S}_n \subseteq N}{\operatorname{argmax}} \{\mathbb{E}_p[u_n]\} \right).$$

Proof. Truthtelling implies that

$$\max_{(\mathcal{S}_1,\ldots,\mathcal{S}_n)} \left\{ \sum_{i=1}^{n} \mathbb{E}_p[u_i] \right\} = \max_{(\mathcal{S}_1,\ldots,\mathcal{S}_n)} \left\{ -\sum_{i=1}^{n} \left(\frac{1-\beta}{k} \sum_{j \in \mathcal{S}_i} \mathbb{E}_p[(\theta_i - \theta_j)^2] \right) \right\}$$
$$= \sum_{i=1}^{n} \left(\max_{\mathcal{S}_i} \left\{ -\frac{1-\beta}{k} \sum_{j \in \mathcal{S}_i} \mathbb{E}_p[(\theta_i - \theta_j)^2] \right\} \right) = \sum_{i=1}^{n} \left(\max_{\mathcal{S}_i} \{\mathbb{E}_p[u_i]\} \right).$$

□

Next, we derive the distribution of θ conditional on the messages observed in user i's personalized feed if there is truthful reporting, i.e., conditional on $\{\theta_j\}_{j \in \mathcal{S}_i}$.

Lemma 3.2. *The posterior distribution of θ conditional on $\{\theta_j\}_{j \in \mathcal{S}_i}$ is given by*

$$\theta | \{\theta_j\}_{j \in \mathcal{S}_i} \sim \mathcal{N} \left(\frac{\mathbb{1} \boldsymbol{\Sigma}_{\mathcal{S}_i}^{-1} \boldsymbol{\theta}_{\mathcal{S}_i}}{\mathbb{1} \boldsymbol{\Sigma}_{\mathcal{S}_i}^{-1} \mathbb{1}^t}, \frac{1}{\mathbb{1} \boldsymbol{\Sigma}_{\mathcal{S}_i}^{-1} \mathbb{1}^t} \right),$$

where $\mathbb{1}$ is an n-vector of ones, $\mathbf{\Sigma}_{\mathcal{S}_i}$ is the submatrix of $\mathbf{\Sigma}$ induced by $\mathcal{S}_i$, and $\boldsymbol{\theta}_{\mathcal{S}_i}$ is the subvector of $(\theta_j)_{j\in N}$ induced by $\mathcal{S}_i$.

Proof. See Appendix C. □

Next, we derive the equilibrium of the game.

Proposition 3.3. *The unique linear Bayesian Nash equilibrium is given by users playing $m_i^* = \theta_i$, $a_i^* = \frac{\mathbb{1}\mathbf{\Sigma}_{\mathcal{S}_i}^{-1}\boldsymbol{\theta}_{\mathcal{S}_i}^t}{\mathbb{1}\mathbf{\Sigma}_{\mathcal{S}_i}^{-1}\mathbb{1}^t}$ and the platform playing $\mathcal{C}$, the closest algorithm.*

Proof. We analyze first the optimal message for the user.[22] Given user i's type θ_i, the algorithm $\mathcal{F}$, the matrix $\mathbf{\Sigma}$, and other users' messages $m_{-i}(\theta_{-i})$, she chooses a message $m_i \in \mathbb{R}$ that maximizes her expected within-the-platform utility, as action utility is not affected by this choice:

$$\mathbb{E}_i[u_i(m_i, m_{-i}, \mathcal{S}_i, \theta_i)|\theta_i, \mathcal{F}] = v - \beta(\theta_i - m_i)^2 - (1-\beta)\frac{1}{k}\mathbb{E}_i\left[\sum_{j\in\mathcal{S}_i}(m_i - m_j(\theta_j))^2|\theta_i, \mathcal{F}\right]$$

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

The first order condition with respect to m_i yields

$$m_i = \beta\theta_i + (1-\beta)\frac{1}{k}\sum_{j\in\mathcal{S}_i}\mathbb{E}_i\left[m_j(\theta_j)|\theta_i, \mathcal{F}\right]. \tag{3}$$

Assuming linear messaging strategies for all users except i (i.e., $m_j(\theta_j) = \gamma_j\theta_j + \delta_j$ for some $\gamma_j, \delta_j \in \mathbb{R}$ and all $j \neq i$), we can work further on the expectation term in (3) for each $j \in \mathcal{S}_i$ (note that the user knows both the algorithm and the covariance matrix Σ, so she also knows which neighbors will appear in her feed) and obtain:

$$\mathbb{E}_i\left[m_j(\theta_j)|\theta_i, \mathcal{F}\right] = \mathbb{E}_i\left[\gamma_j\theta_j + \delta_j|\theta_i, \mathcal{F}\right] = \gamma_j\mathbb{E}_i[\theta_j|\theta_i, \mathcal{F}] + \delta_j = \gamma_j\theta_i + \delta_j.$$

Plugging this into (3) yields

$$m_i = \beta\theta_i + (1-\beta)\frac{1}{k}\sum_{j\in\mathcal{S}_i}(\gamma_j\theta_i + \delta_j) = \beta\theta_i + (1-\beta)(\bar{\gamma}_i\theta_i + \bar{\delta}_i),$$

where $\bar{\gamma}_i := \frac{1}{k}\sum_{j\in\mathcal{S}_i}\gamma_j$ and $\bar{\delta}_i := \frac{1}{k}\sum_{j\in\mathcal{S}_i}\delta_j$. Hence, user i's optimal strategy is linear: $m_i(\theta_i) = \gamma_i\theta_i + \delta_i$. This leads to the system of equations

$$\begin{cases} \gamma_i &= \beta + (1-\beta)\bar{\gamma}_i \\ \delta_i &= (1-\beta)\bar{\delta}_i \end{cases} \quad \forall i \in \{1, \ldots, n\}.$$

22 As the utility function is additive separable and the choice of m_i does not affect that of a_i and vice versa, we can study the optimal decisions independently.

Its unique solution is $\gamma_i = 1$ and $\delta_i = 0$ for all $i \in N$. Thus, the optimal message is $m_i^* = \theta_i$ and every user reports truthfully her type. Now, leveraging its knowledge of Σ, the platform chooses an algorithm $\mathcal{F}$ such that

$$\mathcal{F}^* = \underset{\mathcal{F}=(\mathcal{S}_1,\ldots,\mathcal{S}_n)}{\operatorname{argmax}} \left\{ -\sum_{i=1}^{n} \left(\frac{1}{k} \sum_{j\in\mathcal{S}_i} \mathbb{E}_p \left[(\theta_i - \theta_j)^2 \right] \right) \right\}.$$

But by Lemma 3.1, this is equivalent to maximizing each user's within-the-platform utility, i.e., to finding

$$\begin{aligned} \mathcal{S}_i^* &= \underset{\mathcal{S}_i\subseteq N}{\operatorname{argmax}} \left\{ -\sum_{j\in\mathcal{S}_i} \mathbb{E}_p \left[(\theta_i - \theta_j)^2 \right] \right\} = \underset{\mathcal{S}_i\subseteq N}{\operatorname{argmax}} \left\{ -\sum_{j\in\mathcal{S}_i} (2\sigma^2 - 2\sigma_{ij}) \right\} \\ &= \underset{\mathcal{S}_i\subseteq N}{\operatorname{argmax}} \left\{ \sum_{j\in\mathcal{S}_i} \sigma_{ij} \right\} \quad \forall\, i. \end{aligned}$$

The algorithm that induces such feeds is precisely $\mathcal{C}$, the closest algorithm. The platform displays to each user the messages of those k neighbors who exhibit the highest correlation with her. Thus, in equilibrium, $m_i^* = \theta_i$ for all i and $\mathcal{F}^* = \mathcal{C}$. Next, we calculate user i's optimal action, which maximizes $\mathbb{E}[(a_i - \theta)^2]$ given the observed messages $\theta_{\mathcal{S}_i}$. Hence, the optimal action is $a_i^* = \mathbb{E}_i[\theta|\theta_{\mathcal{S}_i}] = \frac{\mathbb{1}\boldsymbol{\Sigma}_{\mathcal{S}_i}^{-1}\boldsymbol{\theta}_{\mathcal{S}_i}^t}{\mathbb{1}\boldsymbol{\Sigma}_{\mathcal{S}_i}^{-1}\mathbb{1}^t}$ by Lemma 3.2. □

For an explicit example of how the platform designs $\mathcal{C}$ leveraging $\boldsymbol{\Sigma}$, please refer to Appendix A. Note that the second-order moments of θ_j still depend on j (the posterior variance is $\sigma^2(1-\rho_{ij})$), but the first-order moments do not (the posterior mean is always θ_i). Therefore, the platform cannot affect the message choice by providing feeds, but only the action.

Corollary 3.4. *For any algorithm $\mathcal{F} \in \mathcal{A}$, users play truthtelling.*

Indeed, even if the algorithm is unknown and the user does not know who she will be matched to, truthtelling is still the equilibrium strategy. In expectation, every other user is equivalent for user i, and then $\mathbb{E}_i [m_j(\theta_j)|\theta_i] = \gamma_j \mathbb{E}_i[\theta_j|\theta_i] + \delta_j = \gamma_j\theta_i + \delta_j$. Hence, user i plays $m_i^* = \theta_i$ and it is the platform's best reply to implement $\mathcal{C}$.

Corollary 3.5. *If the algorithm is not publicly disclosed, users play truthtelling and the platform implements the closest algorithm $\mathcal{C}$.*

The platform's practice of pairing each user with those who are most similar to her contributes to the formation of echo chambers: each individual learns the messages of her like-minded neighbors. These echo chambers align with the concept of "filter bubbles" in Pariser (2011). Pariser argued that individualized personalization through algorithmic filtering could lead to intellectual isolation and social fragmentation. In the next section, we delve into the consequences of the closest algorithm in terms of social welfare and social learning.

It is worth noting that these echo chambers emerge due to the inclination of truth-telling users to seek conformity. However, even when we significantly reduce the influence of conformity within the utility function (β close to 1), the equilibrium outcome remains unaltered, and echo chambers arise in the same manner.[23] Furthermore, in Subsection 5.2 we show that the platform has stronger preferences for feed similarities than the user. Finally, the result of engagement being maximized in equilibrium is consistent with the findings of Guess et al. (2023), who empirically demonstrate that personalization algorithms significantly increased user engagement compared to chronological feeds.

4 Algorithms and learning

Under $\mathcal{C}$, each agent observes a feed composed of her most similar neighbors, her echo chamber. Still, she benefits from the k messages received to learn about θ. Opting to join the platform always improves expected action utility compared to an outside option in which she only knows her private signal θ_i. However, we are curious about the extent to which this improvement occurs, particularly in relation to population growth. The monopolistic platform prioritizes user engagement and disregards the effect of a_i on the user's total utility. As a result, it selects the most similar users to design the feed. When the pool from which the platform selects becomes infinitely large, it ends up choosing *copies* of the user to maximize conformity. Hence, the induced echo chamber provides her with no new information. We show that it is possible to improve on the closest algorithm by artificially adding to the user's feed an individual with opposite views, which enables her to learn the state of the world. We refer to this enhanced algorithm as the "breaking echo chambers" algorithm.

Now, let us study the economic consequences of the implementation of the closest algorithm. We call learning to the reduction on the expected value of the action utility resulting from reading messages. Choosing action $a_i = \mathbb{E}_i[\theta|\{\theta_j\}_{j\in\mathcal{S}_i}]$ implies that the expected value of the action utility is precisely the posterior variance of θ:

$$\mathbb{E}\left[(a_i-\theta)^2|\{\theta_j\}_{j\in\mathcal{S}_i}\right] = \mathbb{E}\left[(\mathbb{E}_i[\theta|\{\theta_j\}_{j\in\mathcal{S}_i}]-\theta)^2|\{\theta_j\}_{j\in\mathcal{S}_i}\right] = \mathrm{Var}\left[\theta|\{\theta_j\}_{j\in\mathcal{S}_i}\right].$$

We can compare learning through different algorithms by comparing the posterior variances they induce. In this section, we will compare $\mathcal{C}$ with the random algorithm, (that we denote $\mathcal{R}$), which is the chronological algorithm that used to be implemented before personalization algorithms gained traction in the late 2010s. The random algorithm provides user i with a feed comprising k randomly chosen messages from N. To explicitely compare both algorithms, we use Lemma 3.2: $\mathrm{Var}\left[\theta|\{\theta_j\}_{j\in\mathcal{S}_i}\right] = \frac{1}{\mathbb{1}\boldsymbol{\Sigma}_{\mathcal{S}_i}^{-1}\mathbb{1}^t}$.

23 When $\beta = 1$ (and only sincerity matters), the platform cannot influence engagement through its algorithm, and thus, it is indifferent regarding algorithm choice.

Given that the closest algorithm matches a user with her most similar neighbors, one might initially assume it to be the worst algorithm in terms of learning. Perhaps surprisingly, this is not always the case. There are certain signal structures where the closest algorithm can actually enhance learning compared to a particular feed (which can be seen as a specific realization of the random algorithm). A simple example illustrates such a scenario. Consider a small network composed by four individuals ($N = 4$), and a feed length of $k = 3$. 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 $\mathcal{S}_1^{\mathcal{C}} = \{1, 2, 3\}$. Assume that a particular draw from the random algorithm induces $\mathcal{S}_1^{\mathcal{R}} = \{1, 3, 4\}$. 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 random algorithm. In this case, learning is better under $\mathcal{C}$, the algorithm provided by the engagement-maximizing platform. This occurs because there are two forces that affect user 1's learning: her covariances with the others in her feed, but also the covariances among the others. These two forces may interact in such a way that a more similar individual to user 1 yields better learning based on her relationship with the rest of the users in the feed. However, this phenomenon only occurs when N is small. As the population size grows, the similarities to user 1 become the dominant force, and the closest algorithm performs worse. We can observe this in Figure 1, where we plot realizations of learning under $\mathcal{R}$ and $\mathcal{C}$ as N increases for a population growing from $N = k$ to $N = 5000$ and parameters $\lambda = 0.5$, $\beta = 0.2$ and $k = 30$.

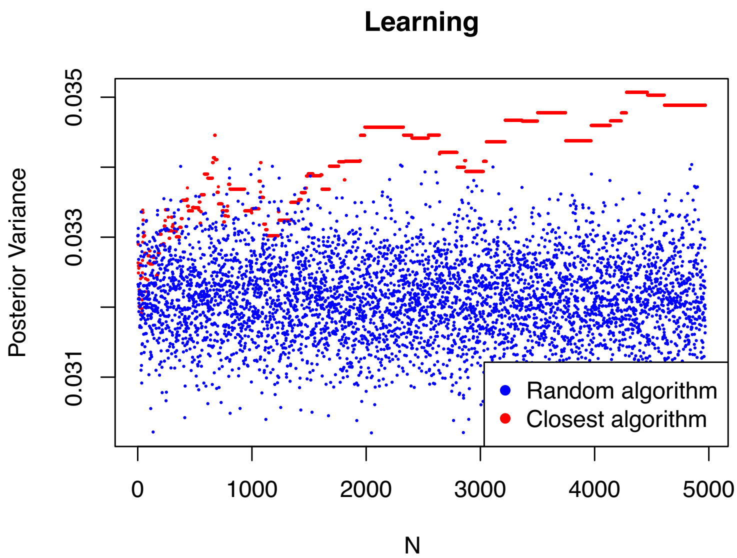

Figure 1: User's posterior variance as the population grows.

Let us focus now on algorithms' performance in large populations. Regarding the society expansion process, we assume that the covariances between new users (that are added one by one) and existing users are drawn from a continuous distribution with a cumulative distribution function that has support in $[-\sigma^2, \sigma^2]$ and is centered at 0. The resulting covariance matrix (the expanded $\mathbf{\Sigma}$) must be symmetric and positive definite.

Proposition 4.1. *Under the closest algorithm, user i's learning becomes negligible as $N \to \infty$:*

$$\lim_{N\to\infty} Var[\theta | \{\theta_j\}_{j\in\mathcal{S}_i^\mathcal{C}}] = \sigma^2.$$

Proof. See Appendix C. □

In general, learning does not monotonically change with neighborhood growth for a fixed k under the closest algorithm (see Figure 1). Although the platform can select *more* similar users, the degree of similarity between them also matters. As demonstrated in the example above, the correlations among the users appearing in the feed may counterbalance the increasing similarities with user i. Therefore, it is possible that adding a more similar neighbor enhances learning. However, this phenomenon disappears as N becomes larger, and learning non-monotonically diminishes, eventually leading to its absence in the limit. When the platform has an infinite pool of users to choose from, it provides a feed formed by *copies* of the user, so that she conforms with everyone. However, the user's posterior variance converges to σ^2 because, in expectation, she only reads the information she already has. Thus, learning vanishes and the user learns the same as if she were alone.

On the other hand, the expected posterior variance of the random algorithm is $\frac{\sigma^2}{k}$, and realizations vary around this value. Given that the probability that a realization of the random algorithm coincides with the closest feed is $\left(\frac{k!(N-k)!}{N!}\right)$, which vanishes as N grows large, we have that:

Corollary 4.2. *As N grows large, the probability that the random algorithm outperforms the closest algorithm in terms of learning converges to 1:*

$$\lim_{N\to\infty} \mathbb{P}\left[Var[\theta | \{\theta_j\}_{j\in\mathcal{S}_N^\mathcal{R}}] < Var[\theta | \{\theta_j\}_{j\in\mathcal{S}_N^\mathcal{C}}] \right] = 1.$$

When society grows large, within-the-platform utility is maximized at the expense of learning, which is minimized. The next result shows for which values of λ and β the user is better-off under $\mathcal{C}$ than under $\mathcal{R}$.

Proposition 4.3. *If $N \to \infty$, total expected utility under $\mathcal{C}$ is greater than total expected utility under $\mathcal{R}$ if and only if*

$$\lambda > \frac{1}{2-\beta}.$$

Proof. When $N \to \infty$, total expected utility under $\mathcal{C}$ is $-(1-\lambda)\sigma^2$, and total expected utility under $\mathcal{R}$ is $-\lambda(1-\beta)\frac{k-1}{k}\sigma^2 - (1-\lambda)\frac{\sigma^2}{k}$. □

Remember that the weight of conformity in the utility function is $\lambda(1-\beta)$. When λ is large and β is small, conformity is highly important for the user and she is willing to sacrifice learning. In this scenario, $\mathcal{C}$ is preferred. Otherwise, the user prefers the more balanced combination given by $\mathcal{R}$.

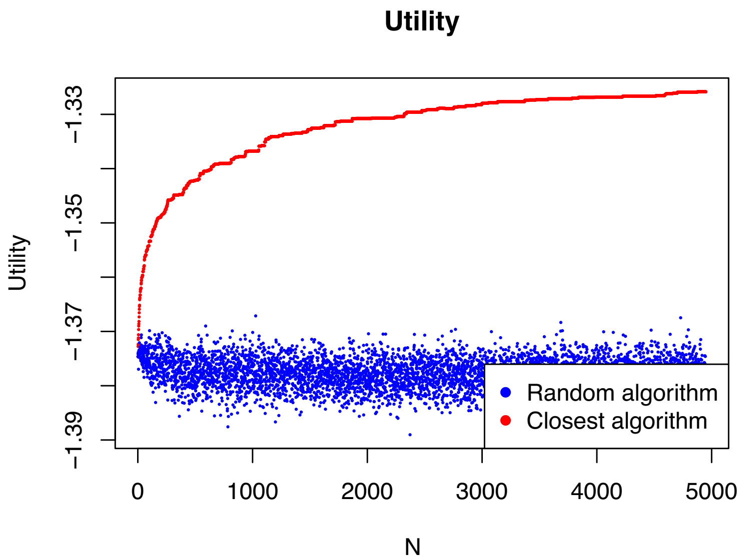

Figure 2: User's total utility as the population grows.

In Figure 2, we observe that total utility under $\mathcal{C}$ increases concavely with N for the same parameters as in Figure 1. This phenomenon is primarily driven by network effects, which occur when a platform's value increases as more people use it. The increase in total utility occurs because the growth of N widens the pool from which the platform can select users to construct the personalized feed. As a result, the platform can choose more similar neighbors, thereby increasing utility by improving the conformity term. While action utility may be negatively impacted, this effect is not sufficient to offset the benefits of increased conformity resulting from greater similarity. In contrast, the random algorithm is not influenced by the size of the network, so it does not exhibit network effects. Network effects are key to understand platform competition, as we will see in Section 6.

Moving forward, since the closest algorithm performs poorly in terms of learning and there is no learning at all in the limit, we shift our focus to potential measures to enhance the engagement-maximizing algorithm. A regulator could enforce the inclusion of neighbors in the feed based on different selection criteria. We examine the effects on user utility when an extra user is added to a feed that has already been selected by the closest algorithm and the population grows large. We consider three different possibilities: (i) selecting a user following the closest algorithm (i.e., $\mathcal{C}$ with $k+1$ feed length), (ii) adding a user uncorrelated with any other user in the feed, and (iii) introducing a user who is maximally negatively correlated with every other user in the feed, the breaking echo chambers algorithm.

We find that if conformity is not disproportionaly weighted compared to learning (i.e., if λ is not too large), the breaking echo chambers algorithm outperforms $\mathcal{C}$. It leads to maximal learning with a small loss in conformity. While this algorithm may be implementable with regulatory enforcement, its long-term viability in the real world remains questionable. Although opposite content can be enforced, whether through sponsored public service announcements with regular frequency or by directly incorporating dissimilar views into the feed, any user may simply choose to disregard artificially added content and, perhaps naively, opt not to engage with it.

Now, for the sake of technical simplicity, let us assume that when population size is sufficiently large, the closest algorithm can select a pool of k users with arbitrarily high correlations satisfying that these correlations are identical among all users within the pool. Formally, for every $\varepsilon > 0$, there is a $\tilde{N}$ such that if $N > \tilde{N}$, then there are $k-1$ users in the network such that $\rho_{jl} = 1 - \varepsilon$ for all j, l in the pool and user i.

Proposition 4.4. *Assuming the closest algorithm can select a pool of k users with arbitrarily high correlations verifying that these correlations are identical among all users within the pool as $N \to \infty$, the effects on user's expected utility of adding an extra user are as follows:*

(i) *If the user is chosen according to the closest algorithm, both conformity and the posterior variance of θ do not vary and are kept to 0 and σ^2 respectively.*

(ii) *If the chosen user is uncorrelated with everyone else appearing in the feed, conformity becomes $\frac{-\sigma^2}{k+1}$, and the posterior variance of θ decreases to $\frac{\sigma^2}{2}$.*

(iii) *Breaking echo chambers algorithm: if the chosen user is maximally negatively correlated to everyone else appearing in the feed, conformity becomes $\frac{-2\sigma^2}{k+1}$ and the posterior variance decreases to 0.*

Finally, the breaking echo chambers algorithm is always better than $\mathcal{R}$. It outperforms (ii) and $\mathcal{C}$ if and only if

$$\lambda \leq \frac{k+1}{2(1-\beta)+k+1}.$$

Proof. See Appendix C. □

Even though the breaking echo chambers algorithm could improve the overall performance of both $\mathcal{C}$ and $\mathcal{R}$ while providing perfect learning, it is still weakly worse than the utilitarian optimal algorithm. We define the user optimal algorithm ($\mathcal{U}$) as the one that maximizes each user's expected utility, representing the best service a platform can offer to each user. For each user i, this algorithm induces the following feed (remember that users always report truthfully, as stated in Corollary 3.4):

$$\mathcal{S}_i^{\mathcal{U}} = \underset{\mathcal{S}_i \subseteq N}{\operatorname{argmax}} \left\{ -\lambda(1-\beta)\frac{1}{k}\sum_{j\in\mathcal{S}_i} \mathbb{E}_p[(\theta_i - \theta_j)^2] - (1-\lambda)\,\mathbb{E}_p[(a_i - \theta)^2] \right\}.$$

Given that $a_i = \mathbb{E}_i[\theta | \{\theta_j\}_{j \in \mathcal{S}_i^{\mathcal{U}}}]$ and the only feed that affects user i's action utility is $\mathcal{S}_i^{\mathcal{U}}$, we can establish the following corollary to Lemma 3.1:

Corollary 4.5. *The user optimal algorithm is the utilitarian optimal algorithm:*

$$\underset{\mathcal{U}=(\mathcal{S}_1^{\mathcal{U}},\dots,\mathcal{S}_n^{\mathcal{U}})}{\operatorname{argmax}} \left\{ \sum_{i=1}^{n} \mathbb{E}_p[U_i] \right\} = \left(\underset{\mathcal{S}_1^{\mathcal{U}} \subseteq N}{\operatorname{argmax}} \{\mathbb{E}_p[U_1]\}, \dots, \underset{\mathcal{S}_n^{\mathcal{U}} \subseteq N}{\operatorname{argmax}} \{\mathbb{E}[U_n]\} \right).$$

For an example of the user optimal algorithm, please refer to Appendix A. Since it is socially preferred, it is worth discussing its potential implementation. Moreover, in this section we have shown that the closest algorithm is not convenient from a learning perspective. Section 6 will delve into how competition may encourage platforms implementing $\mathcal{U}$. However, before exploring that, the next section is dedicated to addressing the model's robustness: first, examining how the model responds when an entry stage is introduced; second, exploring modifications to the assumptions concerning the signal structure; and third, studying the game for naïve users.

5 Extensions

5.1 Outside option and entry problem

In the main model, we assume users join the monopolistic platform at the outset. Hence, it does not need to worry about attracting them and platform's profits are the sum of all users' engagement. Bursztyn et al. (2023) show empirically that the "Fear of Missing Out" (FOMO) induce users to join social platforms even though they would prefer them not to exist. In our model, this translates as follows. The outside option, which consists of being alone outside the platform, yields expected action utility $-(1-\lambda)\sigma^2$ (the user only learns her signal) and within-the-platform utility $-\lambda\tilde{v} < 0$ (because of FOMO). If $v+\tilde{v} \geq (1-\beta)\frac{2k}{k-1}\sigma^2$, then every user would join the platform, irrespective of the algorithm (a payoff of $\lambda(v-(1-\beta)\frac{2k}{k-1}\sigma^2)$ is the worst possible case, so if the outside option is even worse, no user would opt for it). A high level of FOMO (i.e., a large $\tilde{v}$) would result in all users joining a monopolistic platform. Moreover, if $N \to \infty$ and $\mathcal{C}$ is implemented, every user would join provided that $v + \tilde{v} \geq 0$.

Now, let us consider a scenario where there is no FOMO when users decide whether to join the platform or stay out. In this case, users know of the algorithm the platform intends to implement, but they remain uncertain about the similarity matrix $\boldsymbol{\Sigma}$; a user lacks information about the similarities with her peers in a platform until she becomes aware of their identities. The outside option yields utility $-(1-\lambda)\sigma^2$, representing the payoff derived from being alone. As the platform's algorithm is public, users can compute the expected payoff from entering under that algorithm and compare it to the outside option. The user enters if and only if the following inequality holds:[24]

$$\mathbb{E}\left[\lambda\left(v-(1-\beta)\sum_{j\in\mathcal{S}_i}\frac{(\theta_i-\theta_j)^2}{k}\right)-(1-\lambda)(a_i-\theta)^2\right]>-(1-\lambda)\sigma^2.$$

Since the matrix $\mathbf{\Sigma}$ remains unknown to users prior to entry, they must estimate it.[25] Note that under the closest algorithm and assuming a uniform distribution for peers covariances (i.e., if $\tilde{\sigma}_{ij}$ is user i's estimated covariance between her and user j, then $\tilde{\sigma}_{ij}\sim\mathcal{U}[-\sigma^2,\sigma^2]$), this equation becomes:

$$\lambda v-\lambda(1-\beta)\frac{2\sigma^2(k+1)}{N}-(1-\lambda)\mathbb{E}[(a_i-\theta)^2]>-(1-\lambda)\sigma^2.$$

Estimating action utility for a finite N prior entry is unfeasible, so we use Proposition 4.1 to set a bound: $\mathbb{E}[(a_i-\theta)^2]\leq\sigma^2$. This simplifies the inequality to a sufficient condition for user entry. When

$$v-(1-\beta)\frac{2\sigma^2(k+1)}{N}>0,$$

the user chooses to enter. Hence, for any given v, there exists a threshold $\bar{N}$ such that if $N>\bar{N}$, all users opt to join. In a monopolistic scenario with a large population (which is the focus of our analysis), all users join, making the closest algorithm the one that maximizes the platform's profits. Notably, other scenarios exist, such as instances where the population is not as large, prompting the platform to implement an algorithm that considers learning to attract users. Remember that when users decide which platform to join, they take into account total utility. Hence, when the platform cares about attracting users, learning must be taken into account. This is the case when multiple platforms compete in the market: they need to be attractive enough in terms of total expected utility to be joined and then they also need to maximize engagement. This situation is explored in Section 6.

5.2 Other signal structures

So far, we have assumed equal variances across users with heterogeneous correlations. This assumption appears most plausible as differences in opinion or ideology may exert a stronger influence than disparities in learning accuracy. Nevertheless, we find it worthwhile to investigate alternative signal structures within our model. Should we relax the assumption of equal variances ($\sigma_i^2\neq\sigma_j^2$), the platform would not only consider similarity

24 Here, we abstract from considering multiple equilibria scenarios and solely focus on the situation where all users coordinate on entering. This outcome would also arise if the equilibrium concept were strong Nash equilibrium or coalition-proof Nash equilibrium.

25 The task of estimating an entire covariance matrix surpasses our capabilities. Hence, we allow users to estimate the covariances their peers have with them, i.e., a row of the matrix. For simplicity, we assume a uniform distribution for these covariances; under the closest algorithm, a user will be paired with those possessing the k highest covariances: $\{\frac{N-k}{N}\sigma^2,\frac{N-k+1}{N}\sigma^2,...,\frac{N-1}{N}\sigma^2,\sigma^2\}$.

but also accuracy when selecting users:

$$\begin{aligned}\mathcal{S}_i^* &= \underset{\mathcal{S}_i}{\text{argmax}} \left\{ -\sum_{j\in\mathcal{S}_i} \mathbb{E}_p\left[(\theta_i-\theta_j)^2\right]\right\} = \underset{\mathcal{S}_i}{\text{argmax}}\left\{-\sum_{j\in\mathcal{S}_i}(\sigma_i^2+\sigma_j^2-2\sigma_{ij})\right\}\\ &= \underset{\mathcal{S}_i}{\text{argmax}}\left\{-\sum_{j\in\mathcal{S}_i}(\sigma_j^2-2\sigma_{ij})\right\} \quad \forall\, i.\end{aligned}$$

Note that similarity carries twice the weight of accuracy. If it was the user who chose a feed that maximizes her within-the-platform utility, such a feed (let us denote it by $\tilde{\mathcal{S}}_i$) would be given by:

$$\tilde{\mathcal{S}}_i = \underset{\mathcal{S}_i}{\text{argmax}}\left\{-\sum_{j\in\mathcal{S}_i}\mathbb{E}_i\left[(\theta_i-\theta_j)^2|\theta_i\right]\right\} = \underset{\mathcal{S}_i}{\text{argmax}}\left\{-\sum_{j\in\mathcal{S}_i}(\sigma_j^2-\sigma_{ij})\right\}.$$

Hence, the platform has a stronger preference for similarity than the user, who weights similarity and accuracy equally. This implies that the echo chambers result is slightly more subtle than a trivial consequence of conformity. Once we introduce heterogeneous variances, we observe that the platform prefers to build an echo chamber, while the user might prefer to receive a more balanced feed.

Finally, we analyze the simpler scenario where variances can differ, but users' signals are uncorrelated. In this case, the platform implements

$$\begin{aligned}\mathcal{S}_i^* &= \underset{\mathcal{S}_i}{\text{argmax}} \left\{ -\sum_{j\in\mathcal{S}_i} \mathbb{E}_p\left[(\theta_i-\theta_j)^2\right]\right\} = \underset{\mathcal{S}_i}{\text{argmax}}\left\{-\sum_{j\in\mathcal{S}_i}(\sigma_i^2+\sigma_j^2-2\sigma_{ij})\right\}\\ &= \underset{\mathcal{S}_i}{\text{argmax}}\left\{-\sum_{j\in\mathcal{S}_i}\sigma_j^2\right\} \quad \forall\, i.\end{aligned}$$

In this scenario, the platform aims to match users with those who exhibit the highest learning accuracy. This stands in stark contrast to the primary model, as the platform's interests are now fully aligned with individual learning.

5.3 Naïve users

Naïve users are mechanical individuals who share their beliefs and update them using the DeGroot rule (DeGroot, 1974). Therefore, they mechanically report their private signal (so they post $m_i^* = \theta_i$) and then update their beliefs taking the average of the observed messages (so their action is $a_i = \frac{1}{k}\sum_{j\in\mathcal{S}_i} m_j = \frac{1}{k}\sum_{j\in\mathcal{S}_i}\theta_j$). Following the reasoning in Proposition 3.3, an engagement-maximizer platform will implement the closest algorithm $\mathcal{C}$ in equilibrium. However, when applied to naïve users, the closest algorithm always harms learning.

Proposition 5.1. *When users are naïve, the posterior variance $Var\left[\theta|\frac{1}{k}\sum_{j\in\mathcal{S}_i^{\mathcal{C}}}\theta_j\right]$ given the closest algorithm is larger than the posterior variance induced by any other algorithm.*

Proof. Before showing that $\text{Var}\left[\theta|\frac{1}{k}\sum_{j\in\mathcal{S}_i^\mathcal{C}}\theta_j\right] \geq \text{Var}\left[\theta|\frac{1}{k}\sum_{j\in\mathcal{S}_i^\mathcal{F}}\theta_j\right]$ for any algorithm $\mathcal{F}$, we need to characterize the posterior variance of θ when just the average of the observed messages is learnt.

The posterior distribution of the average message conditional on θ follows $\frac{1}{k}\sum_{j=1}^k \theta_j|\theta \sim \mathcal{N}\left(\frac{1}{k}\sum_{j=1}^k \theta_j, \frac{\sigma^2}{k} + \frac{1}{k}\sum_{j\in\mathcal{S}_i}\sigma_{ij}\right)$. As we have assumed improper priors for θ,

$$\theta|\frac{1}{k}\sum_{j=1}^{k}\theta_j \sim \mathcal{N}\left(\frac{1}{k}\sum_{j=1}^{k}\theta_j, \frac{\sigma^2}{k} + \frac{1}{k}\sum_{j\in\mathcal{S}_i}\sigma_{ij}\right).$$

Hence, $\text{Var}[\theta|\frac{1}{k}\sum_{j\in\mathcal{S}_i}\theta_j] = \frac{\sigma^2}{k} + \frac{1}{k}\sum_{j\in\mathcal{S}_i}\sigma_{ij}$. The closest algorithm features the largest posterior variance among any algorithm, because by definition it chooses the neighbors whose signals feature the largest covariances with that of i. □

Addressing learning among naïve users is crucial, not only because the literature on platforms and social networks often model users following a boundedly-rational approach, but also due to potential concerns among policymakers regarding how platforms treat them. Here we show that $\mathcal{C}$ is the worst possible algorithm for naïve users in terms of learning. This finding, combined with the results presented in Section 4 for Bayesian users, underscores the importance of considering the implementation of the user optimal algorithm. Therefore, we dedicate the next section to analyzing the challenges that may arise during the implementation of this algorithm through platform competition and exploring the role of interoperability as a potential regulatory solution.

6 Discussion: Interoperability

In Section 4, we introduced the breaking echo chambers algorithm, an improvement over the closest algorithm that provides perfect learning. However, the success of this algorithm might be unrealistic, and it is still weakly worse than the user optimal algorithm in terms of consumer welfare. We propose platform competition as the setting to incentivize the implementation of the user optimal algorithm. In a perfectly competitive market, platforms must implement it, as they would otherwise lose their user base due to an à la Bertrand argument. However, despite the low cost of entry in digital markets, they typically feature high entry barriers created by network effects, that protect incumbents and deter competition. We propose interoperability as a measure to shift platform-level network effects to market-level network effects, reducing entry barriers and promoting competition.

To provide some intuition, alongside this section we consider a large incumbent, which corresponds to a *quasi* monopolist like Twitter, and small entrants, like for example Mastodon.[26] Entry is costless, and platforms attract consumers based on the expected utility they offer. Theoretically, if users prefer the user optimal algorithm, any entrant platform implementing it could attract the incumbent's user base if the incumbent does not modify its algorithm choice (i.e., if Mastodon's algorithm were superior to Twitter's, a significant migration would likely occur).[27] However, two intrinsic aspects of social media platforms complicate this situation and discourage competition: network effects and incumbent data advantage.[28]

Network effects indicate that the expected utility of being on a platform (prior to entry) increases with the size of the network. Consequently, larger platforms find it easier to attract and retain users (they find it difficult to leave Twitter and join Mastodon if almost no friend has done so, even though Mastodon's service might be better). We show below (Proposition 6.1) that both the closest and the user optimal algorithm feature network effects. Hence, total expected utility depends not only on the implemented algorithm but also on size: platforms compete alongside two dimensions, algorithm choice and size. The incumbent, with network effects on its side, holds market power, creating high entry barriers for potential competitors. The entrant's success in the market no longer depends on its technology (here, the algorithm it implements), but on breaking through the incumbent's network advantage.

Now, let us show formally that both $\mathcal{C}$ and $\mathcal{U}$ feature network effects. We assume that if a user has not joined a platform yet, she expects the posterior variance of θ under the closest algorithm to be weakly increasing with network size.[29] In other words: as more users join the platform, her expected posterior variance will rise, as she expects herself to be matched with more similar friends and hence to learn less about θ. Under this assumption and for λ not too small, both the user optimal and the closest algorithms feature network effects.

Proposition 6.1. *Assuming user i's expected posterior variance to be weakly increasing in N, then*

$$\lambda \geq \frac{k-1}{k(1+(1+\beta)2(k+1))}$$

ensures that the closest algorithm exhibits network effects. The user optimal algorithm features network effects for all $\lambda \in (0,1)$.

Proof. For simplicity, we denote U_i as U. Let $g(N)$ represent the expected posterior variance of user i when the network size is N, prior to entry. If $N = k$, the neighbor pool

26 As of July 2023, Twitter had 237 million active users, while Mastodon accounted for 1.5 million users.

27 Remember that when users make the decision to join a platform, whether they are comparing its benefits to other platforms or the outside option, they take into account their overall utility.

28 As mentioned in the literature review, Biglaiser et al. (2022) provides a microfoundation for incumbent advantage.

29 We have seen that learning non-monotonically diminishes when N is expanded. We assume that, in expectation, for a user who has not joined the platform yet and hence does not know $\mathbf{\Sigma}$, it behaves monotonically.

is i.i.d., making $g(k) = \frac{\sigma^2}{k}$. By Proposition 4.1, we also know that $\lim_{N\to\infty} g(N) = \sigma^2$. The weakly increasing nature of $g(N)$ implies that $g(N+1) - g(N) < \sigma^2 - \frac{\sigma^2}{k} < \frac{k-1}{k}\sigma^2$ for all N.

The closest algorithm features network effects if the expected utility of joining a platform implementing this algorithm increases with population size, i.e., if $\mathbb{E}[U(N+1)] - \mathbb{E}[U(N)] > 0$ for all N. As derived in Subsection 5.1, the expected utility prior to entry under $\mathcal{C}$ is $\mathbb{E}[U(N+1)] = v - \lambda(1-\beta)\frac{2\sigma^2(k+1)}{N+1} - (1-\lambda)g(N+1)$. Thus,

$$\mathbb{E}[U(N+1)] - \mathbb{E}[U(N)] = \lambda(1-\beta)\frac{2\sigma^2(1+k)}{N(N+1)} - (1-\lambda)(g(N+1) - g(N)).$$

Using that $g(N+1) - g(N) < \frac{k-1}{k}\sigma^2$, we obtain that if

$$\lambda \geq \frac{k-1}{k(1+(1-\beta)2(k+1))},$$

then $\mathcal{C}$ features network effects. (Note that there always exist such λ, as the right-hand-side of the inequality is less than 1.)

Finally, $\mathcal{U}$ also exhibits network effects for all $\lambda \in (0,1)$: larger neighborhood sizes result in a larger algorithm selection pool, leading to higher expected user utility since the platform aims to maximize individual welfare. $\square$

Note that if we set $\beta = 0.5$, the bound for λ is below 0.1 for $k \geq 10$ and decreases with k. For scenarios where sincerity is highly valued ($\beta = 0.9$), the threshold for λ is below 0.2 for $k \geq 20$, and even when sincerity is of utmost importance ($\beta = 0.1$), the threshold for λ remains below 0.1 for $k \geq 5$. Therefore, in most cases, the closest algorithm exhibits network effects.

However, network effects are not the sole deterrent to competition in our model. In our model, we find incumbent data advantage. It refers to the fact that when a user decides to leave the incumbent and join a competing platform, the new platform has no information about past interactions, i.e., it does not know her values in the similarity matrix $\mathbf{\Sigma}$. Initially, the new platform cannot personalize the user's feed because the data does not migrate with her (if a user migrates from Twitter to Mastodon, the data does not migrate with her, and Mastodon needs a few periods to adjust its personalized feed). The presence of network effects and incumbent data advantage discourages competition, leading to market concentration and "winner-takes-all" dynamics that result in a monopoly. Even if a potential entrant adopts the user optimal algorithm, it would not suffice to break through the incumbent's advantage if its size is large enough. Moreover, when under pressure, the incumbent can slightly modify its algorithm (or claim to do so). For example, major monopolists like Facebook or Twitter have heavily campaigned to demonstrate algorithmic changes or an increase in their efforts to combat fake news in response to public outcry, with the aim of deterring potential competitors, as evidenced in Horwitz et al. (2021).

We advocate for interoperability as a tool to shift network effects from the platform level to the market level, making them available to competitors and not just an incumbent advantage. Interoperability refers to complete interaction between different platforms: two platforms become interoperable when their users can interact with each other. Hence, an entrant platform could use the whole population to provide each user with the personalized feed it desires: messages posted in the incumbent platform could be displayed in the entrant platform and viceversa. As a consequence, the only dimension platforms compete within is on the algorithm choice. This forces both platforms to play user optimal algorithm in equilibrium, in the spirit of à la Bertrand competition. Drawing on the example of Twitter and Mastodon, interoperability would mean that a user on Mastodon can be friends with users from both platforms. Mastodon would have access to the public content of all their friends who are on Twitter and use the information to generate a personalized feed that includes messages from both platforms.[30] The same principle applies in reverse: Twitter can display a user's friends' messages, regardless of the platform they are registered on. We could even strenghten this notion of interoperability adding data portability: in the event of migration, data would travel with the user. This would eliminate incumbent data advantage.

Interoperability has been successfully implemented in industries such as cell phones and email. Naturally, the level playing field created by interoperability disadvantages platforms with significant network effects, as consumer adoption decisions are no longer influenced by size. Conversely, smaller platforms would fear losing if they competed *for the market* and thus prefer interoperability to be able to compete *in the market* (Belleflamme and Peitz, 2020).[31] Aiming "at preventing gatekeepers from imposing unfair conditions on business and end users and at ensuring the openess of important digital services",[32] the European Comission has introduced interoperability as a regulatory measure in the European Union. The following subsection delves more into it, also discussing the challenges that might arise, before Section 7 sets the concluding remarks of the paper.

30 In this paper we have considered a complete network. However, results apply for general networks, and they are more convenient in this section: under interoperability, each user could keep their neighborhood irrespective of which platform each of her friends is member of. One would care about which friends are in the market, but not in which platform, identically as it happens with the mobile phone industry. We care whether a friend has a mobile phone, but not which is her company, and we keep our neighborhood independently of which companies we are registered at.

31 Still, becoming interoperable is always a decision for the small platform to make. Regulators just require large platforms to make it possible.

32 This quote is extracted from *Questions and Answers: Digital Markets Act: Ensuring fair and open digital markets*, available at https://ec.europa.eu/commission/presscorner/api/files/document/print/en/qanda_20_2349/QANDA_20_2349_EN.pdf.

6.1 Interoperability in the EU: the Digital Markets Act

In July 2022, the European Council passed the Digital Markets Act (DMA), a significant regulatory measure. Under this act, "gatekeeper" platforms and services are mandated to provide interoperability for chats with users on other services. Gatekeeper platforms, defined as those entities exerting substantial market influence and possessing or expected to possess a firmly established and enduring market position, are designated by the European Commission. The DMA's primary objective is to extend interoperability to various communication tools, with a particular focus on messaging applications. For instance, under the DMA's provisions, a WhatsApp user should be able to send a message to a Telegram user, and an iPhone owner should be able to send an online message to an Android user through the iMessage App. However, these means of communication have traditionally required a shared platform, resulting in the concentration of market power.

Under the DMA, gatekeepers' communication services, including messaging applications, are now required to provide the necessary interfaces for horizontal interoperability. Compliance entails ensuring interoperability in fundamental functionalities, such as end-to-end messaging, voice and video calls, and the sharing of images, voice messages, videos, and files. The compliance timeline varies, ranging from four years for voice and video calls to immediate action for end-to-end messaging between two individuals. WhatsApp has already initiated efforts toward achieving interoperability with other applications.[33] However, this endeavor raises significant concerns, particularly for services that promise end-to-end encryption. Cryptographers widely agree that maintaining encryption between different apps may prove challenging, if not impossible, potentially posing substantial implications for user privacy.

7 Conclusion

We have built a model of communication and learning through personalized feed. An engagement-maximizing monopolist platform lacks incentives to consider social learning; therefore, it chooses to display messages from similar neighbors. Our findings reveal that not only is this outcome suboptimal in terms of social welfare, but asymptotically, learning is persistently hindered, to the point that it vanishes. This observation encourages exploring strategies to motivate platforms to adopt the utilitarian optimal algorithm (the user optimal algorithm), given the challenges in implementing measures that would improve the closest algorithm (the breaking echo chambers algorithm). While fostering platform competition seems a viable approach, the presence of entry barriers created by network effects obstruct its effectiveness. Nevertheless, if interoperability is enforced, this obstacle would vanish.

33 See, for example, this recent article: https://www.theverge.com/2023/9/10/23866912/whatsapp-cross-platform-messaging-eu-dma-meta.

Indeed, interoperability might offer broader benefits than those discussed in this paper. For example, Farronato et al. (2023) show that when users have heterogeneous preferences, a single platform might not be as effective as multiple platforms: network effects and platform differentiation offset each other when market tips. In principle, interoperability might solve this issue: network effects would happen at the market level (so they would be maximized) while at the same time there would still be platform differentiation. Addressing the effects of interoperability in a dynamic setting of competing platforms where heterogeneous users can multi-home is a natural continuation for this paper. Specifically, we aim to address two key questions: firstly, whether the necessary standards for interoperability could restrain innovation, and secondly, whether super-large platforms can maintain their dominance over time due to factors beyond algorithm competition.

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 and Asuman Ozdaglar, "Opinion dynamics and learning in social networks," *Dynamic Games and Applications*, 2011, *1* (1), 3–49.

_ , _ , 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*, 2021.

Agarwal, Saharsh, Uttara M Ananthakrishnan, and Catherine E Tucker, "Deplatforming and the Control of Misinformation: Evidence from Parler," working paper, available at *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4232871*, 2022.

Allcott, Hunt and Matthew Gentzkow, "Social media and fake news in the 2016 election," *Journal of Economic Perspectives*, 2017, *31* (2), 211–36.

Bandy, Jack and Nicholas Diakopoulos, "Curating quality? How Twitter's timeline algorithm treats different types of news," *Social Media+ Society*, 2021, *7* (3), 20563051211041648.

Belleflamme, Paul and Martin Peitz, "The competitive impacts of exclusivity and price transparency in markets with digital platforms," *Concurrences*, 2020, (N° 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.

Benzell, Seth and Avinash Collis, "Regulating Digital Platform Monopolies: The Case of Facebook," working paper, available at *https://www.aeaweb.org/conference/2023/program/paper/SNNd74ni*, 2022.

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.

Biglaiser, Gary, Jacques Crémer, and André Veiga, "Should I stay or should I go? Migrating away from an incumbent platform," *The RAND Journal of Economics*, 2022, *53* (3), 453–483.

Bonatti, Alessandro and Gonzalo Cisternas, "Consumer scores and price discrimination," *The Review of Economic Studies*, 2020, *87* (2), 750–791.

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.

Chitra, Uthsav and Christopher Musco, "Analyzing the impact of filter bubbles on social network polarization," in "Proceedings of the 13th International Conference on Web Search and Data Mining" 2020, pp. 115–123.

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

DeGroot, Morris H, "Reaching a consensus," *Journal of the American Statistical Association*, 1974, *69* (345), 118–121.

Demange, Gabrielle, "Simple visibility design in network games," working paper, available at https://hal.science/hal-03925344v1/document, 2023.

DeMarzo, Peter M, Dimitri Vayanos, and Jeffrey Zwiebel, "Persuasion bias, social influence, and unidimensional opinions," *Quarterly Journal of Economics*, 2003, *118* (3), 909–968.

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

Evens, Tom, Karen Donders, and Adelaida Afilipoaie, "Platform policies in the European Union: Competition and public interest in media markets," *Journal of Digital Media & Policy*, 2020, *11* (3), 283–300.

Faelens, Lien, Kristof Hoorelbeke, Ruben Cambier, Jill van Put, Eowyn Van de Putte, Rudi De Raedt, and Ernst HW Koster, "The relationship between Instagram use and indicators of mental health: A systematic review," *Computers in Human Behavior Reports*, 2021, *4*, 100121.

Farronato, Chiara, Jessica Fong, and Andrey Fradkin, "Dog eat dog: Balancing network effects and differentiation in a digital platform merger," *Management Science*, 2023.

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," *arXiv preprint arXiv:2102.13309*, 2021.

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.

Gottfried, Jeffrey and Elisa Shearer, "News use across social media platforms 2016," 2019.

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

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.

Habib, Hussam, Maaz Bin Musa, Fareed Zaffar, and Rishab Nithyanand, "To act or react: Investigating proactive strategies for online community moderation," *arXiv preprint arXiv:1906.11932*, 2019.

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

Horwitz, Jeff et al., "The Facebook Files," *The Wall St Journal. Available online at: https://www.wsj.com/articles/the-facebook-files-11631713039(accessedNovember01,2021)*, 2021.

Hu, Lin, Anqi Li, and Xu Tan, "A rational inattention theory of echo chamber," working paper, available at https://arxiv.org/pdf/2104.10657.pdf, 2021.

Hwang, Elina H and Stephanie Lee, "A nudge to credible information as a countermeasure to misinformation: Evidence from twitter," working paper, available at https://ssrn.com/abstract=3928343, 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).

Jadbabaie, Ali, Pooya Molavi, Alvaro Sandroni, and Alireza Tahbaz-Salehi, "Non-Bayesian social learning," *Games and Economic Behavior*, 2012, *76* (1), 210–225.

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

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.

Molavi, Pooya, Alireza Tahbaz-Salehi, and Ali Jadbabaie, "A theory of non-Bayesian social learning," *Econometrica*, 2018, *86* (2), 445–490.

Morton, Fiona Scott, Pascal Bouvier, Ariel Ezrachi, Bruno Jullien, Roberta Katz, Gene Kimmelman, A Douglas Melamed, and Jamie Morgenstern, "Market structure and antitrust subcommittee report, George J. Stigler Center for the Study of the Economy and the State," available at https://research.chicagobooth.edu/-/media/research/stigler/pdfs/market-structure-report.pdf, 2019.

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.

Mostagir, Mohamed and James Siderius, "Learning in a post-truth world," *Management Science*, 2022, *68* (4), 2860–2868.

_ and _ , "When do misinformation policies (not) work?," *Available online at: https://siderius.lids.mit.edu/how-do-we-stop-misinformation/ (accessed January 26, 2023)*, 2023.

_ and _ , "When should platforms break echo chambers?," *Available online at: https://siderius.lids.mit.edu/wp-content/uploads/sites/36/2023/01/quarantine-v6.pdf (accessed January 26, 2023)*, 2023.

Mudambi, Maya and Siva Viswanathan, "Prominence Reduction versus Banning: An Empirical Investigation of Content Moderation Strategies in Online Platforms," 2022.

Mueller-Frank, Manuel and Claudia Neri, "A general analysis of boundedly rational learning in social networks," *Theoretical Economics*, 2021, *16* (1), 317–357.

_ **, 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.

Pennycook, Gordon and David G Rand, "Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning," *Cognition*, 2019, *188*, 39–50.

_ **and** _ , "The psychology of fake news," *Trends in Cognitive Sciences*, 2021, *25* (5), 388–402.

Popiel, Pawel, "Addressing platform power: The politics of competition policy," *Journal of Digital Media & Policy*, 2020, *11* (3), 341–360.

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.

Sagioglou, Christina and Tobias Greitemeyer, "Facebook's emotional consequences: Why Facebook causes a decrease in mood and why people still use it," *Computers in Human Behavior*, 2014, *35*, 359–363.

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.

Solon, Olivia, "Facebook's failure: Did fake news and polarized politics get Trump elected?," *The Guardian,* available at *https://www.theguardian.com/technology/2016/nov/10/facebook-fake-news-election-conspiracy-theories*, November 10, 2016.

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

Törnberg, Petter, "Echo chambers and viral misinformation: Modeling fake news as complex contagion," *PLoS one*, 2018, *13* (9).

A Example

Here we present the feeds that user 1, with a network (Figure 3; we only draw the links corresponding to user 1 for simplicity) composed by 20 users ($N = 20$), would receive under the closest algorithm (Figure 4), the random algorithm (Figure 5), and the user optimal algorithm (Figure 6). Note that we assume $k = 6$, $\lambda = 0.5$ and $\beta = 0.5$.

$$\boldsymbol{\Sigma} = \left(\begin{matrix}
1.00 & -0.20 & -0.15 & 0.24 & 0.20 & 0.05 & 0.14 & 0.01 & 0.13 & -0.12 \\
-0.20 & 1.00 & -0.00 & -0.12 & 0.21 & 0.08 & -0.13 & -0.07 & -0.07 & 0.13 \\
-0.15 & -0.00 & 1.00 & -0.38 & -0.20 & -0.06 & -0.17 & 0.02 & -0.09 & -0.24 \\
0.24 & -0.12 & -0.38 & 1.00 & -0.23 & -0.20 & 0.04 & 0.05 & 0.03 & 0.07 \\
0.20 & 0.21 & -0.20 & -0.23 & 1.00 & -0.00 & 0.11 & -0.09 & -0.09 & 0.04 \\
0.05 & 0.08 & -0.06 & -0.20 & -0.00 & 1.00 & 0.27 & -0.17 & 0.06 & 0.06 \\
0.14 & -0.13 & -0.17 & 0.04 & 0.11 & 0.27 & 1.00 & 0.23 & 0.21 & -0.02 \\
0.01 & -0.07 & 0.02 & 0.05 & -0.09 & -0.17 & 0.23 & 1.00 & 0.10 & 0.17 \\
0.13 & -0.07 & -0.09 & 0.03 & -0.09 & 0.06 & 0.21 & 0.10 & 1.00 & 0.02 \\
-0.12 & 0.13 & -0.24 & 0.07 & 0.04 & 0.06 & -0.02 & 0.17 & 0.02 & 1.00 \\
0.21 & 0.06 & 0.04 & 0.05 & 0.01 & 0.13 & -0.02 & 0.16 & -0.02 & 0.14 \\
0.17 & 0.05 & -0.29 & 0.06 & 0.39 & -0.05 & 0.14 & -0.22 & -0.14 & -0.00 \\
-0.14 & 0.24 & 0.23 & -0.15 & -0.07 & 0.28 & 0.20 & 0.08 & -0.01 & 0.08 \\
0.14 & -0.18 & 0.20 & 0.02 & -0.11 & -0.29 & -0.34 & -0.16 & -0.04 & -0.01 \\
0.01 & 0.03 & 0.22 & 0.02 & -0.23 & -0.02 & -0.39 & -0.33 & -0.11 & 0.15 \\
-0.16 & 0.25 & -0.24 & 0.09 & 0.06 & -0.04 & -0.13 & -0.16 & 0.18 & 0.08 \\
-0.26 & 0.21 & 0.15 & -0.16 & 0.04 & -0.04 & 0.03 & -0.01 & -0.09 & 0.11 \\
-0.12 & 0.10 & -0.06 & -0.23 & 0.13 & 0.09 & -0.07 & 0.20 & -0.13 & 0.30 \\
0.35 & -0.01 & 0.15 & -0.04 & 0.06 & -0.02 & -0.22 & -0.19 & -0.01 & -0.12 \\
-0.16 & 0.10 & -0.22 & -0.16 & 0.05 & -0.02 & -0.02 & -0.09 & 0.10 & 0.06
\end{matrix}\right.$$

$$\left.\begin{matrix}
0.21 & 0.17 & -0.14 & 0.14 & 0.01 & -0.16 & -0.26 & -0.12 & 0.35 & -0.16 \\
0.06 & 0.05 & 0.24 & -0.18 & 0.03 & 0.25 & 0.21 & 0.10 & -0.01 & 0.10 \\
0.04 & -0.29 & 0.23 & 0.20 & 0.22 & -0.24 & 0.15 & -0.06 & 0.15 & -0.22 \\
0.05 & 0.06 & -0.15 & 0.02 & 0.02 & 0.09 & -0.16 & -0.23 & -0.04 & -0.16 \\
0.01 & 0.39 & -0.07 & -0.11 & -0.23 & 0.06 & 0.04 & 0.13 & 0.06 & 0.05 \\
0.13 & -0.05 & 0.28 & -0.29 & -0.02 & -0.04 & -0.04 & 0.09 & -0.02 & -0.02 \\
-0.02 & 0.14 & 0.20 & -0.34 & -0.39 & -0.13 & 0.03 & -0.07 & -0.22 & -0.02 \\
0.16 & -0.22 & 0.08 & -0.16 & -0.33 & -0.16 & -0.01 & 0.20 & -0.19 & -0.09 \\
-0.02 & -0.14 & -0.01 & -0.04 & -0.11 & 0.18 & -0.09 & -0.13 & -0.01 & 0.10 \\
0.14 & -0.00 & 0.08 & -0.01 & 0.15 & 0.08 & 0.11 & 0.30 & -0.12 & 0.06 \\
1.00 & -0.22 & -0.04 & 0.10 & 0.13 & 0.19 & -0.22 & 0.05 & 0.07 & -0.20 \\
-0.22 & 1.00 & -0.13 & 0.05 & -0.08 & -0.03 & 0.14 & 0.02 & -0.01 & 0.13 \\
-0.04 & -0.13 & 1.00 & -0.29 & -0.00 & -0.23 & 0.14 & 0.06 & -0.13 & -0.15 \\
0.10 & 0.05 & -0.29 & 1.00 & 0.45 & 0.14 & -0.06 & -0.03 & 0.29 & 0.11 \\
0.13 & -0.08 & -0.00 & 0.45 & 1.00 & -0.02 & -0.03 & 0.12 & 0.32 & -0.02 \\
0.19 & -0.03 & -0.23 & 0.14 & -0.02 & 1.00 & 0.03 & -0.16 & -0.07 & 0.26 \\
-0.22 & 0.14 & 0.14 & -0.06 & -0.03 & 0.03 & 1.00 & 0.00 & -0.04 & 0.21 \\
0.05 & 0.02 & 0.06 & -0.03 & 0.12 & -0.16 & 0.00 & 1.00 & 0.08 & 0.08 \\
0.07 & -0.01 & -0.13 & 0.29 & 0.32 & -0.07 & -0.04 & 0.08 & 1.00 & -0.02 \\
-0.20 & 0.13 & -0.15 & 0.11 & -0.02 & 0.26 & 0.21 & 0.08 & -0.02 & 1.00
\end{matrix}\right)$$

The closest algorithm selects the most similar users to 1, and we have colored in salmon their covariances in the matrix $\mathbf{\Sigma}$ as well as their nodes in Figure 4. The user optimal algorithm selects the pool of neighbors that maximize user 1's expected utility (conditional on the platform's information). We have colored the background of their covariances with user 1 in yellow, as well as their nodes in Figure 6. For this specific realization of $\mathbf{\Sigma}$, $\mathcal{C}$ provides a posterior variance of 0.2061, while $\mathcal{U}$ induces a posterior variance of 0.1033, hence increasing learning. Regarding conformity, user 1 loses 0.6376 utils due to conformity under $\mathcal{C}$, while she loses 0.7177 utils under $\mathcal{U}$. In overall terms, the total utility is -0.2624 under $\mathcal{C}$ and -0.2311 under $\mathcal{U}$. Note that, on average, $\mathcal{R}$ penalizes conformity with 0.9825 utils and induces a posterior variance of 0.1667, providing an overall expected utility of -0.3289.

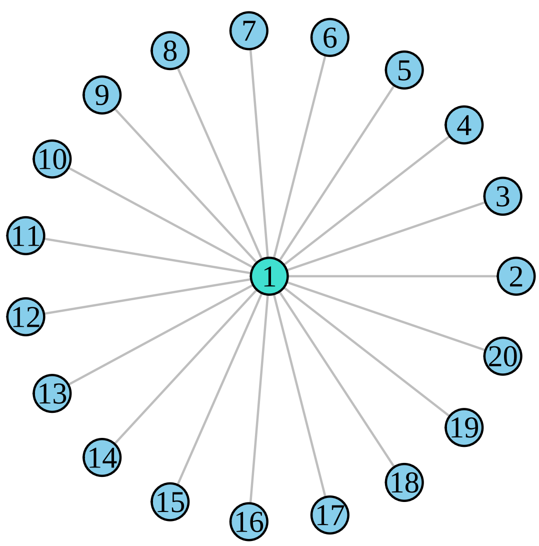

Figure 3: User i's network

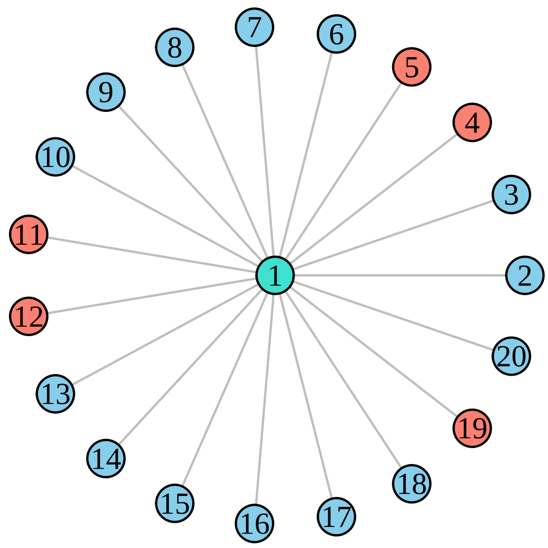

Figure 4: Closest feed

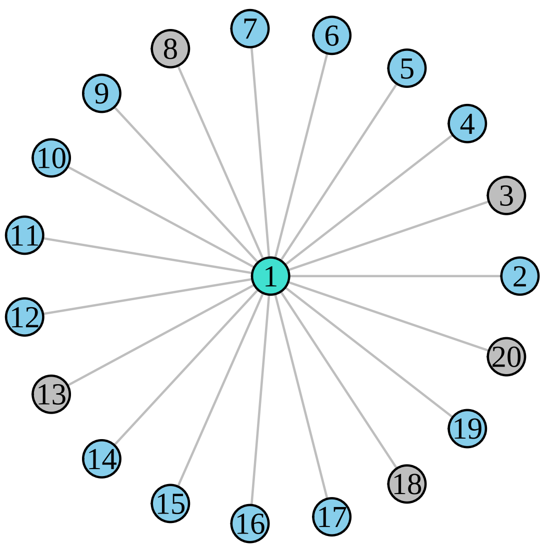

Figure 5: Random feed

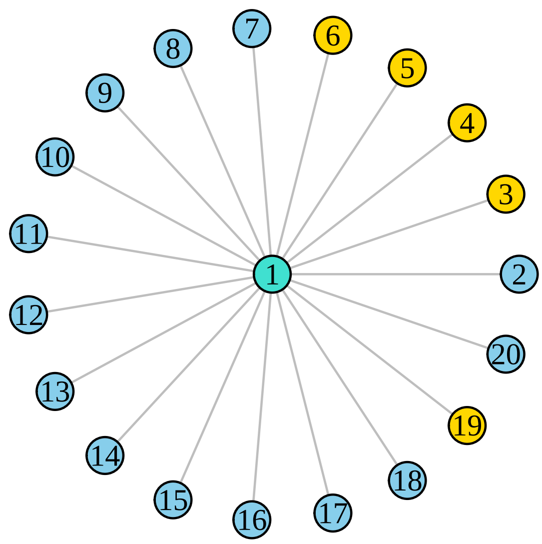

Figure 6: User optimal feed

In summary: $\mathcal{C}$ minimizes the loss coming from conformity, but this comes at the price of learning. $\mathcal{R}$ enhances learning, but, driven by conformity, overall utility is worse. Finally, $\mathcal{U}$ finds the pool of neighbors that maximize expected utility by balancing conformity and learning.

B Endogenous choice of k

Here, we allow a user who prioritizes the immediate rewards of interacting within the platform to determine her level of engagement. It remains optimal for the platform to implement the closest algorithm. Hence, fixing k and focusing on the platform maximizing within-the-platform utility as a proxy for engagement is equivalent to letting the user choose her engagement level in terms of platform's optimal algorithm decision.

Proposition B.1. *If a user who cares about within-the-platform utility could choose k endogenously, $\mathcal{C}$ still arises in equilibrium.*

Proof. Assume that k is endogenously chosen by user i. The reward for being in the platform is now $v(k)$, which we assume to be an increasing and concave function. Now, a feed $\mathcal{S}_i$ is a (personalized) order among all other users, so that for all $k \in \mathbb{N}$ there is some $\mathcal{S}_i(k) \subseteq N$ such that $|\mathcal{S}_i(k)| = k$ and $\mathcal{S}_i(k) \subseteq \mathcal{S}_i(k+1)$.

Given a feed $\mathcal{S}_i$, the user chooses the optimal number of messages to read following

$$
\begin{aligned}
k^* &= \underset{k \in \mathbb{N}}{\operatorname{argmax}} \left\{ v(k) - (1-\beta)\frac{1}{k} \sum_{j \in \mathcal{S}_i(k)} \mathbb{E}[(\theta_i - \theta_j)^2] \right\} \\
&= \underset{k \in \mathbb{N}}{\operatorname{argmax}} \left\{ v(k) + (1-\beta)\frac{1}{k} \sum_{j \in \mathcal{S}_i(k)} \sigma_{ij} \right\} \qquad (4)
\end{aligned}
$$

This k^* yields some within-the-platform utility $\bar{u}_i$. Now fix $\bar{u}_i$ and modify $\mathcal{S}_i$ by taking the most similar user to user i (say, user r) and putting her at the top of the feed (i.e., redefining $\mathcal{S}_i(2) = \{i, r\}$ and keeping the rest of the users in the same order). This implies that when the user reoptimizes and chooses $(k^*)'$, necessarily $(k^*)' \geq k^*$ because the second term in (4) is now larger. Repeating this process yields to the platform implementing the closest algorithm $\mathcal{C}$, as we wanted to show. □

C Omitted proofs

Proof of Lemma 3.2

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

$\theta_{\mathcal{S}_i}$ is proportional to the likelihood function:

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

Manipulating the expression:

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

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

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

as we wanted to show. □

Proof of Proposition 4.1

Proof. The feed length k is fixed. 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 user i's neighbors $j_1, ..., j_{k-1}$ such that $\rho_{i,j_r} > 1 - \varepsilon$ for all $r \in \{1, ..., k-1\}$. 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:

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

Using the bounds derived above, we obtain:

$$\rho_{j_r,j_l} \geq (1-\varepsilon)^2 - 2\varepsilon = 1 - 4\varepsilon + \varepsilon^2 \quad \forall\, j_r, j_l.$$

Let us define $\delta = 4\varepsilon - \varepsilon^2$. For every $\delta > 0$, there is some $\tilde{N}$ such that if $N > \tilde{N}$, the feed induced by the closest algorithm $\mathcal{S}_N^{\mathcal{C}} \subset N$ verifies that if $j_r, j_l \in \mathcal{S}_N^{\mathcal{C}}$, $\rho_{j_r,j_l} > 1 - \delta$ (it is enough to choose ε accordingly). Hence, we have that for the matrix $\mathbf{A}$ defined as

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

$\mathbf{A} \leq \mathbf{\Sigma}_{\mathcal{S}_N^{\mathcal{C}}}$, where $\leq$ refers to element-wise ordering and $\mathbf{\Sigma}_{\mathcal{S}_N^{\mathcal{C}}}$ is the covariance matrix for the users in $\mathcal{S}_N^{\mathcal{C}}$. Now, we need an auxiliary result:

Lemma C.1. *In this particular case, $\boldsymbol{A} \leq \boldsymbol{\Sigma}_{\mathcal{S}_N^{\mathcal{C}}}$ implies $\boldsymbol{\Sigma}_{\mathcal{S}_N^{\mathcal{C}}}^{-1} \leq \boldsymbol{A}^{-1}$.*

Proof. Let $\mathbf{A}$ be the covariance matrix selected by the closest algorithm, i.e., $\mathbf{A} = \boldsymbol{\Sigma}_{\mathcal{S}_N^{\mathcal{C}}}$:

$$\mathbf{A} = \begin{pmatrix} 1 & a_{12} & a_{13} & \dots & a_{1n} \\ a_{12} & 1 & a_{23} & \dots & a_{2n} \\ a_{13} & a_{23} & 1 & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & a_{3n} & \dots & 1 \end{pmatrix}$$

Let $\mathbf{B}$ be the following matrix

$$\mathbf{B} = \begin{pmatrix} 1 & b & b & \dots & b \\ b & 1 & b & \dots & b \\ b & b & 1 & \dots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b & b & b & \dots & 1 \end{pmatrix}$$

with $b = 1 - \delta$ such that $\mathbf{B} \leq \mathbf{A}$ element-wise. We denote the elements of the inverse matrices $\mathbf{A}^{-1}$ and $\mathbf{B}^{-1}$ as follows:

$$\mathbf{A}^{-1} = \begin{pmatrix} \bar{a}_{11} & \bar{a}_{12} & \bar{a}_{13} & \dots & \bar{a}_{1n} \\ \bar{a}_{12} & \bar{a}_{22} & \bar{a}_{23} & \dots & \bar{a}_{2n} \\ \bar{a}_{13} & \bar{a}_{23} & \bar{a}_{33} & \dots & \bar{a}_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{a}_{1n} & \bar{a}_{2n} & \bar{a}_{3n} & \dots & \bar{a}_{nn} \end{pmatrix},$$

and

$$\mathbf{B} = \alpha \begin{pmatrix} 1 & \bar{b} & \bar{b} & \dots & \bar{b} \\ \bar{b} & 1 & \bar{b} & \dots & \bar{b} \\ \bar{b} & \bar{b} & 1 & \dots & \bar{b} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{b} & \bar{b} & \bar{b} & \dots & 1 \end{pmatrix}.$$

Now, as $\mathbf{A}\mathbf{A}^{-1} = \mathrm{Id}$, $\bar{a}_{11} + a_{12}\bar{a}_{12} + a_{13}\bar{a}_{13} + ... + a_{1k}\bar{a}_{1k} = 1$. Moreover, $\mathbf{A} \geq \mathbf{B}$ implies that $\bar{a}_{11} + b\sum_{j=2}^{k} \bar{a}_{1j} \leq 1$. On the other hand, as $\mathbf{B}\mathbf{B}^{-1} = \mathrm{Id}$, $\alpha(1 + b\bar{b}(k-1)) = 1$. Hence,

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

This implies that $\bar{a}_{11} \leq \alpha$ and $\sum_{j=2}^{k} \bar{a}_{1j} \leq \alpha(k-1)\bar{b}$. Following the same reasoning, we obtain

$$\bar{a}_{ii} \leq \alpha \quad \forall\, i \text{ and } \bar{a}_{ij} \leq \alpha \bar{b} \quad \forall\, j \neq i.$$

Then, $\mathbf{A}^{-1} \leq \mathbf{B}^{-1}$ as we wanted to show. ◻

Therefore,

$$\mathbb{1}\mathbf{\Sigma}_{\mathcal{S}_N^{\mathcal{C}}}^{-1}\mathbb{1}^t \leq \mathbb{1}\mathbf{A}^{-1}\mathbb{1}^t \Rightarrow \frac{1}{\mathbb{1}\mathbf{A}^{-1}\mathbb{1}^t} \leq \frac{1}{\mathbb{1}\mathbf{\Sigma}_{\mathcal{S}_N^{\mathcal{C}}}^{-1}\mathbb{1}^t} \Rightarrow \frac{1}{\mathbb{1}\mathbf{A}^{-1}\mathbb{1}^t} \leq \mathrm{Var}[\theta|\{\theta_j\}_{j\in\mathcal{S}_{N_i}}].$$

On the other hand, we have that $\mathrm{Var}[\theta|\{\theta_j\}_{\mathcal{S}_N^{\mathcal{C}}}] \leq \sigma^2$ by construction (note that $\mathrm{Var}[\theta|\theta_i] = \sigma^2$). Consequently, after calculating $\mathbb{1}\mathbf{A}^{-1}\mathbb{1}^t = \frac{k}{\sigma^2(1+(k-1)(1-\delta))}$, we finally get:

$$\frac{\sigma^2(1+(k-1)(1-\delta))}{k} \leq \mathrm{Var}[\theta|\{\theta_j\}_{\mathcal{S}_N^{\mathcal{C}}}] \leq \sigma^2$$

for every $\delta \in (0,1)$. Finally, we have that $\delta \to 0$ as $N \to \infty$. Then, taking limits in the above expression we obtain that $\mathrm{Var}[\theta|\{\theta_j\}_{\mathcal{S}_N^{\mathcal{C}}}] = \sigma^2$.

◻

Proof of Proposition 4.4

Proof. We have assumed that there is a pool of $k-1$ users in N such that $\rho_{jl} = 1-\varepsilon(N)$ and $\lim_{N\to\infty}\varepsilon(N) = 0$ for all j,l belonging to such pool and user i. This pool is precisely $\mathcal{S}_i^{\mathcal{C}}$.

Let us now prove the result. First, we will compute within-the-platform utility for the three cases, and then learning. Expected within-the-platform utility is given by;

$$\mathbb{E}[u_i|\theta_i,\mathcal{F}] = v - (1-\beta)\sum_{j\in\mathcal{S}_i}\frac{\mathbb{E}[(\theta_i-\theta_j)^2|\theta_i]}{k+1}.$$

Given that $\mathbb{E}[(\theta_i-\theta_j)^2|\theta_i] = \mathbb{E}[(\mathbb{E}[\theta_j|\theta_i]-\theta_j)^2|\theta_i] = \mathrm{Var}[\theta_j|\theta_i] = \sigma^2(1-\rho_{ij})$, it holds that

$$\mathbb{E}[u_i|\theta_i,\mathcal{F}] = v - (1-\beta)\sum_{j\in\mathcal{S}_i}\frac{\sigma^2(1-\rho_{ij})}{k+1}.$$

In case (i), the k user in i's feed satisfy $\rho_{ij} = 1-\varepsilon(N)$. Hence,

$$\lim_{N\to\infty}\mathbb{E}[u_i|\theta_i,\mathcal{F}] = \lim_{N\to\infty}\left(v - (1-\beta)\sum_{j\in\mathcal{S}_i}\frac{\sigma^2(k\varepsilon(N))}{k+1}\right) = v$$

In case (ii), there are $k-1$ users in i's feed satisfying $\rho_{ij} = 1-\varepsilon(N)$, while the chosen uncorrelated extra user (let us denote this user as $k+1$) features $\rho_{i,k+1} = 0$. Then,

$$\lim_{N\to\infty}\mathbb{E}[u_i|\theta_i,\mathcal{F}] = \lim_{N\to\infty}\left(v - (1-\beta)\sum_{j\in\mathcal{S}_i}\frac{\sigma^2((k-1)\varepsilon(N)+1)}{k+1}\right) = v - (1-\beta)\frac{\sigma^2}{k+1}.$$

In case (iii), there are $k-1$ users in i's feed satisfying $\rho_{ij} = 1-\varepsilon(N)$, while the chosen maximally negatively correlated extra user features $\rho_{i,k+1} = \delta(N)-1$ with $\lim_{N\to\infty}\delta(N) = 0$. Hence,

$$\lim_{N\to\infty}\mathbb{E}[u_i|\theta_i,\mathcal{F}] = \lim_{N\to\infty}\left(v-(1-\beta)\sum_{j\in\mathcal{S}_i}\frac{\sigma^2((k-1)\varepsilon(N)+(2-\delta(N)))}{k+1}\right) = v-(1-\beta)\frac{2\sigma^2}{k+1}.$$

Next, we calculate how adding an extra user affects learning. Case (i) is already solved in Proposition 4.1: $\lim_{N_i\to\infty}\mathrm{Var}[\theta|\{\theta_j\}_{j\in\mathcal{S}_i}] = \sigma^2$. Cases (ii) and (iii) require some steps more. The similarities between the k users (including user i) that originally form the feed induce the following matrix

$$\mathbf{\Sigma} = \sigma^2\begin{pmatrix} 1 & 1-\varepsilon & \dots & 1-\varepsilon \\ 1-\varepsilon & 1 & \dots & 1-\varepsilon \\ \vdots & \vdots & \ddots & \vdots \\ 1-\varepsilon & \dots & 1-\varepsilon & 1 \end{pmatrix},$$

where $\lim_{N\to\infty}\varepsilon = 0$ and we have written ε instead of $\varepsilon(N)$ for convenience. In case (ii), adding an extra user implies that the extended matrix is

$$\mathbf{\Sigma}' = \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \sigma^2 \end{bmatrix}, \text{ and } (\mathbf{\Sigma}')^{-1} = \begin{bmatrix} \mathbf{\Sigma}^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sigma^2} \end{bmatrix}.$$

The posterior variance is given by:

$$\begin{aligned}\lim_{N\to\infty}\mathrm{Var}[\theta|\{\theta_j\}_{j\in\mathcal{S}_i}] &= \lim_{N\to\infty}\frac{1}{\mathbb{1}(\mathbf{\Sigma}')^{-1}\mathbb{1}^t} = \lim_{N\to\infty}\frac{1}{\mathbb{1}\mathbf{\Sigma}^{-1}\mathbb{1}^t+\frac{1}{\sigma^2}} \\ &= \lim_{N\to\infty}\frac{1}{\frac{k}{\sigma^2(1+(k-1)(1-\varepsilon))}+\frac{1}{\sigma^2}} = \frac{\sigma^2}{2}.\end{aligned}$$

In case (iii), the extra user is maximally negatively correlated with all users in the pool. Following a similar argument of that from Proposition 4.1, $\rho_{j,k+1} = \delta(N)-1$ where $\lim_{N\to\infty}\delta(N) = 0$ for all j in the pool. Hence, the extended matrix is

$$\mathbf{\Sigma}'' = \sigma^2\begin{pmatrix} 1 & 1-\varepsilon & \dots & 1-\varepsilon & \delta-1 \\ 1-\varepsilon & 1 & \dots & 1-\varepsilon & \delta-1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1-\varepsilon & \dots & 1-\varepsilon & 1 & \delta-1 \\ \delta-1 & \dots & \delta-1 & \delta-1 & 1 \end{pmatrix}.$$

Hence,

$$\lim_{N\to\infty}\mathrm{Var}[\theta|\{\theta_j\}_{j\in\mathcal{S}_i}] = \lim_{N\to\infty}\frac{1}{\mathbb{1}(\mathbf{\Sigma}'')^{-1}\mathbb{1}^t} = \lim_{N\to\infty}\frac{\sigma^2((k-1)(\delta-2)\delta+(k-2)\varepsilon)}{2\delta(k-1)+\varepsilon(k-2)-4(k-1)} = 0.$$

Finally, we compute the total expected utility from each case and compare them. Let us denote by $\mathbb{E}[U_i|\theta_i, \mathcal{F}^{BCE}]$ user i's total expected utility under the breaking echo chambers algorithm. Then, $\mathbb{E}[U_i|\theta_i, \mathcal{F}^{BCE}] = \lambda v - (1-\beta)\lambda \frac{2\sigma^2}{k+1}$. $\mathbb{E}[U_i|\theta_i, \mathcal{C}] = \lambda v - (1-\lambda)\sigma^2$ and $\mathbb{E}[U_i|\theta_i, \mathcal{R}] = \lambda v - \lambda(1-\beta)\frac{k}{k+1}\sigma^2 - (1-\lambda)\frac{\sigma^2}{k+1}$. Finally, the total expected utility under (ii) is $\mathbb{E}[U_i|\theta_i, \mathcal{F}^{(ii)}] = \lambda v - \lambda(1-\beta)\frac{\sigma^2}{k+1} - (1-\lambda)\frac{\sigma^2}{2}$. Simple algebra yields the desired result. □