Digital Platforms and Social Networks

Three Essays in Microeconomic Theory

Inaugural-Dissertation

zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften

durch

die Rechts- und Staatswissenschaftliche Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Miguel Risco

aus Salamanca, Spanien

Bonn

2025

Dekan:	Prof. Dr. Jürgen von Hagen
Erstreferent:	Prof. Dr. Sven Rady
Zweitreferent:	Prof. Dr. Francesc Dilmé
Tag der mündlichen Prüfung:	18. November 2024

A mis padres, Gabriel y Julia,
pues nada me enorgullecerá más que ser digno hijo suyo.

Acknowledgements

Many years ago, my father taught me the poem Ithaca by Kavafis. Like Ulysses, I envisioned my goal clearly on the horizon, and the journey toward it became my true destiny. Fortunately, I have never rowed alone on this voyage. Like almost everything in life, these pages are not an individual product but rather a collective one. I am infinitely grateful to those who, in one way or another, have contributed to my learning, growth, and happiness, to enjoying life with all its colors and complexity, and, in particular, to the completion of this chapter after six years. I will humbly try to express my gratitude in the following lines.

To Sven Rady, thank you for skillfully guiding a stuttering student whose first writings read like "a toothache." I deeply admire your elegance as a teacher and even more your integrity and honesty as a person. Thank you for being demanding yet understanding. Your teachings will always accompany me. To Francesc Dilmé, thank you for your infinite kindness, always able to understand my models in a couple of minutes more deeply than I do. Talking with you always made me feel close to home. To Alexander Frug, thank you for taking such good care of me during my adventure at UPF; our conversations marked a turning point in my understanding of the profession. Your contribution far exceeds what I deserve. Thank you from the bottom of my heart. To David Jiménez-Gómez, thank you for your commitment and dedication since 2017, and for your sincerity and patience. Of course, I also thank the Bonn Graduate School Economics, the DAAD, and the CRC TR 224 because, without them, this journey would have been impossible.

It would have also been impossible without my travel companions, the 2018 cohort. I do not remember a moment when I did not receive a helping hand when I needed it. In particular, thank you, Andrea and Ege, for your complicity and generosity, and thank you, Lenard, for everything. My gratitude towards you cannot fit in these lines. You are the friend anyone would wish to have. Our anecdotes fill my heart. Some others arrived later, but I am sure they are here to stay. Thank you, Moritz, Carl-Christian, and Jacopo, for every conversation and exchange (usually one-sided) of ideas. Listening to you is extremely enriching and one of the greatest pleasures of the profession. And, of course, thank you, Manuel, co-author and friend, with whom everything flows as if we had known each other

since childhood. May our dreams come true: "[...] the point, however, is to change it."

If anything made the stress more bearable and provided an oasis in the harsh winter, it was meeting every day at Sportpark Nord with my friends from SSF Bonn, the triathlon team with the highest density of PhD students in history. Thank you, Christoph, for treating me as one of the family; thank you all for your affection and understanding. I have been immensely happy swimming, cycling, and running, sometimes alone, sometimes accompanied, sometimes in good weather, and many other times not, through the landscapes of this beautiful city. And I cannot forget, of course, my dear neighbor Simon, without whom I literally would not be writing these lines.

"It takes a village to raise a child," the saying goes. Thank you to those who are part of my adventure from every part of Spain. You are too many to name all of you. My eternal friends from college, Carlos and Miriam, my family in Alicante, who always make me feel as if time has not passed, Emilio, Miguel, Marc, Adrián, Ana; the Catalan-Mallorcan, Núria, Pau, Raquel, Llorenç, Guillem, filled with unforgettable memories, and the adopted family, Valencian, Jesús, and the whole C.E.A. Bétera. Also Ander, José, and of course, my friends from Salamanca, Iván, Víctor, Sergio, Marcos, and Pablo. Thank you all. I am truly honored to consider you my friends. Thank you also, Belén, for your teachings about love and feminism, and for taking such good care of me when winter was too scary. Thank you, Víctor, for living far but close all these years. Thank you for always being in my corner.

Having said this, nothing for me can compare to the love of my family. None of my successes, if they can be called that, would be possible without them. Thank you, Javier, my little brother, for your loyalty and unconditional support. I will always bet everything on our pair. Your love far exceeds what I deserve. Gabriel, Julia, my parents. This thesis and everything that makes me proud is dedicated to you. I have no words to describe the gratitude I feel and the admiration I have for you: humble, hardworking, and brave. Thank you for building our home, for instilling your values in us, and above all, for holding my hand and never letting go.

Finally, thank you, Sara, for your pure love and that happy sparkle in your eyes that could overcome any difficulty. Thank you for your unwavering trust and for accompanying me on this adventure. I hope to keep reaching unseen ports, like Ulysses, but in your company, for then no Cyclops or Laestrygonian will be able to scare me.

Contents

Acknowledgements **v**

Introduction **1**

1 Feed for Good? On the Effects of Personalization Algorithms in Social Media Platforms **3**

1.1 Introduction **3**

1.1.1 Related literature 9

1.2 A model of communication and learning through personalized feeds **11**

1.3 Platform-optimal algorithm **16**

1.4 The reverse-chronological algorithm and other alternatives **21**

1.5 The need for horizontal interoperability **26**

1.6 Conclusion **30**

Appendix 1.A Omitted proofs **32**

Appendix 1.B Example **38**

References **41**

2 Based on the Papers You Liked: Designing a Rating System for Strategic Users **47**

2.1 Introduction **47**

2.1.1 Related literature 51

2.2 Likes and Double-Likes: a Model of Quality Feedback **52**

2.2.1 Baseline rating system 53

2.2.2 More granular rating system 57

2.3 Framework: Two-Sided Platform **59**

2.3.1 Equilibrium analysis 61

2.4 Monopolistic Platform **62**

2.4.1 Different rating systems 63

2.4.2 Comparison of rating technologies 65
2.5 Price-taker platform 67
2.6 Conclusion 69
Appendix 2.A Omitted proofs 71
Appendix 2.B Concavity of $\tilde{b}^2(k)$ 75
References 76

3 Network Effects on Information Acquisition by DeGroot Updaters 79
3.1 Introduction 79
3.2 Model 84
3.3 Equilibrium 86
3.3.1 Examples 89
3.3.2 Equilibrium characterization 94
3.3.3 Equilibrium uniqueness 95
3.4 Social Welfare 105
3.5 Extension to multiple periods 107
3.5.1 Examples: Infinitely many communication periods 109
3.6 Conclusion 111
Appendix 3.A Quasi-Bayesian Foundation 113
References 114

Introduction

This dissertation consists of three self-contained chapters that employ theoretical methods to shed light on the effects of personalization algorithms in social media platforms, the role of rating systems on streaming platforms and information acquisition in social networks, respectively.

In an increasingly digitized society, it is crucial to theoretically understand the strategic interplay between profit-maximizing platforms and users. This understanding will eventually lead to policy recommendations. This is the spirit of the first two chapters. Specifically, the first chapter examines how a social media platform governs informational exchanges between users. Once logged out, these users interact with people in their real-life networks, discussing relevant topics and influencing others. Their decisions regarding information provision are crucial for society. The third chapter precisely analyzes how individuals acquire information based on the networks they are part of.

In the first chapter, which is joint work with Manuel Lleonart-Anguix, Ph.D. student at Universitat Autónoma de Barcelona, we build a theoretical model of communication and learning in a social media platform, and describe and characterize the algorithm an engagement-maximizing platform implements in equilibrium. Such algorithm excessively exploits similarity, locking users in echo-chambers. Moreover, learning vanishes as platform size grows large. As this is far from ideal, we explore alternatives. The reverse-chronological algorithm the DSA mandated to reincorporate turns to be not good enough, so we build the "breaking echo chambers" algorithm, a modification of the platform-optimal algorithm that improves learning by promoting opposite thoughts. Additionally, we seek a natural implementation path for the utilitarian optimal algorithm. This is why we advocate for horizontal interoperability. Horizontal interoperability compels platforms to compete based on algorithms. In the absence of platform-specific network effects that entrench users within dominant platforms, the retention of user bases hinges on implementing algorithms that outperform those of competitors.

In the second chapter, which is joint work with Jacopo Gambato, Ph.D. student at Universität Mannheim, we focus on streaming platforms and the rating systems

they implement. Such rating systems allow streaming platforms to leverage users' experience to signal quality of the third party products they host. We study the interaction between strategic rating by users, granularity of the rating technology, and streaming platform size. Users rate products to be grouped with other users with similar tastes, to then be able to receive high quality recommendations, an effort that is more effective the larger the platform and the more granular the rating system are. Users become more demanding as the selection of potentially available content grows. The platform's need to generate value for users and remunerate sellers upfront leads to a trade-off: a platform with limited reach prefers more granular systems to employ users' ratings efficiently; a large one prefers a less informative and less taxing system to increase engagement. If the platform is large enough to affect competition intensity on the outside market, she has an incentive to limit access to sellers to minimize operational costs.

In the third chapter, I analyze how social networks impact information processes, shape individuals' beliefs and influence their decisions. This chapter proposes a model to understand how boundedly rational (DeGroot) individuals behave when seeking information to make decisions in situations where both social communication and private learning take place. The model assumes that information is a local public good, and individuals must decide how much effort to invest in costly information sources to improve their knowledge of the state of the world. Depending on the network structure and agents' positions, some individuals will invest in private learning, while others will free-ride on the social supply of information. The model shows that multiple equilibria can arise, and uniqueness is controlled by the lowest eigenvalue of a matrix determined by the network. The lowest eigenvalue roughly captures how two-sided a network is. Two-sided networks feature multiple equilibria. Under a utilitarian perspective, agents would be more informed than they are in equilibrium. Social welfare would be improved if influential agents increased their information acquisition levels.

Chapter 1

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

Joint with Manuel Lleonart-Anguix

1.1 Introduction

On May 25, 2024, a video went viral on TikTok after showing the stark difference in comments displayed on Instagram to Eli, the user who posted it, and her boyfriend when reading the same post.[1] In the post, which is public, we see a girl waiting for her boyfriend, who was supposed to meet her at 3 p.m. after playing golf. The post shows the girl recording herself after each extra half-hour she has to wait for him. When Eli read the comments below the post, which were displayed

* I acknowledge financial support by the German Research Foundation (DFG) through CRC TR 224 (Project B05) and funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 949465). We thank Sven Rady, Pau Milán, Francesc Dilmé, Alexander Frug, Carl-Christian Groh, David Jiménez-Gómez, Daniel Krähmer, Philipp Strack, Marc Bourreau, Mikhail Drugov, Manuel Mueller-Frank, Robin Ng, Raquel Lorenzo, Malachy Gavan, Francisco Poggi, Fernando Payro, Jordi Caballé, Inés Macho-Stadler, Iván Rendo and Diego Fica and the participants at the BSE Summer Forum in Digital Platforms, the EAYE 2024, the Paris Digital Economics Conference, the MaCCI Annual 2024, the CRC TR 224 retreat, the 6th Workshop on The Economics of Digitalization, the EDP Jamboree, the 2nd ECONtribute YEP Workshop, the CoED, the JEI, various CRC TR 224 YRWs, the ENTER Jamboree, the BSE Jamboree, as well as seminars and workshops at the University of Bonn, UPF, TSE, UAB, UV, UA, EUI and UB for their helpful comments and discussions.

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

under the *most relevant* tag, they were along the lines of "oh this is so rude!" or "it is disregard of her time"—as she expected, as she literally says. Eli then sent the post to her boyfriend, who was sitting next to her. He opened it at the same time, but surprisingly, the *most relevant* comments were strikingly different: "or you could get your own hobby instead of waiting around for him", "he meant 3 a.m., he is ahead of schedule" or "God forbid he has a good time". People look at the comments on a video to gain perspective and see how others feel about it. However, now, with personalized feeds, each user gets tailor-made content based on her interactions and behavior in the app. "The only difference about our interactions with Instagram is that he is a guy and I am a girl", Eli says. She tried to find the comments that appeared to her boyfriend in her own list, but she could not.

Eli's video went viral, reaching almost three million views and adding fuel to the social debate on the effects of personalized social media feeds on people's beliefs and perspectives. However, these concerns are not new, as platforms 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 elections were likely the turning point where public opinion began to question the suitability of personalization and its consequences on beliefs and decisions; Facebook was accused of failing to combat fake news (Solon, 2016). Since then, much evidence on this has been collected at the academic, empirical level (Allcott, Gentzkow, and Song, 2022; Bursztyn, Handel, Jiménez-Durán, and Roth, 2023), showing how harmful engagement-maximizing platforms are and how they trap users in their echo-chambers (in particular, Bursztyn et al. (2023) show that users would be willing to pay to have others, including themselves, deactivating their TikTok and Instagram accounts). Moreover the journalistic investigations by Horwitz et al. (2021) called "The Facebook Files" revealed that Meta internally acknowledges the harmful effects of its algorithms on users, specifically on female teenagers ("[t]ime and again, the documents show, Facebook's researchers have identified the platform's ill effects"). However, platforms can create value through their superior level of information and,[2] then, it is essential to investigate how personalization algorithms affect social welfare, as their repercussions have emerged as a significant economic concern.

The *feed* is a customized scroll of friends' content and news stories that appears on most social media platforms. Until around 2015, it was reverse-chronological.[3]

2. Quoting Scott Morton, Bouvier, Ezrachi, Jullien, Katz, 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."

3. Social media platforms began transitioning from reverse-chronological feeds to personalized feeds at different times. Facebook started implementing personalized feeds in 2009, while

Now, a proprietary algorithm controls what appears on the screen, based on user behavior on the platform. Since platforms' revenues come from advertising, their primary goal is to maximize engagement, which may not align with promoting informative communication. If, as Eli's video shows, a biased set of comments will maximize the probability you stay on the platform longer, this is what you will receive. Personalized algorithms account for the increase in engagement and addictive behavior in social media platforms, regardless of the field (Guess, Malhotra, Pan, Barberá, Allcott, et al., 2023).

The approval of the Digital Service Act (DSA) and the Digital Markets Act (DMA) by the European Commission in 2022 represents one of the first efforts to address the problems arising from algorithm personalization through regulation. In particular, the DSA requires platforms to reinstate the reverse-chronological algorithm as an option for their users, thereby providing an alternative to personalization. Some platforms, like X, were very compliant, while others, like Instagram, were less so: it is not only complicated to find the button reverting the feed to reverse-chronological, but the feed goes personalized again once you log out. Still, it does not seem that the availability of the reverse-chronological algorithm is really alleviating any of the urgent media-related problems society faces. Moreover, personalization need not be detrimental to social welfare; it could be used for an improvement, as the following quote illustrates (Lauer, 2021): "If Facebook employed a business model focused on efficiently providing accurate information and diverse news, rather than addicting users to highly engaging content within an echo chamber, the algorithmic outcomes would be very different". To achieve this, however, we would need to find a way to align the platform's incentives with social well-being, so that naturally, the optimal algorithm for the platform would also be optimal for the users.

With all this in mind, we claim that there is a need for theoretical research that guides the optimal regulatory approach, understanding the incentives of platforms in designing their optimal algorithm and how they would respond to regulation. It is crucial to understand the strategic interplay between an engagement-maximizing platform and users who value not just the instantaneous joy coming from scrolling down and posting their thoughts, but also the reward from learning. This is precisely what this chapter intends to achieve.

To do so, we build a theoretical model of communication and learning on a social media platform, describing and characterizing the algorithm a platform implements in equilibrium. We find that such algorithm exploits similarity too much,

---

Twitter (now X) and Instagram transitioned between 2015 and 2016. Younger platforms, like TikTok, have provided curated content since their launch.

locking users in echo chambers. Moreover, learning disappears as platform size grows large. As this is far from ideal, we explore alternatives. As we could expect, the reverse-chronological algorithm the DSA mandated to reincorporate is not that proficient: in general, it cannot compete against the platform-optimal algorithm. In light of recent efforts from platforms to *give context* or *promote fact-checked content*, we build the "breaking echo chambers" algorithm, a modification of the platform-optimal algorithm that improves learning significantly as platform size grows large. Still, we look for a natural way for platforms to implement the optimal algorithm for the users, the utilitarian optimal algorithm. This is why we propose horizontal interoperability. Under horizontal interoperability, platforms are forced to compete on algorithms because, absent platform-specific network effects that capture users in the dominant site, the only way to retain the user base is for platforms to implement an algorithm that is preferred over those implemented by other competing platforms. This simple argument à la Bertrand leads to platforms opting for the utilitarian optimal algorithm.

We highlight four main contributions of this chapter. First, we build a model where users post messages and learn through a feed designed by the engagement-maximizing platform. We assume that users derive instantaneous utility from engaging in communication with peers about some underlying topic, and we call this utility stream *within-the-platform utility*. It has three channels: satisfaction is brought by reading a post written by a friend, expressing one's own views (in the sense of being loyal to own innate opinions; *sincerity*), and conforming with the rest (in the sense of matching the opinions that others have shared; *conformity*).[4] The strength of these incentives depends on model parameters. In particular, we encompass situations in which conformity is almost negligible. The second utility stream comes from gathering valuable information on the platform to improve a decision, termed *action utility*.[5] The effectiveness of the Covid-19 vaccine, which

4. Conformity is a driving-force in social media behavior (Mosleh, Martel, Eckles, and Rand, 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.

5. The first component of the utility function is similar to the payoffs in Galeotti, Golub, Goyal, and Rao (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".

triggered significant public debate,[6] is our leading example, as each part of the previously described utility function can be easily identified. People, driven by both a desire for sincerity and conformity, used social media to express their views about the benefits and risks of vaccination. Note that individuals sought to communicate their personal viewpoints because vaccination was a pivotal societal concern, but at the same time expressing dissenting opinions proved to be socially taxing. Additionally, gathering information was crucial to deciding whether to get vaccinated or not.

Engagement is defined, in this chapter, as the number of posts a user reads. It is equivalent to the time spent scrolling down before logging out. Crucially, we assume that users do not rationally decide how much to engage, but that their engagement is controlled by a stochastic process driven by the instantaneuos joy of consuming content. After reading each post, a user continues scrolling down with some probability depending on the instantaneous within-the-platform utility derived. Otherwise, she logs out. Scrolling is, then, seen as an addictive behavior the user does not rationally control (Allcott, Gentzkow, and Song, 2022): it is a rather automatic process corresponding with the intrinsic happiness derived within the platform.[7] However, the explicit decision to post a message is seen as a rational move in which the user consciously acts to achieve a goal. Users post messages and then observe those which appear in their feeds until they log out. Afterwards, they take an action based on the information gathered. Feeds are the product of an algorithm designed by the platform which, as explained earlier, has no incentives to promote learning: engagement purely depends on within-the-platform utility. The platform, which does not read messages, designs the algorithm leveraging its information on users' similarities in views. We assume the platform knows perfectly how similar users are, and utilizes this information to maximize its profits, i.e., to maximize total engagement. We think of similarities being derived from past interactions and users' personal data by using sophisticated machine learning techniques.[8]

6. Loomba, Figueiredo, Piatek, Graaf, and Larson (2021) find that the acceptance of the Covid-19 vaccine in US and UK declined an average of 6 percentage points due to misinformation.

7. This assumption could be also interpreted following the *Dual Process Theory* as in Benhabib and Bisin (2005). For an overview on Dual Process Theory, see Grayot (2020). The fact that engagement depends only on within-the-platform utility can be seen under the light of present bias: the user weights disproportionally low the benefits of learning when reading posts.

8. 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, Stillwell, and Graepel (2013) already showed that less sophisticated techniques could predict a wide range of personal attributes by just using data on "likes".

Our second contribution is to identify the platform-optimal algorithm and study its properties. As expected, the platform-optimal algorithm is driven by the desire to maximize expected conformity, because it is the main force behind engagement. However, the fact that each user knows her own signal creates an information friction and the platform brings *too much* similarity to the feeds. The user would have preferred to have more diverse views, but is locked in an echo chamber, precisely as Eli shows in her video. This excess of similarity in the feeds becomes more pronounced as the platform size grows. Then, the feed becomes flooded with close *copies* of a user and consequently learning vanishes, contrasting with classical results where large societies learn better (Golub and Jackson, 2010).

The third contribution consists in studying alternatives to the platform-optimal algorithm. We start with the reverse-chronological algorithm brought back by the DSA. This algorithm is generally not good enough to be considered a suitable alternative, so we analyze a variation of the platfom-optimal algorithm that maximizes learning when platform size is large. This is the breaking echo chambers algorithm, which adds a user with opposite views at the top of each feed induced by the platform-optimal algorithm. It improves learning but slightly decreases conformity and consequently engagement. While it still outperforms the platform-optimal algorithm for many types of users, it is plausible that real-world individuals may disregard information from a completely opposing source, complicating its practical implementation.

Regardless, the utilitarian optimal algorithm is the only one that maximizes social welfare, so then we must explore its implementation. This leads us to the fourth and last contribution of this chapter, the discussion whether implementing horizontal interoperability would suffice for platforms to opt for the utilitarian optimal algorithm through competition. Without horizontal interoperability, the network effects that social media platforms feature (i.e., the fact that the more users join a platform, the more valuable its service becomes) create high barriers to entry and induce winner-takes-all (or most) market dynamics. Horizontal interoperability compels platforms to connect, so that users from different platforms can be linked. A user's feed would then be an ordered list of the posts coming from all her friends, regardless of which platform they are registered on, designed by the platform she joined. Crucially, network effects will be shared, and platforms will have to compete along the non-interoperable dimension, i.e., they will have to compete in algorithms. Each user will join the platform whose algorithm offers the highest expected utility, disregarding platform size. And this algorithm is, of course, the utilitarian optimal algorithm. Then, competing platforms would be *forced* to implement this algorithm; otherwise, they risk losing their user base. The pursuit of this goal aligns with the intentions of EU reg-

ulators, as reflected in the Digital Markets Act,[9] which mandates certain large social platforms to achieve interoperability in their messaging communications in the immediate future. Quoting Kades and Scott Morton (2020): "Interoperability eliminates or lowers the entry barrier, which is the anticompetitive advantage the platform has maintained and exploited. Users will not switch to a new social network until their friends and families have switched. [...] Interoperability causes network effects to occur at the market level—where they are available to nascent and potential competitors—instead of the firm level where they only advantage the incumbent."

The rest of the chapter is organized as follows. After the literature review, each section corresponds to each of the contributions described above: Section 1.2 develops the model, Section 1.3 finds the platform-optimal algorithm and characterizes it, Section 1.4 analyzes alternative algorithms, and Section 1.5 discusses horizontal interoperability. Finally, Section 1.6 concludes.

1.1.1 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.

In broad terms, our chapter 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, Pai, Reggini, Saporiti, and Simantujak (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, Ozdaglar, and Siderius (2023) study 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

9. 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.

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", also appear 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 analyze the effects of filter bubbles 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, Li, and Tan (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 chapter is closely related. Franck and Peitz (2023) study competition between social media platforms, claiming that market power (mainly represented by the network effects) leads to suboptimal outcomes for society. In particular, it may not be the platform with the best offer that dominates the market. Biglaiser, Crémer, and Veiga (2022) offer a micro-foundation for incumbent advantage. Essentially, network effects prevent users from migrating to even Pareto-superior equilibria when they receive stochastic opportunities to migrate to an entrant. Kades and Scott Morton (2020) also examine network effects in digital platforms and offer an overview of interoperability. Popiel (2020) and Evens, Donders, and Afilipoaie (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, Malladi, and McAdams (2022) examine 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, Ananthakrishnan, and Tucker (2022) provide empirical evidence of the negative consequences of deplat-

forming (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, Musa, Zaffar, and Nithyanand (2019), Hwang and Lee (2021) or Mudambi and Viswanathan (2022). Mostagir and Siderius (2023b) model community formation and show 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) demonstrate that cognitive sophistication matters when faced with misinformation, and Mostagir and Siderius (2023a) find 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) and Pennycook and Rand (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, Vayanos, and Zwiebel (2003), Acemoglu and Ozdaglar (2011), Jadbabaie, Molavi, Sandroni, and Tahbaz-Salehi (2012), Molavi, Tahbaz-Salehi, and Jadbabaie (2018) or Mueller-Frank and Neri (2021).

1.2 A model of communication and learning through personalized feeds

Here we present the baseline model of the chapter. We start by providing an overview, then delve into the formal details, and finally discuss some of the assumptions made along the way.

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

Users' utility comprises two components: the *within-the-platform utility*, influenced by engagement, conformity, and sincerity, and the *action utility*, which depends on learning, i.e., how close users can get to the state of the world after

communicating on the platform. The platform's revenues, in turn, are contingent upon user engagement. Hence, the platform designs an algorithm seeking to maximize such engagement by leveraging information on similarities between users' worldviews. This algorithm curates a personalized feed for each user, determining the order in which messages appear on the scrolling screen.

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

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

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

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

Users know their private signals, the distribution of all signals, the covariance matrix $\boldsymbol{\Sigma}$, and the distribution of the state of the world, for which we crucially assume improper priors.[10] Thus, conditional on θ_i, the posterior distributions of θ_j and θ are normal and centered on θ_i, namely $\theta_j|\theta_i \sim \mathcal{N}\left(\theta_i, \sigma_{jj} - \frac{\sigma_{ij}^2}{\sigma_{ii}}\right)$ for all $j \in N$ and $\theta|\theta_i \sim \mathcal{N}\left(\theta_i, \sigma_{ii}\right)$.

Once logged in the platform, each user i posts a message $m_i \in \mathbb{R}$ and then observes $e_i \in \mathbb{N}$ messages that appear in her personalized feed, which is provided

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

by the platform. The number $e_i \leq n$ represents her engagement, and platform profits depend precisely on the sum of all users' engagement, $\sum_{i=1}^{n} e_i$. In order to maximize user engagement, the platform designs an algorithm consisting of an assigment that, given a pair of users i, j, tells which position user j occupies in user i's feed. Given engagement e_i, the feed is the set of users from whom messages will be observed. Formally, an algorithm $\mathscr{F}$ is a collection $(\mathscr{F}_i)_{i\in\mathscr{U}}$ where $\mathscr{F}_i \in \text{Bij}\big(\{1,...,n-1\}, \mathscr{U}\backslash\{i\}\big)$. Given $k \leq n$, $\mathscr{F}_i(k)$ is the k-th user in i's ranking induced by $\mathscr{F}$, so i's feed for engagement e_i is precisely

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

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

$$u_i(e_i, m_i, m_{-i}, \mathscr{F}_i, \theta_i) \;=\; \alpha e_i - \beta \overbrace{(\theta_i - m_i)^2}^{\text{Sincerity}} - (1-\beta) \overbrace{\sum_{j\in\mathscr{F}_i^{e_i}} \frac{(m_i - m_j)^2}{e_i}}^{\text{Conformity}}, \quad (1.2.1)$$

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

$$U_i(e_i, m_i, m_{-i}, a_i, \mathscr{F}_i, \theta_i, \theta) \;=\; \lambda\, u_i(e_i, m_i, m_{-i}, \mathscr{F}_i, \theta_i) - (1-\lambda) \overbrace{(a_i - \theta)^2}^{\text{Action utility}}, \quad (1.2.2)$$

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

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

Along the lines of digital addiction theory, we assume that the user does not rationally control her scrolling time but, after reading k posts, reads the next message with probability $g(u_i(k-1, m_i, m_{-i}, \mathscr{F}_i, \theta_i))$, where $g : \mathbb{R} \to [0,1]$ is some continuous and increasing function. With probability $1-g(u_i(k-1, m_i, m_{-i}, \mathscr{F}_i, \theta_i))$, the user discontinues scrolling down and exits. Hence, user i features engagement e_i with probability $(1-g(u_i(e_i, m_i, m_{-i}, \mathscr{F}, \theta_i)))\prod_{r=1}^{e_i-1} g(u_i(r, m_i, m_{-i}, \mathscr{F}_i, \theta_i))$. In particular, we assume that $\forall\, x \in \mathbb{R}$, $g(x) \in (0,1)$, i.e., no feed guarantees either continuation or abandonement. Note that because of the addictive nature of e_i, the user sees it as something exogenous and given.

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

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

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

Complete network. This model could be extended to any network given by some undirected graph $\mathscr{G}$. In such a case, each user i belongs to a neighborhood n_i and hence $e_i \leq |n_i|$. All the results presented below hold. Thus, we prefer to work with the complete network for ease of notation and exposition.

Monopolistic platform, all users on board. In this baseline model, we assume there is only one platform, and all users are already on board. Hence, the platform does not need to care about capturing users, but only about their engagement. This is, of course, a simplifying assumption, but the main social media platforms (Facebook, Instagram, TikTok or X) are monopolists of their fields:[12] even though they can be broadly described as social media platforms that enable public posting

12. Regarding monopoly structures in the social media platform market, the Bundeskartellamt (the German competition protection authority) states in its case against Facebook (B6-22/16, "Facebook", p. 6): "The facts that competitors are exiting the market and there is a downward trend in the user-based market shares of remaining competitors indicate a market tipping process that will result in Facebook becoming a monopolist." (Franck and Peitz, 2023).

and private communication, they differ in their core functionality. Each site dominates a specific field: photography (Instagram), short videos (TikTok), reciprocal communication with friends (Facebook), and micro-blogging (X). In most cases, there is no realistic alternative for the average user but to stay out, and then, as Bursztyn et al. (2023) show, fear of missing out makes users join even when they would prefer the platform not to exist.[13]

Improper priors. Users' prior distribution is uniform along $\mathbb{R}$. Intuitively, this means that none of them understands whether her signal is extreme. Indeed, every user believes her opinion is central (Greene, 2004). This assumption is made for the sake of model tractability. Under normal priors, we can only determine the users' optimal linear messaging strategies, but we cannot derive an explicit expression for the platform-optimal algorithm.

Non-rational engagement. Following the literature on digital addiction (Allcott, Gentzkow, and Song, 2022), we dismiss a rational framework for engagement, and opt for a simplified setting in which digital addiction is captured as a by-product of habit formation and self-control problems. The user irrationally continues scrolling down depending on the instantaneous within-the-platform utility experienced so far. Our configuration encapsulates addictive behavior in a reduced form, capturing some essential features: the probability of engaging for k periods is always higher than the probability of engaging for k' periods if $k < k'$, a higher utility derived from reading a message implies a greater probability of staying, and engagement does not depend on action utility. This last feature could be intuitively conceptualized as an extreme form of present bias: when scrolling down, the user heavily discounts the long-run reward from learning (Guriev, Henry, Marquis, and Zhuravskaya, 2023). This is also in line with the main case in Bonatti and Cisternas (2020), where consumers ignore the link between their current actions and the future consequences.

Platform's profits as a function of total engagement. Social media platforms are generally free to access, and their revenues come from advertisers' payments for product placement. These payments depend on user engagement: the larger the engagement, and hence the greater the exposure to their content, the more an advertiser is willing to pay. For simplicity, in this model the platform objective is to maximize total engagement, so its profit function is $\Pi_p(\mathscr{F}, \Sigma) = \sum_{i=1}^{n} \mathbb{E}_p[e_i]$.

13. As already commented above, Bursztyn et al. (2023) show that users would be willing to pay to have others, including themselves, deactivating their TikTok and Instagram accounts.

1.3 Platform-optimal algorithm

In this section, we obtain and characterize the algorithm the platform implements in equilibrium. First, we show that users find it optimal to report their private signals truthfully. The platform, in turn, designs a feed for each user that is excessively driven by similarity. Intended to maximize engagement, such a feed worsens user learning as the population grows until it asymptotically vanishes.

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

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

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

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

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

Proposition 1.3.1. *Given any algorithm $\mathscr{F}$, every user plays truthtelling in equilibrium, i.e., $m_i^* = \theta_i$ for all $i \in \mathscr{U}$.*

Proof. See Appendix 2.A. □

Because of improper priors, the platform cannot affect first-order moments through the feed it designs, and hence, user i believes that, in expected terms, every other user will play θ_i. Deviating from truthtelling is then not profitable.

Having shown that users play truthtelling in equilibrium, we derive the platform-optimal algorithm, denoted by $\mathscr{P}$. First, we show that maximizing profits, or total expected engagement $\sum_{i=1}^n \mathbb{E}_p[e_i]$, is equivalent to maximizing each user's expected engagement $\mathbb{E}_p[e_i]$.

Lemma 1.3.2. *It is equivalent for the platform to maximize total user engagement and maximizing each user's individual engagement separately.*

Proof. We know from Proposition 1.3.1 that, given user i, engagement e_i and a feed $\mathscr{F}_i^{e_i}$, user i plays $m_i = \theta_i$ in equilibrium: messages are not affected by the feed. Hence, there are no interdependencies across feeds: the order in which the platform ranks user j in i's feed does not affect anyone else. Finally, user i's expected engagement is a function of her expected within-the-platform utility, affected by the truthful message θ_i and her feed $\mathscr{F}_i^{e_i}$. Hence, maximizing the

sum of all users' expected engagement is equivalent to maximizing each of them individually. □

Now, we intuitively explain how the platform designs the platform-optimal algorithm $\mathscr{P}$, and what the optimal action a_i^* taken by user i is after reading her feed $\mathscr{P}_i^{e_i}$. The formal details are left to the Appendix 2.A as part of the proof of Proposition 1.3.1 below. From the point of view of the platform, and because of truthful reporting, user i's within-the-platform utility simplifies to $\mathbb{E}_p[u_i(k-1,\theta_i,\theta_{-i},\mathscr{F},\theta_i)] = \mathbb{E}_p[(v(k)-(1-\beta)\frac{1}{k}\sum_{j\in\mathscr{F}_i^{k-1}}(\theta_i-\theta_j)^2]$ when she has read $k-1$ messages. The probability of staying after reading the k-th message is then $\mathbb{E}_p[g(u_i(k,\theta_i,\theta_{-i},\mathscr{F},\theta_i))]$. To maximize this probability, the platform chooses a user j to be included next in the feed among those who have not been chosen yet, i.e., $j\in\mathscr{U}\setminus\mathscr{F}_i^k$. As g is increasing in u_i, maximizing g is equivalent to maximizing u_i. Moreover, note that conformity is the only term in which the platform can affect user i's within-the-platform utility at this stage. Hence, j is chosen according to

$$j = \underset{j\in\mathscr{U}\setminus\mathscr{F}_i^k}{\operatorname{argmax}}\{-\mathbb{E}_p[(\theta_i-\theta_j)^2]\},$$

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

$$\begin{aligned}
\mathscr{P}_i^1 &= \underset{j\in N}{\operatorname{argmax}}\{-\mathbb{E}_p[(\theta_i-\theta_j)^2]\},\\
\mathscr{P}_i^2 &= \mathscr{P}_i^1 \cup \underset{j\in N\setminus\mathscr{P}_i^1}{\operatorname{argmax}}\{-\mathbb{E}_p[(\theta_i-\theta_j)^2]\},\\
&\vdots \qquad\qquad \vdots\\
\mathscr{P}_i^k &= \mathscr{P}_i^{k-1} \cup \underset{j\in N\setminus\mathscr{P}_i^{k-1}}{\operatorname{argmax}}\{-\mathbb{E}_p[(\theta_i-\theta_j)^2]\}.
\end{aligned} \tag{1.3.1}$$

For an explicit example of how the platform designs $\mathscr{P}$ leveraging $\boldsymbol{\Sigma}$, please refer to Appendix 1.B.

Proposition 1.3.3. *In equilibrium, the platform chooses the algorithm $\mathscr{P}$ as specified in Equation (1.3.1). In other words, the algorithm that maximizes user engagement is the one that, for each user i, designs a feed in which others appear in reverse order regarding the expected loss in conformity with user i they induce.*

Proof. The formal derivation of $\mathscr{P}$ can be found in Appendix 2.A. □

The information friction between the platform and the users is crucial. On the one hand, the platform chooses the feed so as to maximize the loss in conformity, which effectively means maximizing $-\mathbb{E}_p[(\theta_i-\theta_j)^2]$ through the choice of j. But $-\mathbb{E}_p[(\theta_i-\theta_j)^2]=-\sigma_{ii}-\sigma_{jj}+2\sigma_{ij}$, so

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

On the other hand, from the user's perspective, expected conformity is $-\mathbb{E}_i[(\theta_i-\theta_j)^2|\theta_i]=-\sigma_{jj}+\frac{\sigma_{ij}^2}{\sigma_{ii}}$. User i's knowledge of θ_i notably changes the expression compared to that of the platform, and we observe that, given σ_{jj}, user i would prefer to be matched with some $j\in\mathscr{U}$ either very similar or very opposite to her. As the platform is less informed, it only selects very similar users to user i and, on top of that, fixes the weight of similarity to 2, when the user would prefer it to depend on $\frac{1}{\sigma_{ii}}$. All this drives the user to an *excessive* similarity bubble, while she would prefer to observe a more diverse feed.

Proposition 1.3.4. *The platform excessively weights similarity between users when designing its optimal algorithm.*

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

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

□

When variances are homogeneous, meaning each user has an equally precise posterior of θ, the feed becomes a reverse ranking based on similarities. Then, users are displayed according to how correlated they are to the one reading the feed.

Corollary 1.3.5. *If variances are homogeneous, i.e., $\sigma_{ii}=\sigma_{jj}=\sigma^2$ for all $i,j\in\mathscr{U}$, the platform-optimal algorithm $\mathscr{P}$ ranks uniquely in terms of similarity.*

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

Note that the user's expected conformity is $-\mathbb{E}_i[(\theta_i-\theta_j)^2|\theta_i]=-\sigma^2+\frac{\sigma_{ij}^2}{\sigma^2}$ in this particular case, and the main interpretation of the difference between what

the platform maximizes and what the user would like to be maximized remains the same. Crucially, the way messages are displayed in the feed influences users actions. The next result provides a formal expression for user i's optimal action a_i^*.

Proposition 1.3.6. *User i's optimal action after reading e_i messages, for any algorithm $\mathscr{F}$ is*

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

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

Proof. User i's optimal action maximizes $\mathbb{E}_i[(a_i-\theta)^2]$ given the observed messages $\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}$. Hence, the optimal action is

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

by Lemma 1.A.1. □

To understand the impact of the platform-optimal algorithm on social welfare, we must examine its effect on *learning*, which refers to how information gathered on the platform improves decision-making. Before doing so, we make an additional assumption for tractability purposes. We assume that users' variances are homogeneous, i.e., that $\sigma_{ii} = \sigma_{jj}$ for all users $i, j \in \mathscr{U}$. Thus, we are in the case described by Corollary 1.3.5. Note that the disparity between what users prefer to observe in their feed and what the platform provides is maintained. To clearly indicate that we are now working under homogeneous variances, we will denote the platform-optimal algorithm as $\mathscr{C}$, referring to the "closest" algorithm, as now the platform simply matches users with those who are most similar, or closest, to them.

Now, let us analyze how personalization algorithms affects information gathering. The scenario we study is precisely that of a large platform size and high engagement values, reflecting the substantial growth in social media usage in recent years, both in terms of the number of users and the time spent on platforms.[14] Learning is defined as the increase in expected action utility resulting

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

from reading messages. When a user picks the optimal action, its expected value is the posterior variance of θ conditional on the messages in the feed:

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

Applying Lemma 1.A.1, we explicitely obtain the posterior variance:

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

This expression allows us to calculate the improvement in decision-making after reading a feed for any algorithm and any society characterized by $\mathscr{U}$ and $\boldsymbol{\Sigma}$. Note that, in this model, the posterior variance is weakly lower than σ_{ii} for each user i, meaning users cannot be worse off in terms of learning after reading their feeds. We let the platform grow large now, expanding a given user base $\mathscr{U}$ by assuming that the covariances between new users and existing users are drawn from a continuous distribution with a cumulative distribution function supported in $[-\sigma^2,\sigma^2]$ and centered at 0. The resulting covariance matrix (the expanded $\boldsymbol{\Sigma}$) is symmetric and positive definite.

The manner in which $\mathscr{C}$ selects the feed becomes very apparent when platform size grows large. With a vast pool of users, conformity is maximized by choosing someone almost identical to the user. This creates a feed of close copies, resulting in an echo chamber where learning diminishes. Note that this is not what the user desires: she would prefer matches with very similar or very different individuals, which would additionally increase learning. However, the next result formally shows that, asymptotically, the platform-optimal algorithm induces no learning. This is independent of the specific engagement level of the user.

Proposition 1.3.7. *Under the closest algorithm $\mathscr{C}$, and, for any engagement level e_i, user i's learning becomes negligible as $n\to\infty$:*

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

Proof. See Appendix 1.A. □

This finding is in stark contrast to classic learning models where the wisdom of the crowd enhances learning as the population grows. Here, the platform's strategic role in feed selection undermines learning, making it vanish. In conclusion, the optimal algorithm for the platform not only creates excessive echo chambers but also harms long-term learning in large populations. These issues are significant in public debate, raising concerns about the impact of social media platforms on social welfare. The approval of the DSA and DMA in the European Union addresses these concerns. In particular, the DSA forces platforms to include the non-strategic reverse-chronological algorithm that was used before personalization algorithms

as an option for users. The next section is devoted to an analysis of alternative algorithms, including the already mentioned reverse-chronological algorithm, the user-optimal algorithm, and the breaking-echo-chambers algorithm.

1.4 The reverse-chronological algorithm and other alternatives

The reverse-chronological algorithm, which will be denoted by $\mathscr{R}$, displays friends' posts in the (reverse) order they were written. Before the implementation of personalization algorithms, every social media platform relied on this simple method of presenting the feed, which is not strategic at all. In this model, we understand the reverse-chronological algorithm as a *random* algorithm in which a post will be at the top of the feed with probability $\frac{1}{n-1}$. Consequently, this algorithm does not create echo chambers, and, as we show below, when the platform size is large, it outperforms the platform-optimal algorithm in terms of learning, but not in terms of conformity. The effect on overall utility depends on how users weight sincerity, conformity and learning. For small populations, however, it is not even the case that individuals learn better under the reverse-chronological algorithm, so we are far from stating unambiguously that the reverse-chronological algorithm is a feasible substitute for the platform-optimal algorithm, which motivates our search for a better alternative.

Given that the closest algorithm, $\mathscr{C}$, shows a feed of likeminded users when platform size grows large, learning vanishes (Proposition 1.3.7). The random nature of $\mathscr{R}$ yields better learning asymptotically:

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

This is, of course, not surprising. However, note that the higher the engagement (the e_i), the better for learning, but that $\mathscr{R}$ yields lower engagement than $\mathscr{C}$ because it is worse for conformity. This trade-off arises when we compare the expected utility under both algorithms.

Proposition 1.4.1. *Given $\boldsymbol{\Sigma}$, the closest algorithm outperforms the reverse-chronological algorithm in large populations if and only if*

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

For a general $\boldsymbol{\Sigma}$ and assuming the expected correlation between every pair of users i and j is zero, i.e., $\mathbb{E}[\rho_{ij}] = 0$ for all i,j, the condition is given by:

$$\lambda > \frac{1}{1 + \left(\frac{\mathbb{E}_i[e_i^{\mathscr{R}}](\alpha(\mathbb{E}_i[e_i^{\mathscr{C}} - e_i^{\mathscr{R}}]) + (1-\beta)\sigma^2(1 - \mathrm{Var}[\rho_{ij}]))}{\sigma^2(\mathbb{E}_i[e_i^{\mathscr{R}}] - 1)} \right)}.$$

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

Proof. The second part of the proposition was already shown above. Regarding the first result, we just compare expected utilities for the user when n grows large. For each user i, they are given, respectively, by:

$$\lim_{n\to\infty} \mathbb{E}_i[U_i(\mathscr{C})] = \lambda\, \alpha\, \mathbb{E}_i[e_i^{\mathscr{C}}] - (1-\lambda)\sigma^2;$$

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

□

Crucially, the closest algorithm maximizes user engagement, so in expectation user i reads more posts and derives higher intrinsic utility from doing so: $\alpha\, \mathbb{E}_i[e_i^{\mathscr{C}} - e_i^{\mathscr{R}}] > 0$. Hence, both terms in the numerator of

$$\frac{\mathbb{E}_i[e_i^{\mathscr{R}}]\alpha\, \mathbb{E}_i[e_i^{\mathscr{C}} - e_i^{\mathscr{R}}] + (1-\beta)\sigma^2 \sum_{j \in \mathscr{R}_i^{\mathbb{E}_i[e_i^{\mathscr{R}}]}} (1 - \rho_{ij}^2)}{\sigma^2(\mathbb{E}_i[e_i^{\mathscr{R}}] - 1)}$$

are positive. It is likely, then, that for many specifications of the model, the closest algorithm dominates the reverse-chronological algorithms.

Less intuitive is the case of small platform size, in which the comparison between both algorithms is more complicated. Of course, within-the-platform utility is always better under the closest algorithm, but we cannot state unambiguously which of the two algorithms is better for learning. Two effects must be taken into account when it comes to the latter. First, when feeds are of small length, even if engagement is the same under $\mathscr{C}$ and $\mathscr{R}$ (which, in general, will not be the case), we cannot state in general that learning is worse under $\mathscr{C}$. The next example illustrates this.

Consider a tiny network composed of four individuals ($n = 4$), and assume, for this exercise, that the same feed length is the same for both algorithms, $k = 3$ (this will not happen in general, as the closest algorithm will provide a longer feed, but we want to show that even in this case the reverse-chronological algorithm does

not guarantee better learning). Assume that the distribution of signals, conditional on θ, is as follows:

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

The closest algorithm induces, for user 1, a feed given by $\mathscr{C}_1^k = \{1,2,3\}$. Assume that a particular realization of the reverse-chronological algorithm induces the feed $\mathscr{R}_1^k = \{1,3,4\}$. Posterior variances are $\text{Var}[\theta|\{\theta_1,\theta_2,\theta_3\}] = 0.58$ for the closest algorithm and $\text{Var}[\theta|\{\theta_1,\theta_3,\theta_4\}] = 0.68$ for the reverse-chronological algorithm. Surprisingly, $\mathscr{C}$ yields better learning. The covariance to user 1 is not the unique driving force; the correlations between other users in the feed also play a role. In fact, the tension between different forces provokes that learning is not monotonic under the closest algorithm, as we can observe in Figure 1.4.1. However, when platform size grows, the similarities to user 1 become the dominant factor, and as a consequence, the closest algorithm performs worse than the reverse-chronological algorithm. This can also be observed in Figure 1.4.1, where we plot realizations of learning under $\mathscr{R}$ and $\mathscr{C}$ as n increases for a population growing from $n = 30$ to $n = 5000$, constant engagement $k = 30$ and parameters $\lambda = 0.5$ and $\beta = 0.2$.

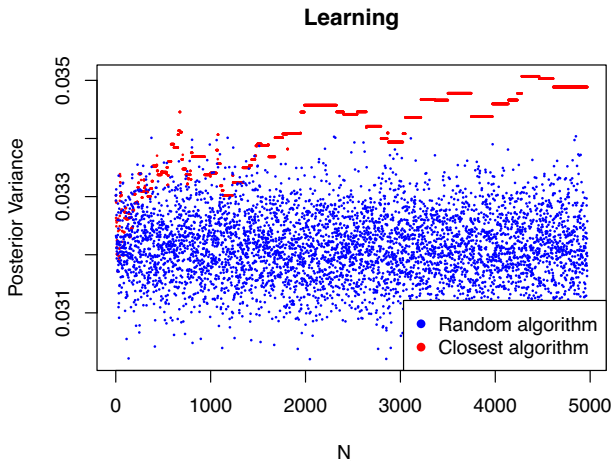


Figure 1.4.1. User's posterior variance as population grows; engagement is fixed to $k = 30$.

The second effect regarding learning when the user base is small relies on the fact that $\mathscr{C}$ induces higher engagement. Reading more posts is weakly better in learning terms, so even if we might intuitively think that learning is

worse under $\mathscr{C}$ because like-minded individuals are brought to the feed, this might be counterbalanced by the greater number of messages to learn from. The example above illustrate this case, too. Assume, as the simplest scenario, that $\mathscr{C}$ induces engagement $e_1^{\mathscr{C}} = 2$ for user 1 and hence shows the posts of users 2 and 3, while $\mathscr{R}$ induces engagement $e_1^{\mathscr{R}} = 1$ and shows the message of user 4. The posterior variances are, respectively, $\text{Var}[\theta|\{\theta_1, \theta_2, \theta_3\}] = 0.58$ and $\text{Var}[\theta|\{\theta_1, \theta_4\}] = 0.75$. Note that the extra message user 1 reads under $\mathscr{C}$ is key, as otherwise $\text{Var}[\theta|\{\theta_1, \theta_2\}] = 0.9$.

The reverse-chronological algorithm might be an alternative to enhance learning, but it does not seem realistic that bringing it back to social media platforms would work: most users are better off under the closest algorithm, even if platform size is large. While the goal of the DSA is to target the harms coming from personalization algorithms (mainly, as we show in Section 1.3, the excessive similarity in the feeds), it does not seem realistic to think of users going back to the reverse-chronological algorithm by themselves.[15] And this can be understandable: personalization algorithms had the objective of maximizing engagament and they certainly succeded (Guess et al., 2023). The platform-optimal algorithm is a sophisticated tool designed to please the user.

Another alternative worth exploring is to offer small modifications of the platform-optimal algorithm. Recently, platforms like X or Facebook have been adding features that "give context" or "promote fact-checked content". While keeping their personalized feeds, they sometimes incorporate a sponsored message, trying to improve users' information. Next, we study how such a modified platform-optimal algorithm might work in this model. We create the breaking echo-chambers algorithm, $\mathscr{B}$, by simply adding a user with opposite views to the closest feed. Formally, for every $i \in \mathscr{U}$, $\mathscr{B}_i(k) = \mathscr{C}_i(k-1)$, and $\mathscr{B}_i(1)$ is precisely the user with the highest negative correlation to user i.

The next result shows that, when platform size grows large, $\mathscr{B}$ allows the user to correctly learn the state of the world, maximizing learning at no cost in conformity. Remember that user i's expected conformity is $\sigma^2 - \frac{\sigma_{ij}^2}{\sigma^2}$. Hence, she is indifferent between a covariance of $\sigma_{ij} = -\sigma^2$ or $\sigma_{ij} = \sigma^2$, so, asymptotically, the breaking echo chambers algorithm incurs no penalty when maximizing learning. Because conformity and learning are simultaneously maximized, this algorithm converges to a utilitarian optimal algorithm. Note that, in contrast, the finite case is ambiguous: if there is a *very* opposite user in the pool, neither conformity

15. This becomes even more complicated when platforms, like Instagram currently, make the button for accessing a reverse-chronological feed difficult to find and provide the personalized feed by default each time a user logs in.

nor engagement will be that harmed and learning will be significantly improved. However, it might be the case that no such user is available, conformity and engagement are punished, and even though there is an improvement in learning, the platform-optimal algorithm provides higher utility.

In fact, when platform size is large, the breaking echo chambers algorithm has an effect on users similar to that of the plataform aggregating information and displaying it publicly. When platform size is not that large, the ability of the breaking echo chambers algorithm to achieve perfect learning should be at least questioned. Summarizing: when platform size is large, the breaking echo chambers algorithm works as a utilitarian optimal one, but it is not the case when platform size is finite. Hence, it is still a must to analyze how to achieve the implementation of the utilitarian optimal algorithm in general.

Proposition 1.4.2. *When platform size grows large, the breaking echo chambers algorithm outperforms the closest algorithm and converges to a utilitarian optimal algorithm.*

Proof. See Appendix 2.A. □

Platforms have already implemented algorithm modifications that promote content intended to improve user information: for example, in 2021, Twitter (now X) launched "Birdwatch", which became widespread in 2023, a feature where contributors could give context under a post. As in our model we do not allow the platform to know the messages of the users, we cannot build a similar feature in which the platform could aggregate them to obtain (and potentially share) an estimator for θ, but this is what the breaking echo chambers algorithm does in practice when platform size grows large. The purpose is the same: each user would read (and learn) the state of the world and at the same time derive some instantaneous utility from interacting with friends.

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

So far, we have explored the current institutional alternative to the platform-optimal algorithm, the reverse-chronological algorithm, and also an artificial improvement to the platform-optimal algorithm, the breaking echo-chambers algorithm. However, none of these alternatives are fully satisfactory, as either their performance or their viability is questionable. There is, however, one alternative

we have not yet explored: the utilitarian optimal algorithm. This algorithm is characterized by maximizing social welfare, and, by Lemma 1.3.2, this is equivalent to maximizing the expected utility of each user. Nonetheless, we cannot provide a closed-form expression for this algorithm. We do, however, offer an example that can be found in Appendix 1.B. We will denote the user-optimal algorithm as $\mathscr{U}$, and the next section is dedicated to exploring how, under the imposition of horizontal interoperability in a competitive market, platforms are compelled to implement it.

1.5 The need for horizontal interoperability

So far, we worked in our baseline model where a monopolist platform caters to a pool of n users who are already on board. In this scenario, the platform is not concerned about user capture but focuses solely on maximizing the time users spend on the platform—their engagement. This mirrors the current landscape of social media platforms. Large platforms like Instagram, TikTok, or X operate as monopolists within their specific niches: for instance, if someone wants to join a community for sharing pictures with friends, she would likely choose Instagram. While other sites may exist, the critical factor is that her friends are on Instagram. Network effects—a key feature of social media platforms wherein the platform's value increases as more users join and engage—protect these large incumbents. Consequently, platforms have strong incentives to grow their user base to offer greater network benefits than their rivals. This creates a high barrier to entry for new competitors, who must offer a vastly superior service to overcome the network effects and attract users.[16]

Network effects are particularly significant when it comes to algorithms, as they heavily rely on platform size.[17] The larger the network, the more possibilities for optimizing feeds and, eventually, the higher the expected utility for users. This is evident for the user-optimal algorithm: since platform and user incentives are aligned, a larger pool from which the platform can curate a feed translates to higher expected utility. However, the situation is more nuanced for the closest algorithm, as two opposing forces come into play when the platform size increases. On one hand, within-the-platform utility increases due to better matching possibilities. On the other hand, learning might decrease (we know that learning does not behave monotonically for small size increases, but that it asymptotically vanishes).

16. This was the case, for example, when Facebook was launched and then replaced MySpace as the leading social networking site in 2009.

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

Intuitively, the strategic role of the platform means the first force should dominate: the feed is chosen to maximize within-the-platform utility, with the effects on learning being a secondary consequence.

The next results provides a necessary and sufficient condition on the parameter λ for the closest algorithm to feature network effects. Before presenting it, let us introduce slight changes in notation. Let us refer to $U_i^n(.)$ to user i's utility when platform size is n, and similarities are captured by $\boldsymbol{\Sigma}$. Then, $U_i^{n+1}(.)$ refers to user i's utility function when platform size has grown to $n+1$, and similarities are captured by the extension of $\boldsymbol{\Sigma}$ as described in Section 1.4. Moreover, we denote e_i user i's engagement when platform size is n, and $\tilde{e}_i$ user i's engagement when platform size is $n+1$. Formally, we say that an algorithm features network effects if and only if the expected utility the user derives from joining the platform increases with platform size, i.e., $\mathbb{E}_i[U_i^{n+1}(\tilde{e}_i, \theta_i, \theta_{-i}, \mathscr{F}_i)] \geq \mathbb{E}_i[U_i^n(e_i, \theta_i, \theta_{-i}, \mathscr{F}_i)]$ for all n. From then on, we will work under the case of $\mathscr{C}$ featuring network effects, as it is the standard in this literature.

Proposition 1.5.1. *Denoting by $\mathscr{C}(n)$ the closest algorithm applied to platform size n, and by $\Delta\text{Var}[\theta|\boldsymbol{\theta}_{\mathscr{C}}] = \mathbb{E}_i[\text{Var}[\theta|\boldsymbol{\theta}_{\mathscr{C}(n+1)}] - \text{Var}[\theta|\boldsymbol{\theta}_{\mathscr{C}(n)}]]$ the expected difference in learning when the platform size increases from n users to $n+1$, we have that the closest algorithm features network effects if and only if*

$$\lambda \geq \frac{1}{1 + \frac{1}{\Delta\text{Var}[\theta|\boldsymbol{\theta}_{\mathscr{C}}]}\left(\alpha\,\mathbb{E}_i[\tilde{e}_i - e_i] + (1-\beta)\mathbb{E}_i[\,\nu(n, \tilde{e}_i, e_i)]\right)},$$

where $\nu(n, \tilde{e}_i, e_i) = \frac{2\tilde{e}_i(\tilde{e}_i - e_i)(3+2(e_i+\tilde{e}_i))+6(e_i-\tilde{e}_i)\tilde{e}_i n+3n^2}{3e_i(2+n)^2} > 0.$

Proof. See Appendix 2.A. □

Network effects are, therefore, platform-specific and proprietary. A platform with a small user base will provide low expected utility to its users, even if implementing the user-optimal algorithm $\mathscr{U}$. Consequently, users gravitate towards the large incumbent, causing the market to tip in its favor. The incumbent platform has no incentives to deviate from the platform-optimal algorithm, effectively trapping users. This is where the need for horizontal interoperability in social media platforms becomes apparent. Horizontal interoperability would enable a user from platform A to see posts from friends on platform B and vice versa. In other words, the algorithm implemented by platform A could match users from A with those from B, while also accessing their previous posts and interactions.[18] Some indus-

18. Although we consider a complete network in this chapter, the results apply to general networks and are more relevant in this section. With interoperability, users can maintain their neighborhood regardless of which platform each friend is a member of. This is similar to the mobile phone industry, where the focus is on whether a friend has a mobile phone, not the company providing the service.

tries, such as the cell and email industries, have already become interoperable: for example, a Yahoo user can send an email to a Gmail user seamlessly.

Although horizontal interoperability is a measure potentially applicable in many markets that feature network effects, it is particularly beneficial here for two main reasons. First, the implementation of interoperability removes entry barriers created by network effects, shifting them from the platform level to the market level and distributing them among all market players. This levels the playing field, increasing competition and contestability (Crémer, Rey, and Tirole, 2000; Kades and Scott Morton, 2020). This argument is applicable to almost any market with network effects but may not hold in markets with few non-interoperable features. For example, if messaging apps like WhatsApp or Telegram were mandated to become interoperable, users not concerned about privacy would have little reason to switch from WhatsApp. Even if switching costs are low, the lack of significant non-interoperable features in messaging apps means users would likely remain with the monopolist, WhatsApp. In social media platforms, however, algorithms are a key non-interoperable feature: while platform A implements $\mathscr{C}$, platform B could implement $\mathscr{U}$. Thus, as network effects are shared, platforms must compete at the algorithm level. This is the second and crucial reason for implementing horizontal interoperability. The primary way a platform can differentiate itself is through its personalized feed algorithm. Without platform-specific network effects, users can freely choose the feed that offers the best expected utility.

In a simplified setting where interoperability eliminates the incumbent's advantage from network effects, platforms are compelled to implement the user-optimal algorithm $\mathscr{U}$. Otherwise, users will migrate to a competitor implementing it. This argument is key: horizontal interoperability would naturally induce platforms to adopt the utilitarian algorithm. The following part of this section discusses the potential benefits and weaknesses of horizontal interoperability in social media platforms, its implementation challenges, and its current status in European legislation, particularly regarding the DMA.

Apart from the benefits already outlined, horizontal interoperability makes network effects a public good and then induces competition in all dimensions of non-interoperable features. Following our example, if two platforms were to implement the user-optimal algorithm, they would be equally attractive to a potential user. However, they could still compete in other dimensions such as service quality, user interface quality, or privacy and security. Hence, interoperability induces innovation in the non-interoperable features. By eliminating entry barriers generated by network effects, and given that entry cost is relatively low in social media, market contestability is also enhanced. Quite intuitively, large platforms will oppose interoperability: it 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).[19]

The main weakness of horizontal interoperability in social media platforms is the challenge its implementation constitutes, both in practical terms and regarding the consequences for privacy. While it does not seem too complicated to develop an standard of the basic features (see Kades and Scott Morton (2020) for an overview on standarization), platforms would need to share private data. This includes not only the messages that their users post (which in most cases are public), but also individual-level data regarding their interactions, as this allows for the calculation of similarities. Opening such data flows to third parties will raise privacy and security concerns.[20] Moreover, interoperability poses a challenge in services that promise end-to-end encryption. Cryptographers widely agree that maintaining encryption between different apps may prove challenging, if not impossible.

Aiming at "preventing gatekeepers from imposing unfair conditions on business and end users and at ensuring the openess of important digital services",[21] the European Comission has introduced interoperability as a regulatory measure in the European Union through the Digital Markets Act (DMA), passed in July 2022. Under this act, "gatekeeper" platforms and services are mandated to provide interoperability for chats with users on other services.[22] However, despite horizontal interoperability gaining traction as a regulatory measure in the EU, its actual implementation in social media platforms remains distant, as currently only messaging services are addressed.[23]

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

20. For the interested reader, we refer to Bourreau and Krämer (2022) for a detailed discussion of privacy and security risks of interoperability in digital markets.

21. 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.

22. 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. They are Alphabet, Amazon, Apple, ByteDance, Meta and Microsoft.

23. For the interested reader, we refer to Bourreau and Krämer (2023) for an overview on horizontal and vertical interoperability in the DMA.

1.6 Conclusion

We have developed a model of communication and learning through a personalized feed. An engagement-maximizing platform excessively weights conformity when designing feeds, aligning with existing evidence on echo chambers and filter bubbles (Pariser, 2011). The platform overemphasizes users' desire for conformity, resulting in severely impaired learning. Pariser argues that individualized personalization through algorithmic filtering could lead to intellectual isolation and social fragmentation as the product of being surrounded only by like-minded individuals. Our chapter theoretically demonstrates that this is the price to pay, as Guess et al. (2023) show empirically, when platforms are free to manage information exchanges to maximize profits and, hence, engagement. Institutional efforts to improve this situation have relied on the reverse-chronological algorithm, but our analysis suggests that it may not be sufficient. Users enjoy receiving recommended content, and while a random selection might enhance learning, the associated disutility may outweigh the benefits. Additionally, the likelihood of users disconnecting prematurely increases, meaning that even if diverse content enhances learning, users do not consume enough of this content.

The breaking echo chambers algorithm is a promising alternative when the platform size is large, which is the case for most social media platforms today. However, its practical implementation may be challenging. We propose horizontal interoperability as a solution, arguing that it is not just a *silver bullet* but a highly advantageous measure for the market we are analyzing. Algorithms are a non-interoperable feature of social media platforms, and the primary way platforms differentiate themselves from competitors is by finding the best algorithm for users. Competition would naturally lead to the implementation of healthier algorithms that fulfill users' desire for conformity while significantly enhancing learning.

Further research avenues emerge from this work. Beyond addressing the technical difficulties in the proper priors version of this model, we aim to explore how the algorithms we study affect *polarization*, defined as the sum of the squares of the differences between each user's beliefs about θ and the average belief. Additionally, we plan to further analyze horizontal interoperability. Interoperability might offer broader benefits than those discussed in this chapter. For example, Farronato, Fong, and Fradkin (2024) 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 the market tips. In principle, interoperability might resolve this issue: network effects would occur at the market level, maximizing them, while platform differentiation would still exist. Analyzing the effects of interoperability in a dynamic setting of competing

platforms where heterogeneous users can multi-home is a natural extension of this work. 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.

Appendix 1.A Omitted proofs

Proof of Proposition 1.3.1

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

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

This is equivalent to maximizing

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

The first order condition with respect to m_i yields

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

As this holds for all $j\in\mathscr{U}$, we substitute in this expression $m_j(\theta_j) = \beta\theta_j + (1-\beta)\frac{1}{e_j}\sum_{l\in\mathscr{F}_j^{e_j}}\mathbb{E}_j\left[m_l(\theta_l)|\theta_j,\mathscr{F}\right]$ for all $j\in\mathscr{F}_i^{e_i}$, and then we repeat the procedure for all $l\in\mathscr{F}_j^{e_j}$ and so on. Users' knowledge of $\mathscr{F}$ and Σ is crucial at this point, allowing us to commute the sum and the expectation operators. We can iterate on this procedure as many times as desired:[24]

$$\begin{aligned}
m_i &= \beta\theta_i + (1-\beta)\frac{1}{e_i}\sum_{j\in\mathscr{F}_i^{e_i}}\mathbb{E}_i\left[m_j(\theta_j)|\theta_i,\mathscr{F}\right]\\
&= \beta\theta_i + (1-\beta)\frac{1}{e_i}\sum_{j\in\mathscr{F}_i^{e_i}}\mathbb{E}_i\left[\beta\theta_j + (1-\beta)\frac{1}{e_j}\sum_{l\in\mathscr{F}_l^{e_j}}\mathbb{E}_j[m_l(\theta_l)|\theta_j,\mathscr{F}]\right]\\
&= \beta\theta_i + (1-\beta)\beta\theta_i + \frac{(1-\beta)^2}{e_ie_j}\sum_{j\in\mathscr{F}_i^{e_i}}\left(\sum_{l\in\mathscr{F}_l^{e_j}}\mathbb{E}_i[\mathbb{E}_j[m_l(\theta_l)|\theta_j,\mathscr{F}]|\theta_i,\mathscr{F}]\right) = \dots\\
&= \beta\theta_i\sum_{r=0}^{m}(1-\beta)^r + \frac{(1-\beta)^m}{e_ie_j\dots e_m}\sum_{j\in\mathscr{F}_i^{e_i}}\left(\dots\left(\sum_{p\in\mathscr{F}_m^{e_m}}\mathbb{E}_i[\dots[\mathbb{E}_m[m_p(\theta_p)|\theta_m,\mathscr{F}]\dots]|\theta_i,\mathscr{F}]\right)\dots\right).
\end{aligned} \tag{1.A.2}$$

24. Abusing notation, we iterate m times and also refer to the m-th user as m.

This expression holds for all $m \in \mathbb{N}$, so we can take limits when $m \to \infty$. On the one hand, $\lim_{m\to\infty} \sum_{r=0}^{m} (1-\beta)^r = \frac{1}{\beta}$, and, hence, the first term in Equation (1.A.2) is simply θ_i. On the other hand, the second term vanishes as $m \to \infty$. Hence, $m_i^* = \theta_i$ for all $i \in \mathscr{U}$ and we have truthtelling for any algorithm $\mathscr{F}$ and any engagement levels $\{e_i\}_{i\in\mathscr{U}}$. □

Proof of Proposition 1.3.3

Proof. The probability that, under algorithm $\mathscr{F}$, user i stays for one more period after staying for k is given by $g(u_i(k, \mathscr{F}))$. To maximize such probability, the platform chooses $\mathscr{F}_i(k) = \arg\max_{j\in\mathscr{U}\setminus\mathscr{F}_i^{k-1}} \{\mathbb{E}\left(g(u_i(k, \mathscr{F}))\right)\}$. As g is strictly increasing on u_i and the expectation preserves the order, this is equivalent to maximizing user i's expected inside-the-platform utility.

Given truthful reporting and noting that $v(.)$ is independent of the algorithm $\mathscr{F}$, we can write the platform's objective as finding the user $j \in \mathscr{U}\setminus\mathscr{F}_i^{k-1}$ that maximizes $-\mathbb{E}_p\left(\sum_{l\in\mathscr{F}_i^{k-1}} (\theta_i - \theta_l)^2 + (\theta_i - \theta_j)^2\right)$ or simply $-\mathbb{E}_p\left(\theta_i - \theta_j)^2\right)$. Thus, maximizing the probability of user i staying for one more period is equivalent to minimizing the conformity cost of such period.

Let us prove next that the algorithm that maximizes expected engagement is the same that maximizes within-the-platform expected utility. Given algorithm $\mathscr{F}$, the probability of staying at least until period e_i is $\Pi_{j=1}^{e_i} g(u_i(j, \mathscr{F}))$, and the probability of staying precisely until period e_i is then

$$\Pi_{j=1}^{e_i} g(u_i(j, \mathscr{F}))\left(1 - \Pi_{j=e_i+1}^{n} g(u_i(j, \mathscr{F}))\right).$$

Now, let us take two feeds, namely $\mathscr{F}_i$ and $\mathscr{F}_i'$, such that they are identical except from two users that are interchanged, i.e., there are users t and t' such that

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

Moreover, they satisfy $-\mathbb{E}_p\left((\theta_i - \theta_t)^2\right) > -\mathbb{E}_p\left((\theta_i - \theta_t')^2\right)$. All this means that in the feed $\mathscr{F}_i$, the user who penalizes conformity the least is shown before. Without loss of generality we can assume that $t = 1$ and $t' = 2$, and then our goal is to show that such feed $\mathscr{F}_i$ yields higher expected engagement. Notice, first, that $g(u_i(1, \mathscr{F})) > g(u_i(1, \mathscr{F}'))$ because conformity is higher. Expected engagement under $\mathscr{F}$ reads as

$$g(1, \mathscr{F})\left[\left(1 - \Pi_{j=2}^{n} g(u_i(j, \mathscr{F}))\right) + g(u_i(2, \mathscr{F}))\left(2(1 - \Pi_{j=3}^{n} g(u_i(j, \mathscr{F})) + 3...)\right)\right].$$

Given that $g(u_i(j,\mathscr{F})) = g(u_i(j,\mathscr{F}'))$ for $j > 2$, we define c, which has the same value for both algorithms, as $c := \left(2(1 - \Pi_{j=3}^{n} g(U_i(j,\mathscr{F}_i)) + 3...)\right)$ to ease notation. Then, given that $v(e_i) = \alpha e_i$, $g(u_i(1,\mathscr{F})) = g(u_i(2,\mathscr{F}'))$ and $g(u_i(1,\mathscr{F}')) = g(u_i(2,\mathscr{F}))$, we have that

$$\begin{aligned}\mathbb{E}_p\left(e_i \mid \mathscr{F}\right) &= g(1,\mathscr{F})\left[\left(1 - \Pi_{j=2}^{n} g(u_i(j,\mathscr{F}))\right) + g(u_i(2,\mathscr{F}))C\right] \\ &\geq g(1,\mathscr{F}')\left[\left(1 - \Pi_{j=2}^{n} g(u_i(j,\mathscr{F}'))\right) + g(u_i(2,\mathscr{F}'))C\right] = \mathbb{E}_p\left(e_i \mid \mathscr{F}'\right).\end{aligned}$$

This shows that any feed that is not reverse-ordered following the expected loss in conformity $\mathbb{E}_p((\theta_i - \theta_j)^2)$ is always dominated. □

Lemma 1.A.1. *The posterior distribution of θ conditional on $\boldsymbol{\theta}_{\mathscr{F}_i^{e_i}}$ is given by*

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

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

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

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

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

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

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

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

as we wanted to show. □

Proof of Proposition 1.3.7

Proof. By assumption, $g(.) \in (0,1)$, so that even if there is no penalty in conformity and within-the-platform utility is just given by $v(e_i)$, user i's engagement is a finite number. Let us call such number k. 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 $j_1, ..., j_k$ such that $\rho_{i,j_r} > 1-\varepsilon$ for all $r \in \{1, ..., k\}$. On the other hand, applying the Cauchy-Schwarz inequality to the correlations between the pairs formed by user i and two other users, say j_r and j_l, we get

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

Using the ε-bounds derived above, we obtain:

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

Now, assume engagement is e_i. As $e_i \leq k$, users from $\mathscr{C}_i^{e_i} \subset \mathscr{U}$ are taken from the set of k users specified above. Let us now 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 $\mathscr{C}_i^{e_i}$ verifies that if $j_r, j_l \in \mathscr{C}_i^{e_i}$,[25] $\rho_{j_r,j_l} > 1-\delta$ (it is enough to choose ε accordingly). Hence, we have that for the matrix $\mathbf{A}$ defined as

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

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

Lemma 1.A.2. *In this particular case, $\boldsymbol{A} \leq \boldsymbol{\Sigma}_{\mathscr{C}_i^{e_i}}$ implies $\boldsymbol{\Sigma}^{-1}_{\mathscr{C}_i^{e_i}} \leq \boldsymbol{A}^{-1}$.*

25. Here we abuse notation slightly, as $\mathscr{U}$ should be $\mathscr{U}(n)$ and $\mathscr{C}_i^{e_i}$ should be $\mathscr{C}_i^{e_i}(.,n)$.

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

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

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:

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

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

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

This implies that $\bar{a}_{11} \leq \alpha$ and $\sum_{j=2}^{e_i} \bar{a}_{1j} \leq \alpha(e_i - 1)\bar{b}$. 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,

$$\mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{C}_i^{e_i}}\mathbb{1}^t \leq \mathbb{1}\mathbf{A}^{-1}\mathbb{1}^t \Rightarrow \frac{1}{\mathbb{1}\mathbf{A}^{-1}\mathbb{1}^t} \leq \frac{1}{\mathbb{1}\boldsymbol{\Sigma}^{-1}_{\mathscr{C}_i^{e_i}}\mathbb{1}^t} \Rightarrow \frac{1}{\mathbb{1}\mathbf{A}^{-1}\mathbb{1}^t} \leq \text{Var}[\theta|\boldsymbol{\theta}_{\mathscr{C}_i^{e_i}}].$$

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

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

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 $\text{Var}[\theta|\boldsymbol{\theta}_{\mathscr{C}_i^{e_i}}] = \sigma^2$.

□

Proof of Proposition 1.4.2

Proof. The following proof consists of two parts. First, we will show that conformity is at least as good under $\mathscr{B}$ than under $\mathscr{C}$. Second, we will show that, asymptotically, learning is perfect under $\mathscr{B}$.

Note that, by assumption, $g(.) \in (0,1)$, so that even if there is no penalty in conformity and within-the-platform utility is just given by $v(e_i)$, user i's engagement is a finite number. Let us call such number k. Now, following the reasoning in the proof of Proposition 1.3.7, for every $\varepsilon > 0$ there exists a $\bar{n}(\varepsilon) \in \mathbb{N}$ such that for every $n > \bar{n}(\varepsilon)$, there exists a set of k users such that $\rho_{ij} \geq 1-\varepsilon$ for every $j = 1,...,k$. Moreover, for every $\delta > 0$, there exists a $\tilde{n}(\delta) \in \mathbb{N}$ such that for every $n > \tilde{n}(\delta)$ there is a user l such that $\rho_{il} < \delta - 1$. Let us now take $n \geq \max\{\bar{n}(\varepsilon), \tilde{n}(\delta)\}$ and define, for any engagement e_i, $\mathscr{B}_i^{e_i} = \{l, 1, ..., e_i - 1\}$, where users in $\{1,...,e_i-1\}$ are taken from the pool of size k obtained before. Then, as $n \to \infty$, we will have that $\rho_{ij} \to 1$ for all $j \in \{1,...,k\}$ and $\rho_{il} \to -1$. Then, expected conformity under the breaking echo chambers algorithm becomes

$$\mathbb{E}_i\left(\sum_{j \in \mathscr{B}_i^{k+1}} (\theta_i - \theta_j)^2 \;\middle|\; \theta_i\right) = \sigma^2 \sum_{j=1}^{k+1}(1-\rho_{ij}^2) + \sigma^2(1-\rho_{il}^2),$$

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

Now, let us study learning. First, we analyze what user i learns from user l's message:

$$\text{Var}[\theta|\theta_i, \theta_l] = \frac{1}{\mathbb{1}\boldsymbol{\Sigma}^{-1}_{il}\mathbb{1}^T} = \frac{\delta(2-\delta)}{4-\delta},$$

which coverges to zero as n grows large. Note that, as $l \in \mathscr{B}_i^{e_i}$,

$$\mathrm{Var}[\theta|\theta_i, \theta_l] \geq \mathrm{Var}[\theta|\theta_{\mathscr{B}_i^{e_i}}] \geq 0,$$

so $\lim_{n\to\infty} \mathrm{Var}[\theta|\theta_{\mathscr{B}_i^{e_i}}] = 0$ and there is perfect learning under the breaking echo chambers algorithm. □

Proof of Proposition 1.5.1

Proof. Again, we assume that covariances are drawn from a uniform distribution $\mathscr{U}[-\sigma^2, \sigma^2]$. The platform matches the user with those featuring the highest covariances to her, and then, in terms of within-the-platform utility it means that we have to compute

$$\begin{aligned}\mathbb{E}_i[u_i^n(e_i, \theta_i, \theta_{-i}, \mathscr{C})] =& \lambda\left(v(e_i) - (1-\beta)\frac{1}{e_i}\left(e_i\sigma^2 - \frac{1}{\sigma^2}\sum_{j\in\mathscr{C}_i^{e_i}}\sigma_{ij}^2\right)\right)\\ =& \lambda\left(v(e_i) - (1-\beta)\left(\sigma^2 - \frac{1}{e_i}\sum_{j=n-e_i+1}^{n}\left(\frac{4j^2}{(n+1)^2} - \frac{4j}{n+1} + 1\right)\right)\right).\end{aligned}$$

Hence, overall user i's expected utility is given by

$$\begin{aligned}\mathbb{E}_i[U_i^n(e_i, \theta_i, \theta_{-i}, \mathscr{C})] =& \lambda\left(v(e_i) - (1-\beta)\left(\sigma^2 - \frac{1}{e_i}\sum_{j=n-e_i+1}^{n}\left(\frac{4j^2}{(n+1)^2} - \frac{4j}{n+1} + 1\right)\right)\right)\\ &- (1-\lambda)\mathbb{E}_i[\mathrm{Var}[\theta|\boldsymbol{\theta}_{\mathscr{C}}]],\end{aligned}$$

and a simple rearrangement of the expression $\mathbb{E}_i[U_i^{n+1}(\tilde{e}_i, \theta_i, \theta_{-i}, \mathscr{C})] - \mathbb{E}_i[U_i^n(e_i, \theta_i, \theta_{-i}, \mathscr{C})]$ yields the desired inequality. □

Appendix 1.B Example

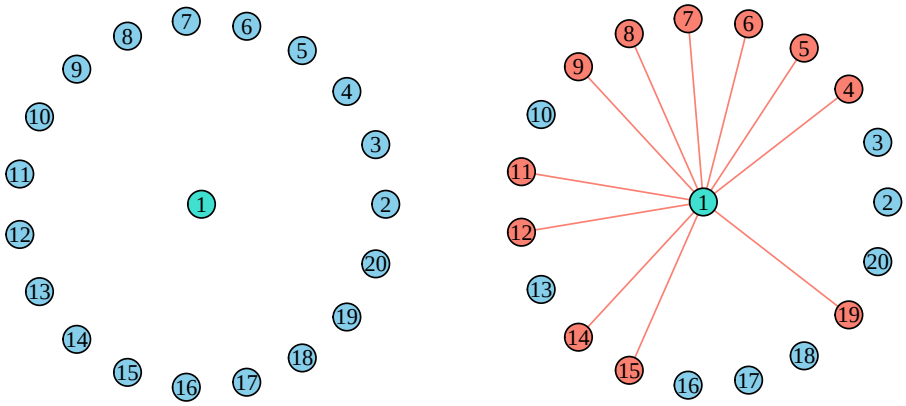
Here we present the feeds user 1 would observe in a platform of size $n = 20$ (Figure 1.B.1a) with similarity matrix $\boldsymbol{\Sigma}$ as displayed below. We fix parameters to $\alpha = 0.001$, $\lambda = 0.5$ and $\beta = 0.2$.

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

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

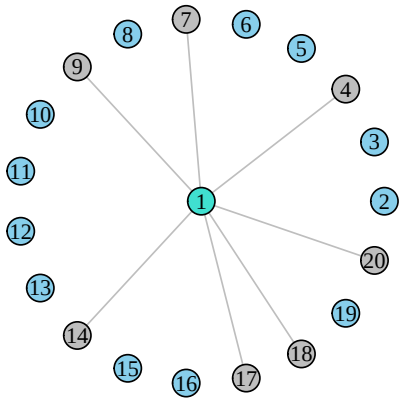
The platform-optimal or closest algorithm, $\mathscr{P}$, ranks users according to their covariances to user i. The ranking is 19, 4, 11, 5, 12, 7, 14, 9, 6, 8, 15, 10, 18, 13, 3, 16, 20, 2, and 17. For the specific configuration of this example, the expected

engagement can be calculated to 10.8 (approximated to 11), and hence user i will learn the messages of the first 11 users in the ranking. We represent this in Figure 1.B.1b, where user 1 is linked to those whose messages will be read. In turn, the user-optimal algorithm, $\mathscr{U}$, ranks users according to their overall contribution to user i's utility. The ranking is 19, 7, 4, 5, 14, 11, 9, 12, 6, 8, 15, 3, 20, 18, 10, 16, 13, 2. The expected engagement is 10.4 (approximated to 10), and hence user i observes the messages of the first 10 users in such ranking, as represented in Figure 1.B.1d. Crucially, even though the order provided by each of these two algorithms is different, the set of users appearing in the realized feeds is almost the same (note that the only difference is that the feed under $\mathscr{P}$ includes user 15). Finally, the reverse-chronological algorithm $\mathscr{R}$ randomly ranks users as 14, 20, 4, 17, 18, 9, 7, 15, 6, 16, 11, 19, 12, 10, 3, 13, 2, 8, 5, yields expected engagement 6.8 (approximated to 7), and induces a feed represented in Figure 1.B.1c.

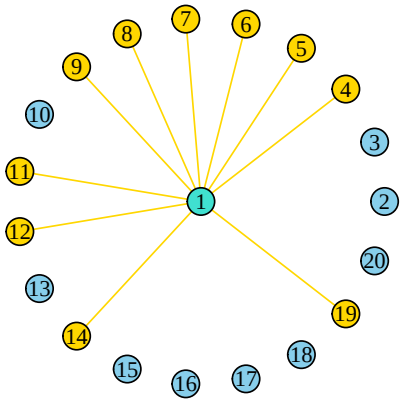


(a) Users in the platform.

(b) Platform-optimal feed.



(c) Reverse-chronological feed.



(d) User-optimal feed.

Figure 1.B.1. User 1's feeds in a platform of size $n = 20$.

References

Abreu, Luis, and Doh-Shin Jeon. 2019. "Homophily in social media and news polarization." working paper, available at *https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=3468416*, [10]

Acemoglu, Daron, and Asuman Ozdaglar. 2011. "Opinion dynamics and learning in social networks." *Dynamic Games and Applications* 1(1): 3–49. [11]

Acemoglu, Daron, Asuman Ozdaglar, and James Siderius. 2023. "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*, [9]

Agarwal, Saharsh, Uttara M. Ananthakrishnan, and Catherine E. Tucker. 2022. "Deplatforming and the control of misinformation: Evidence from Parler." working paper, available at *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4232871*, [10]

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social media and fake news in the 2016 election." *Journal of Economic Perspectives* 31(2): 211–36. [4]

Allcott, Hunt, Matthew Gentzkow, and Lena Song. 2022. "Digital addiction." *American Economic Review* 112(7): 2424–63. [4, 7, 15]

Belleflamme, Paul, and Martin Peitz. 2020. "The competitive impacts of exclusivity and price transparency in markets with digital platforms." *Concurrences*, (1): 2–12. [29]

Benhabib, Jess, and Alberto Bisin. 2005. "Modeling internal commitment mechanisms and self-control: A neuroeconomics approach to consumption–saving decisions." *Games and Economic Behavior* 52(2): 460–92. [7]

Benzell, Seth, and Avinash Collis. 2022. "Regulating digital platform monopolies: The case of Facebook." working paper, available at *https://www.aeaweb.org/conference/2023/program/paper/SNNd74ni*, [10]

Bernheim, B. Douglas. 1994. "A theory of conformity." *Journal of Political Economy* 102(5): 841–77. [6]

Biddle, Sam. 2018. "Facebook uses artificial intelligence to predict your future actions for advertisers, says confidential document." *Intercept* 13(04): 2018. [7]

Biglaiser, Gary, Jacques Crémer, and André Veiga. 2022. "Should I stay or should I go? Migrating away from an incumbent platform." *RAND Journal of Economics* 53(3): 453–83. [10]

Bonatti, Alessandro, and Gonzalo Cisternas. 2020. "Consumer scores and price discrimination." *Review of Economic Studies* 87(2): 750–91. [15]

Bourreau, Marc, and Jan Krämer. 2022. "Interoperability in digital markets." available at https://sitic.org/wordpress/wp-content/uploads/Interoperability-in-Digital-Markets.pdf, [29]

Bourreau, Marc, and Jan Krämer. 2023. "Horizontal and vertical interoperability in the DMA." available at https://cerre.eu/wp-content/uploads/2023/12/ISSUE-PAPER-CERRE-DEC23DMA-Horizontal-and-Vertical-Interoperability-Obligations.pdf, [29]

Bursztyn, Leonardo, Benjamin Handel, Rafael Jiménez-Durán, and Christopher Roth. 2023. "When product markets become collective traps: the case of social media." working paper, available at https://ssrn.com/abstract=4596071, [4, 15]

Chamley, Christophe. 2004. *Rational herds: Economic models of social learning*. Cambridge University Press. [6]

Chitra, Uthsav, and Christopher Musco. 2020. "Analyzing the impact of filter bubbles on social network polarization." In, 115–23. [10]

Cialdini, Robert B., and Noah J. Goldstein. 2004. "Social influence: Compliance and conformity." *Annual Review of Psychology* 55 (1): 591–621. [6]

Crémer, Jacques, Patrick Rey, and Jean Tirole. 2000. "Connectivity in the commercial Internet." *Journal of Industrial Economics* 48 (4): 433–72. [28]

Demange, Gabrielle. 2023. "Simple visibility design in network games." working paper, available at https://hal.science/hal-03925344v1/document, [10]

DeMarzo, Peter M, Dimitri Vayanos, and Jeffrey Zwiebel. 2003. "Persuasion bias, social influence, and unidimensional opinions." *Quarterly Journal of Economics* 118 (3): 909–68. [11]

Ellman, Matthew, and Fabrizio Germano. 2009. "What do the papers sell? A model of advertising and media bias." *Economic Journal* 119 (537): 680–704. [10]

Evens, Tom, Karen Donders, and Adelaida Afilipoaie. 2020. "Platform policies in the European Union: Competition and public interest in media markets." *Journal of Digital Media & Policy* 11 (3): 283–300. [10]

Farronato, Chiara, Jessica Fong, and Andrey Fradkin. 2024. "Dog eat dog: Balancing network effects and differentiation in a digital platform merger." *Management Science* 70 (1): 464–83. [30]

Franck, Jens-Uwe, and Martin Peitz. 2023. "Market power of digital platforms." *Oxford Review of Economic Policy* 39 (1): 34–46. [10, 14]

Galeotti, Andrea, Benjamin Golub, Sanjeev Goyal, and Rithvik Rao. 2021. "Discord and harmony in networks." *available at https://arxiv.org/pdf/2102.13309*, [6]

Golub, Benjamin, and Matthew O. Jackson. 2010. "Naive learning in social networks and the wisdom of crowds." *American Economic Journal: Microeconomics* 2 (1): 112–49. [8]

Grayot, James D. 2020. "Dual process theories in behavioral economics and neuroeconomics: A critical review." *Review of Philosophy and Psychology* 11 (1): 105–36. [7]

Greene, Steven. 2004. "Social identity theory and party identification." *Social science quarterly* 85 (1): 136–53. [15]

Guess, Andrew M., Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, et al. 2023. "How do social media feed algorithms affect attitudes and behavior in an election campaign?" *Science* 381 (6656): 398–404. [5, 9, 24, 30]

Guriev, Sergei, Emeric Henry, Théo Marquis, and Ekaterina Zhuravskaya. 2023. "Curtailing false news, amplifying truth." working paper, available at https://shs.hal.science/halshs-04315924/document, [15]

Habib, Hussam, Maaz Bin Musa, Fareed Zaffar, and Rishab Nithyanand. 2019. "To act or react: Investigating proactive strategies for online community moderation." *available at https://arxiv.org/pdf/1906.11932*, [11]

Hartigan, John A. 1983. *Bayes Theory*. Springer Science & Business Media. [12]

Horwitz, Jeff et al. 2021. "The Facebook files." *Wall Street Journal*, available online at: https://www.wsj.com/articles/the-facebook-files-11631713039, [4]

Hu, Lin, Anqi Li, and Xu Tan. 2021. "A rational inattention theory of echo chamber." working paper, available at https://arxiv.org/pdf/2104.10657.pdf, [10]

Hwang, Elina H., and Stephanie Lee. 2021. "A nudge to credible information as a countermeasure to misinformation: Evidence from twitter." working paper, available at https://ssrn.com/abstract=3928343, [11]

Jackson, Matthew O., Suraj Malladi, and David McAdams. 2022. "Learning through the grapevine and the impact of the breadth and depth of social networks." *Proceedings of the National Academy of Sciences* 119(34): [10]

Jadbabaie, Ali, Pooya Molavi, Alvaro Sandroni, and Alireza Tahbaz-Salehi. 2012. "Non-Bayesian social learning." *Games and Economic Behavior* 76(1): 210–25. [11]

Kades, Michael, and Fiona Scott Morton. 2020. "Interoperability as a competition remedy for digital networks." *Washington Center for Equitable Growth Working Paper Series*, [9, 10, 28, 29]

Kosinski, Michal, David Stillwell, and Thore Graepel. 2013. "Private traits and attributes are predictable from digital records of human behavior." *Proceedings of the National Academy of Sciences* 110(15): 5802–5. [7]

Kranton, Rachel, and David McAdams. 2022. "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, [10]

Lauer, David. 2021. "Facebook's ethical failures are not accidental; they are part of the business model." *AI and Ethics* 1(4): 395–403. [5]

Levy, Ro'ee. 2021. "Social media, news consumption, and polarization: Evidence from a field experiment." *American Economic Review* 111(3): 831–70. [10]

Loomba, Sahil, Alexandre de Figueiredo, Simon J. Piatek, Kristen de Graaf, and Heidi J. Larson. 2021. "Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA." *Nature Human Behaviour* 5(3): 337–48. [7]

Molavi, Pooya, Alireza Tahbaz-Salehi, and Ali Jadbabaie. 2018. "A theory of non-Bayesian social learning." *Econometrica* 86(2): 445–90. [11]

Mosleh, Mohsen, Cameron Martel, Dean Eckles, and David G. Rand. 2021. "Shared partisanship dramatically increases social tie formation in a Twitter field experiment." *Proceedings of the National Academy of Sciences* 118(7): e2022761118. [6]

Mostagir, Mohamed, and James Siderius. 2022. "Learning in a post-truth world." *Management Science* 68(4): 2860–68. [11]

Mostagir, Mohamed, and James Siderius. 2023a. "When do misinformation policies (not) work?" available online at: https://siderius.lids.mit.edu/how-do-we-stop-misinformation/, [11]

Mostagir, Mohamed, and James Siderius. 2023b. "When should platforms break echo chambers?" available online at: https://siderius.lids.mit.edu/wp-content/uploads/sites/36/2023/01/quarantine-v6.pdf, [11]

Mudambi, Maya, and Siva Viswanathan. 2022. "Prominence reduction versus banning: An empirical investigation of content moderation strategies in online platforms." working paper, available at https://core.ac.uk/download/pdf/542548963.pdf, [11]

Mueller-Frank, Manuel, and Claudia Neri. 2021. "A general analysis of boundedly rational learning in social networks." *Theoretical Economics* 16(1): 317–57. [11]

Mueller-Frank, Manuel, Mallesh M. Pai, Carlo Reggini, Alejandro Saporiti, and Luis Simantujak. 2022. "Strategic management of social information." working paper, available at *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4434685*, [9]

Pariser, Eli. 2011. *The filter bubble: How the new personalized web is changing what we read and how we think.* Penguin. [10, 30]

Pennycook, Gordon, and David G Rand. 2021. "The psychology of fake news." *Trends in Cognitive Sciences* 25(5): 388–402. [11]

Pennycook, Gordon, and David G. Rand. 2019. "Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning." *Cognition* 188: 39–50. [11]

Popiel, Pawel. 2020. "Addressing platform power: The politics of competition policy." *Journal of Digital Media & Policy* 11(3): 341–60. [10]

Reuter, Jonathan, and Eric Zitzewitz. 2006. "Do ads influence editors? Advertising and bias in the financial media." *Quarterly Journal of Economics* 121(1): 197–227. [10]

Sagioglou, Christina, and Tobias Greitemeyer. 2014. "Facebook's emotional consequences: Why Facebook causes a decrease in mood and why people still use it." *Computers in Human Behavior* 35: 359–63. [10]

Scott Morton, Fiona, Pascal Bouvier, Ariel Ezrachi, Bruno Jullien, Roberta Katz, Gene Kimmelman, A. Douglas Melamed, and Jamie Morgenstern. 2019. "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, [4]

Silverman, Craig. 2016. "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,): [4]

Solon, Olivia. 2016. "Facebook's failure: Did fake news and polarized politics get Trump elected?" *Guardian*, available at *https://www.theguardian.com/technology/2016/nov/10/facebook-fake-news-election-conspiracy-theories*, (November 10,): [4]

Sunstein, Cass R. 2017. "A prison of our own design: Divided democracy in the age of social media." *Democratic Audit UK*, [10]

Törnberg, Petter. 2018. "Echo chambers and viral misinformation: Modeling fake news as complex contagion." *PLoS one* 13 (9): [10]

Chapter 2

Based on the Papers You Liked: Designing a Rating System for Strategic Users

Joint with Jacopo Gambato

2.1 Introduction

Streaming platforms have become ubiquitous in today's digital landscape. Video streaming services such as Netflix, HBO, Amazon Prime Video, and Disney+ boast 1.8 billion users worldwide. Music streaming platforms like Spotify, Apple Music, and Deezer cater to approximately 524 million users. The global streaming market is currently valued at $84.3 billion. These platforms are characterized by two main features: they offer access to extensive libraries of content in exchange for a (monthly) fee, and employ recommendation algorithms that provide users with tailor-made suggestions. These algorithms enhance user experience, and are designed to draw in and retain subscribers. For instance, Netflix's recommendation system is responsible for 80% of the platform's streaming time. This system gained notoriety in 2006 when Netflix launched the "$1 million Netflix Prize", awarded to anyone who could improve the algorithm's performance by 10%. Recently, Net-

* I acknowledge financial support by the German Research Foundation (DFG) through CRC TR 224 (Project B05) and funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 949465). We are grateful to Sven Rady, Alexander Frug, Carl-Christian Groh and Manuel Lleonart-Anguix for their helpful comments and discussions.

flix has continued to innovate by adding a "double like" button to its "thumbs up/thumbs down" rating system.[1]

Users face asymmetric information problems when trying new products: while they might know which kind of content they want to consume, they may not be able to ascertain quality before consumption. Platforms leverage user-side network effects to mitigate this issue by using aggregate information disclosed by other users to offer tailored recommendations. This is true for both "horizontal" classification of content—its genre—and its "vertical" differentiation—its quality. What distinguishes rating systems from other certification systems is their costless nature: users employ other users' ratings to navigate the catalogue of platforms, and rate content to signal platforms which products to recommend. Platforms have an inherit incentive to allow users to rate the content that they experience because, on top of being free, this sort of certification is less susceptible to gaming compared to direct seller certification systems.[2]

Ratings interact ambiguously with the size a streaming platform. The larger the platform is, the more products are available for recommendation and, crucially, the more aggregated information can be used to generate recommendations. At a glance, because users value high-quality products and because the platform can monetize the quality of the matches she proposes, both users and platform should be interested in a design that generates the best possible recommendations. However, anecdotal evidence shows dominant platforms often selecting less granular rating systems. YouTube, for example, allows users to give thumbs up or thumbs down but only displays the thumbs up count. Spotify lets listeners like songs, and Netflix uses a double like/like/dislike system. A possible explanation is that more granular systems increase cognitive load and reduces engagement with the rating system.[3] This would justify the drastically different approach of more specialized portals, like IMDb or Filmaffinity, who use finer rating systems (e.g., a 0 to 10 Likert scale) anticipating that their users are typically more willing to spend time providing detailed ratings.

1. See Gómez-Uribe and Hunt (2015) for an (internal) overview of the Netflix recommendation system.

2. We argue that the subjectivity of ratings in modern streaming platform is a feature, not a bug: with enough users and sellers, the ratings are more useful as a way to aggregate users by taste to make recommendations more precise than as a way to randomly signal high quality of a product.

3. There is sparse evidence on the effect of the granularity of rating systems on engagement with system itself: Lafky (2014) shows that even a small cost of rating significantly decreases the willingness to rate in an experimental setting. Along these lines, Pommeranz, Broekens, Wiggers, Brinkman, and Jonker (2012) shows that users prefer feedback methods such that less effort is required.

Still, this narrative seems to be incomplete. In time, many platforms have adjusted their rating systems as they have grown. For example, YouTube shifted from a five-star system to a simpler like/dislike system in September 2009, and then again to the current, even less informative system in 2021. This is counter-intuitive: if platforms can better monetize their content with more precise recommendations, information extraction should be paramount. It would seem natural, then, to observe platforms offering *more granular* systems as they grow rather then the opposite. Furthermore, YouTube's 2021 design change is striking because it effectively takes away from users the ability to select high quality content based on others' ratings while, at the same time, maintaining their ability to signal their preferences to the platform for the recommendation system to use. This observation justifies the view that engagement with rating systems goes beyond simple pro-social behavior.[4] This chapter presents a theoretical model that aims at explaining why platforms facing strategic users move to less granular systems as they grow (or, as we show, as the cost of hosting sellers decreases).

We model ratings as thresholds that users select to distinguish products based on their quality. By selecting these thresholds, users disclose their taste and enable the platform to group them with similar other users, allowing for more accurate recommendations of new content. This process is a form of exploration: users consume and rate products and, if deemed good enough, the platform recommends them to others in their category. The main methodological novelty of the chapter is to build a mechanism that allows strategic users to disclose information on tastes to the platform, which then uses this information to provide them with recommendations. This structure significantly distances our chapter from much of the existing literature, which is generally focused on how an informed regulator (the platform) optimizes the disclosure mechanism. We argue that most streaming platforms do not have a credible way to learn and disclose content quality to users without users' rating, and that direct certification systems are more prone to gaming and, therefore, less effective at inducing consumption.

Our model begins with the simplest rating system: a user can like a product after consuming it, thus classifying content as acceptable or not. The user must decide where to set the threshold for acceptability. The platform then recommends only acceptable products (if available), which increases the users' expected utility of consumption and their willingness to pay for the subscription. We then add granularity by allowing users to set two thresholds: the platform attempts to recommend something of quality above the highest threshold, then between

4. Tadelis (2016) argues that prosocial behavior is prevalent on eBay, where approximately 65% of users leave feedback. However, users still engage with YouTube's rating system even though its most recent change made ratings much less informative for other users.

the highest and the lowest, and finally at random if no content is available above either thresholds. We apply these rating systems to a two-sided platform hosting both sellers (who provide one product each) and users (who receive recommendations). We distinguish two configurations. First, we assume that the platform represents a large enough share of the market that she hosts to affect competition intensity on the outside market. In this case, the platform selects the fixed upfront fee for sellers as well as the granularity of the rating system.[5] Second, we assume the platform to be a price-taker that only selects the rating system, and we distinguish between platforms with high and low "reach", which in turn determines her relative bargaining power *vis-à-vis* sellers.

When the platform strategically selects how much she is willing to remunerate sellers, she is able to control the number of users and sellers that join. The outside option for users is better the fewer sellers join, and *vice versa*. We assume users to be heterogeneous in their ability to find acceptable content without the recommendation of the platform, so that the most "sophisticated" users always prefer to consume outside the platform. The combination of granularity of the rating system and number of sellers that join in equilibrium determines the marginal user who is willing to join the platform. The more users join the platform, the fewer remain outside, making sellers' outside option worse as a consequence; at the same time, the more sellers join the platform, the higher are the platform's operational costs. In other words, the model incorporates two-sided indirect network effects, and the platform adjusts its design and equilibrium fee to encourage or discourage participation in a mechanism reminiscent of those studied by Teh (2022) and Gambato and Peitz (2023).

The platform profits from the rating system by extracting the surplus generated through better recommendations. Generally, improved information makes the platform more appealing and, therefore, more popular. Although more granular rating systems provide more information, the decrease in engagement due to cognitive overload of a more complex system can offset or even negate this benefit. However, the marginal gain of hosting another seller diminishes with more granular rating systems, which may induce the platform to strategically restrict the number of active sellers on the platform and reduce operational costs as a consequence. The optimal granularity of large streaming platforms is, therefore, ambiguous.

When the platform is a price-taker, the pivotal dimension of comparison changes: when selecting the optimal number of sellers to host, a platform with

5. This fee structure is akin to that of a streaming platform acquiring the rights to stream some content, like Netflix has to pay to studios to include their movies.

high marginal operational costs cannot afford to attract a large supply side. Consequently, users tend to select more forgiving thresholds than if they were to join a platform with a larger supply side. Because forgiving thresholds generate relatively low value for users, the platform selects a more granular rating system to encourage stricter ratings, use the fewer sellers that she can afford to host more efficiently, and monetize on the users' side more aggressively. The more granular a rating system is, however, the faster the marginal gain of attracting one more sellers decays. When the platform has the means to attract a large number of sellers, she has an incentive to make the rating system less informative but, crucially, less straining for the users. Because users become naturally more demanding the more sellers are available, the loss in precision is more than compensated by the volume of participation, making the simpler rating system the one that generates more value for the platform to appropriate.

The rest of the chapter is organized as follows. After a review of the relevant literature, we make our model of strategic user rating and platform environment explicit in Sections 2.2 and 2.3. We then obtain the optimal platform design when the platform is large enough to affect competition on the outside market (Section 2.4) and when it is a price-taker (Section 2.5). Finally, we draw our conclusions and detail future plans in Section 2.6.

2.1.1 Related literature

We contribute to the platforms literature on information mechanisms, ratings, and product quality. Regarding the informational role of ratings, previous research has extensively shown that ratings drive demand predictably, leading to higher prices and fees (the surplus is extracted from the user side) and to active steering towards high-quality sellers. These intuitive results are replicated under our proposed rating mechanism. For an overview of this literature, see Tadelis (2016), which focuses on two-sided platforms like Amazon, eBay or Airbnb. More broadly, the paper relates to the literature on platform design as a non-price instrument of governance (Teh, 2022). As discussed in Hagiu and Wright (2015), platforms are special in their ability to shape the interaction between users and sellers. The granularity of the rating system can be thought of as a discrete design choice that contributes to governing this interaction in the spirit of Gambato and Peitz (2023), Choi and Jeon (2023), and Bedre-Defolie, Johansen, and Madio (2024). Next, we

contribute to the literature on asymmetric information in digital platforms. We explore the particular case of a two-sided platform whose incentives are aligned with those of the users, but the platform is uninformed about product quality. Similar settings have been studied in platform economics: early work by Lizzeri (1999), for example, shows that, in an adverse selection environment, a platform only discloses whether a product satisfies a minimum quality threshold. Che and Hörner

(2018) investigate the trade-off between exploration and exploitation, finding that it is not always optimal to recommend the best available product but to improve overall information. Vellodi (2018) shows that an optimal rating system would implement upper censorship, pooling high-quality sellers together to stimulate participation and reduce the "cold-start" problem. Our modeling choices implicitly reflect (and are instructed by) these results. The closest work to ours might be by

Haraguchi and Yasui (2023), who show that the effects of a platform implementing a more precise rating system depend on whether seller entry is fixed or free. If entry is fixed, users always benefit; otherwise, low-quality sellers are pushed out, decreasing user surplus. The main difference in our study is the rating system design: we propose a mechanism for strategic users to disclose information, whereas Haraguchi and Yasui (2023) allow the platform to choose the signal granularity, making her implicitly aware of the quality of all content she hosts. Conversely, we consider a platform that remains wholly agnostic regarding what content to define acceptable, letting the threshold or thresholds emerge endogenously through active user engagement with the rating system. We find that the effect of granularity on match value depends on the platform's popularity, interacts with the number of sellers, and is, therefore, ambiguous. Additionally, we contribute to the literature

on platform quality analysis. Hopenhayn and Saeedi (2023) show that when an informed platform can signal users about sellers' quality, demand shifts from low-quality to high-quality firms. The aforementioned Bedre-Defolie, Johansen, and Madio (2024), instead, studies platform-driven certification systems and shows that two-sided platforms tolerate low-quality content under a binary rating system because it increases sellers' effort and quality. More specifically on streaming platforms, Jeon and Nasr (2016) and Gambato and Sandrini (2023) argue that these platforms may bias recommendations toward low-quality, cheap content, negatively impacting users and high-quality sellers to induce more favorable rates *vis-à-vis* the seller she hosts, a result that finds an empirical counterpart in the investigation of news aggregators' incentives to restrict access to high-quality content by Freimane (2022). Empirically, Hui, Liu, and Zhang (2023) identify a trade-off between encouraging high-quality entrant sellers to distinguish themselves from low-quality sellers and incentivizing established sellers to maintain quality. Vatter (2022), finally, shows that in Medicare Advantage, a binary quality certification achieves 94% of the welfare obtained under optimal conditions. In such a case, firms produce quality at the scoring threshold.

2.2 Likes and Double-Likes: a Model of Quality Feedback

We first characterize the liking and "double liking" technology in a vacuum. If the former is the selected rating technology, users can rate content as acceptable or

not acceptable by leaving a "like" or "thumbs up" to content they enjoy. This is the rating technology employed in platforms such as Spotify, Twitter, Instagram, and, more recently, YouTube. If the latter is selected, instead, users can rate content as good, neutral, or bad. YouTube's rating system before the last update in 2021 worked as such: leaving a like, nothing, or a dislike (now hidden and, therefore, uninformative for other users) effectively "binned" content in three levels of acceptability.

We assume, in this section, that the monopoly streaming platform we consider in the remainder of the chapter hosts k sellers each producing and providing one product, with k exogenously given.[6] Each one of the k products has a privately known inherent quality x_j, $j \in \{1, 2, ..., k\}$, drawn from some well-behaved distribution $F(x_j)$. Users value quality according to a value function $v : [0, 1] \to [0, 1]$ that: *i*) is continuous, *ii*) is increasing (i.e., if $x_j > x_r$, then $v(x_j) > v(x_r)$; products of higher quality are always better valued); and *iii*) satisfies $v(0) = 0$ and $v(1) = 1$. We can say that user i is more demanding than user i' if $v_i(x_j) \leq v'_i(x_j)$ for all $x_j \in [0, 1]$.

2.2.1 Baseline rating system

Under the less granular, "binary" rating technology, users disclose information on product quality by liking products: each user chooses a threshold $z \in [0, 1]$ such that if $v(x_j) > z$, j is rated as acceptable. A critical assumption in this model is that, when the user chooses a threshold, the k available products are divided in acceptable and not acceptable automatically. We justify this choice through the signaling role of ratings: users value being grouped together with other like-minded users. The division, then, follows from the classification performed by these other users.

Given any threshold z, the platform looks for products that were liked by similar users: she recommends, at random, an acceptable product (that is, a product such that $x_j \geq z$) if at least one is available. Otherwise, she recommends at random out of the necessarily not acceptable options. When selecting the optimal threshold, then, the user faces the following trade-off: if the threshold is too high, the set of acceptable products becomes too narrow, and not receiving any recommendation too likely. If, however, the threshold is too low, the expected quality of an acceptable product would not be high enough to qualify as such. User's ex-ante expected utility function is, then, $u(z)$, and it is given by the expected value of the

6. We endogeneize k and model entry decision of users explicitly in the next section.

recommendation received after disclosing threshold z. Formally, the user chooses such z to maximize:

$$\begin{aligned}
u(z) =& \mathbb{P}[\text{ there is a product } j \text{ valued above } z]\,\mathbb{E}[v(x)|v(x) > z] \\
&+ \mathbb{P}[\text{ there is no product } j \text{ valued above } z]\,\mathbb{E}[v(x)|v(x) \leq z] \\
=& \left(1 - \prod_{j=1}^{k} \mathbb{P}[v(x_j) < z]\right) \mathbb{E}[v(x)|v(x) > z] \\
&+ \left(\prod_{j=1}^{k} \mathbb{P}[v(x_j) < z]\right) \mathbb{E}[v(x)|v(x) \leq z] \\
=& \frac{1 - F(v^{-1}(z))^k}{1 - F(v^{-1}(z))} \int_z^1 v(x) f(x) dx + \frac{F(v^{-1}(z))^k}{F(v^{-1}(z))} \int_0^z v(x) f(x) dx.
\end{aligned}$$

Note that if a user sets her threshold to $z = 0$ or $z = 1$, she would get a random recommendation with probability one.

Before delving in the characterization of the user's optimal threshold, we clarify a few modeling choices and their implications:

Ratings as a disclosure mechanism (categorization). As mentioned, users rate content to align themselves with a group of similar individuals, enabling the platform to provide accurate recommendations based on their preferences. Consider, for example, two fans of *noir* films: when one of them consumes a good piece of content, and signals it to the platform with a like, the platform learns the preferences of that user and will recommend other films liked by the other fan if he, too, liked it. The opposite is clearly also true. This creates a strong incentive for users to engage with the rating system. Intuitively, we could think of a heterogeneous population among which there are subgroups of individuals with similar tastes. A user belonging to a subgroup is precisely interested in telling the platform that he belongs to that subgroup, so that the platform can effectively recommend him suitable products.

Constant like threshold. Implicitly, we assume that each user has a unique threshold and does not change it. This is done for simplicity: since we model recommendation and consumption as a one-shot interaction, we ignore any learning and evolution of tastes that might happen in reality.

Platform's and users' incentives are aligned. We assume the platform has no information on products' quality. In reality, the platform might have some way to determine whether some content is of high or low quality (for example, through past performance of similar content, or content by the same author). We argue, however, that platforms rely heavily on ratings, i.e., on aggregate user-generated information for recommendations, and that users are generally more susceptible

to peer reviews than they are to a platform's possibly biased certification system. Because it is in her interest to provide users with good recommendations, the platform's goal is generally aligned with those of the users who join her.

Retaliation. It is well known in the rating literature that retaliation is one of the major causes of users being reluctant to leave negative feedback; see Bolton, Greiner, and Ockenfels (2013) or Fradkin, Grewal, and Holtz (2021). We ignore this dimension altogether: retaliation is generally associated with product rating on e-commerce platforms, while we explicitly model streaming platforms.

Noisy ratings and reviews. Belleflamme and Peitz (2018) note that ratings can suffer from a lack of informativeness due to factors such as misunderstandings, idiosyncratic tastes, uncontrollable shocks, and price variations. Let us address these concerns. First, ratings in our model are not influenced by idiosyncratic tastes because the platform groups users by similar preferences. Second, uncontrollable shocks are minimized since, on a streaming platform, content is directly delivered from the platform to the user, something that cannot be said for e-commerce platforms for which poor delivery, for example, might skew ratings downwards. Third, price variations are not a concern in our setting because the platform charges a uniform fee for access to all content, ensuring that fee changes do not affect ratings. Fourth, and finally, misunderstandings may be common in reality; however, we are interested in comparing the performance of different levels of rating granularity. For this reason, we choose to normalize the cognitive load of a binary rating system, the simplest and most user-friendly system, to zero. We incorporate misunderstanding in the more granular (and hence more complex) rating systems in the next subsection.

Let us call $\tilde{F} = F \circ v^{-1}$. Because of v's properties, $\tilde{F}$ is also a distribution function. We now derive analytically the first-order conditions with respect to z in the user's problem above.

$$\frac{-k\tilde{F}(z)^{k-1}(1-\tilde{F}(z)) + (1-\tilde{F}(z)^k)}{(1-\tilde{F}(z))^2}\int_z^1 v(x)\tilde{f}(x)dx + \frac{(1-\tilde{F}(z)^k)}{(1-\tilde{F}(z))}(-v(z)\tilde{f}(z))$$
$$+(k-1)\tilde{F}(z)^{k-2}\int_0^z v(x)\tilde{f}(x)dx + \tilde{F}(z)^{k-1}v(z)\tilde{f}(z) \;=\; 0.$$

If $z^* \neq 1$, we can multiply the whole expression by $(1-\tilde{F}(z))^2$ to get:

$$g(k,z) \;=\; (k-1)\tilde{F}(z)^k\mathbb{E}[v(x)] + (k-1)\left(\tilde{F}(z)^{k-2} - 2\tilde{F}(z)^{k-1}\right)\int_0^z v(x)\tilde{f}(x)dx \tag{2.2.1}$$
$$+ (1-k\tilde{F}(z)^{k-1})\int_z^1 v(x)\tilde{f}(x)dx + (-1+\tilde{F}(z)-\tilde{F}(z)^k+\tilde{F}(z)^{k+1})v(z) \;=\; 0.$$

In general, an increase in product quantity makes the user more stringent.

Proposition 1. *It is always profitable to set a threshold $z \in (0,1)$. Moreover, $\frac{\partial z^*(k)}{\partial k} > 0$, i.e., the optimal threshold increases with product availability.*

Proof. See Appendix 2.A. □

Corollary 2.2.1. *If $k = 2$ and $v(x) = x$, then the optimal threshold is $z^* = \mathbb{E}[x]$.*

Proof. Under these assumptions, the user's utility simplifies to

$$u(z) = (1 + F(z)) \int_z^1 xf(x)dx + F(z) \int_0^z xf(x)dx = \int_z^1 xf(x)dx + F(z) \int_0^1 xf(x)dx.$$

The first order condition reads as $-zf(z) + f(z)\mathbb{E}[x] = 0$, and hence $z^* = \mathbb{E}[x]$. □

The second part of Proposition 1 states that the more products (that is, the more sellers) are present on the platform, the higher is the optimal threshold z selected by users. This is intuitive: an increase in k corresponds to an increase in random draws from the distribution of quality $F(x)$. Since the user and the platform are both interested in the highest realizations out of these draws, the standard property of order statistics (that is, the expected highest realization is increasing and concave in the number of draws) applies.

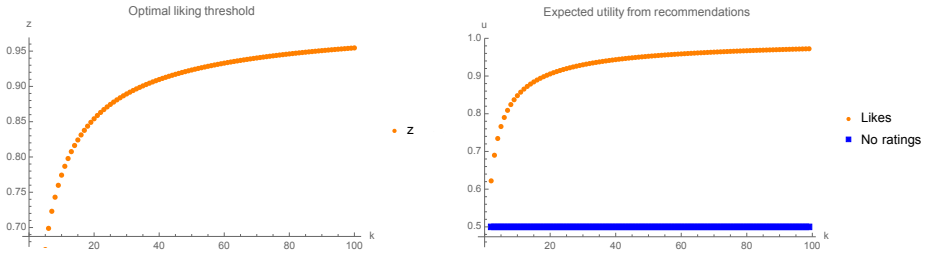
In the simplest case, when there are just two sellers (and hence, two products), and the value for quality is $v(x) = x$, a user sets her threshold precisely at the expected quality $\mathbb{E}[x]$. This shows that the rating system is useful to users even for a small number of products. Because the platform can always try to recommend something above the average of the distribution of quality, users always have an incentive to receive recommendations, which increases the value they get from joining the platform and, therefore, the subscription fee they are willing to pay.

Uniform distribution. We choose to proceed by assuming quality to be uniformly distributed. While this is not without loss of generality,[7] the complexity of modeling the platform system in which to embed the rating technology, especially the more granular rating system in the next subsection, requires some simplification. The restriction to the uniform distribution allows us to determine tractable thresholds, a necessary condition for comparisons between the equilibrium results under different granularity levels.We also assume, from now on, that users value function is $v(x) = x$, so that $v^{-1}(x) = x$ and $\tilde{F} = F$.

Corollary 2.2.2. *If $F(x) = x$ and $v(x) = x$, the optimal threshold is*

$$z^* = k^{\frac{1}{1-k}}.$$

7. The loss of generality lies in the speed at which z changes as a function of the increase in k. However, for any distribution, the direction of the change is the same: it is increasing (Proposition 1).



(a) Optimal threshold z^*.

(b) Expected utility comparison.

Figure 2.2.1. Uniformly distributed product quality.

Proof. User utility is given by:

$$u(z) = \frac{1-z^k}{1-z}\int_z^q x dx + z^{k-1}\int_0^z x dx = \frac{1+z-z^k}{2},$$

and the first order condition with respect to z leads to

$$kz^{k-1} = 1 \Rightarrow z^* = k^{\frac{1}{1-k}}.$$

□

As per Figure 2.2.1a the resulting optimal threshold is "concave" in k.[8] Notice that, as per Figure 2.2.1b, if $k > 34$, $z > 0.9$. While our exercise is merely numerical, comparing this number with that of Corollary 2.2.1 highlights how impactful a rudimentary recommendation system can be, but also how reliant it is on having enough products to make a noticeable difference.

2.2.2 More granular rating system

Suppose now that the platform allows users to signal their preferences with two thresholds ($z_L < z_H$) instead of one.[9] Users are now able to divide products into three categories. Products with quality below z_L are still unacceptable. Products with quality in between thresholds are now acceptable, but are only recommended if the platform does not find any "preferred" product, i.e., a product above the highest threshold z_H. With the same steps as above, we can define the expected

8. Since k is assumed to be a natural number for the sake of the rating system to work as specified, concavity translates to an increase in the optimal threshold that decreases in width as k increases.

9. We maintain the distributional assumption $F(x) = x$.

utility for two generic thresholds and then obtain the optimal thresholds selected by users through direct maximization:

$$\begin{aligned} u(z_L, z_H) =& \mathbb{P}[\text{there is a product above } z_H]\frac{1}{1-z_H}\int_{z_H}^{1} x dx \\ &+ \mathbb{P}[\text{there is a product above } z_L, \text{ but not above } z_H]\frac{1}{z_H - z_L}\int_{z_L}^{z_H} x dx \\ &+ \mathbb{P}[\text{there is no product above } z_L]\frac{1}{z_L}\int_{0}^{z_L} x dx \\ =& \frac{1-z_H^k}{2}(1+z_H) + \frac{(z_H^k - z_L^k)}{2}(z_H + z_L) + \frac{z_L^{k+1}}{2} \\ =& \frac{1}{2}(1 - z_H^k + z_H - z_H z_L^k + z_H^k z_L). \end{aligned}$$

Then, taking first-order conditions:

$$\begin{aligned} z_L - z_H k^{1/(1-k)} &= 0, \\ z_H^{k-1}\left(-k + (k^{(2-k)/(1-k)} - k^{k/(1-k)})z_H\right) + 1 &= 0. \end{aligned}$$

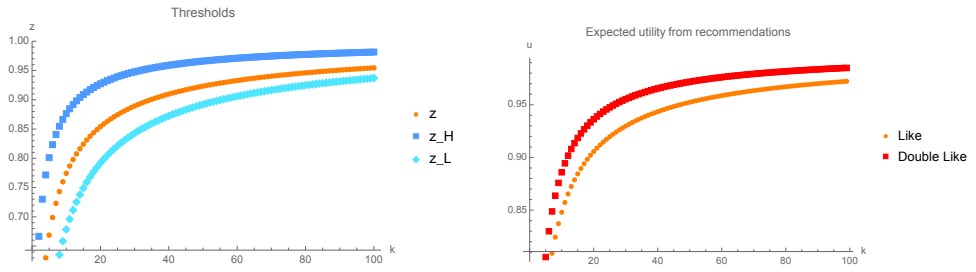


Figure 2.2.2. Likes vs double likes.

Implementing a more granular rating system allows the platform to extract more information from her users, and hence to provide better recommendations. In principle, expected utility from joining the platform should increase if the platform goes from the baseline to the more granular rating system, all else equal. Figure 2.2.2 illustrates the optimal thresholds for the double-like system and compares them to that found for the baseline rating system. It also shows how the more granular rating system provides better recommendations in expected terms. While we assume that the cost of implementing a more granular system compared to the the baseline binary system is negligible, and therefore normalized to zero, it is often argued that a more granular rating system is associated with higher cognitive effort, so participation could decay. In this setting, we assume that a fraction $\alpha > 0$ of users that join the platform does not leave a rating if the more granular rating technology is selected. Because the remaining fraction $1-\alpha$ of

users engages with the rating system, every user can still discern content quality based on the ratings of others. However, since not all users engage with the rating system, the overall quality of recommendations decreases in α. This will play a crucial role in the next section, in which we endogenize users' and sellers' participation decision.

2.3 Framework: Two-Sided Platform

We now embed the rating technology and the strategic considerations of users who engage with it in a platform setting and, in doing so, we endogenize the entry decision of both users and sellers. The existence of a rating system generates direct positive network effects for users: the more users engage with the rating system, the more accurate are the recommendations of the platform.[10] The framework crucially generates indirect network effects as well: the more users (or sellers) join the platform, the stronger the incentive becomes for sellers (or users) to join as well.

Sellers. We consider a finite number of sellers, $K \geq 2$, indexed by j, each producing at zero cost some content of quality $x_j \sim U[0,1]$. Content quality is private information: we assume that the platform cannot discern how good a product is without users consuming it and then engaging with the rating system. Sellers $j \in \{1,2,..,k\}$, with $k < K$, join the platform. The remaining $K-k$ sellers remain outside. The k sellers on the platform cash in a one-time fee p_{in} from the platform in exchange for the rights to stream their content.[11] The $K-k$ sellers outside, instead, sell to users who do not join the platform at price p_{out}.

Users. There is a mass m of users interested in consuming content either on the platform or on the outside market. Users are characterized by a private variable $b \sim U[0,1]$ that reflects their "sophistication": the higher b is, the more sophisticated the user. We assume that sophistication affects the outside option of users. In particular, the more sophisticated a user is, the better is his outside option. For example, one can imagine sophisticated users to know how and where to find high quality content without the platform recommendation, while unsophisticated users do not.

10. This is one of the crucial characteristic of any platform. Quoting Belleflamme and Peitz (2018): "in an e-commerce context, platforms have the potential to generate network effects, as a user is better off the more fellow users that are around."

11. This modeling choice mirrors contracts between most streaming platforms and sellers. Netflix, for example, buys the rights to stream content for a finite amount of time through a one-time *license agreement*.

Because users' gains from joining the platform depend on their sophistication, only a proportion $\bar{b} \in (0,1)$ of users will join the platform. The remaining $(1-\bar{b})m$ users, instead, purchase content from one of the sellers outside the platform at the aforementioned price p_{out}. Users who join the platform pay a subscription fee f to get access to the streamed content and the recommendation system. Users join the platform anticipating how they will engage with the rating system and, in particular, where they will optimally set the quality threshold $z_i \in [0,1]$ (or, $z_{i,L}, z_{i,H} \in [0,1]$) and, therefore, their expected utility derived from the recommendation system. Notice that, by construction, $z_i = z$, $z_{i,L} = z_L$, and $z_{i,H} = z_H$ for all i. So, we henceforth drop the subscript i and directly refer to the optimal common thresholds z, z_L, and z_H. Users join if and only if the fee is weakly lower than the expected utility generated through strategic engagement with the rating system.

Platform. The platform puts users and sellers in contact and mediates their interaction through the rating system. The platform receives the subscription fee for all users who join, and pays p_{in} for the right to stream content to k sellers. The platform always selects the optimal number of sellers to host as part of her strategic choices. In Section 2.4, we assume that the platform can affect the degree of competition on the residual market: the more sellers join the platform, the laxer is the competition on the residual market and, therefore, the higher the expected profits from not joining, all else equal. In Section 2.5, instead, we assume the platform to be a price-taker: when selecting the optimal number of sellers to attract, the platform cannot affect the expected profits sellers would make on the residual market. In the latter case, her only leverage to attract sellers manifests through the mass of users she manages to attract.

As mentioned, the platform cannot observe quality directly, but through the likes and dislikes can group products by quality. Given threshold z and a proportion of users that join $\bar{b}$, with probability $\bar{b}$ the platform recommends something acceptable if available, and at random if more than one product is available. Otherwise, the platform recommends at random out of the unacceptable options. With probability $1-\bar{b}$, instead, the platform recommends at random from the whole library. In other words, the more users join and engage with the rating system, the better the recommendations. Assuming that the platform selects the single like as her rating system, then, recommendations yield users expected utility from joining the platform equal to[12]

12. Henceforth, $u(z)$ refers to the utility a user derives from recommendations, while $U(z)$ refers to the utility a user derives from joining the platform.

$$\begin{aligned} U(z,\bar{b}) &= \bar{b}\,u(z) + (1-\bar{b})\frac{1}{2} \\ &= \bar{b}\left[\frac{(1-z^k)}{1-z}\int_z^1 q dq + z^{k-1}\int_0^z q dq\right] + (1-\bar{b})\frac{1}{2} \\ &= \frac{1}{2}\left(\bar{b}\,(1+z-z^k) + (1-\bar{b})\right) \\ &= \frac{1}{2}\left(1+\bar{b}(z-z^k)\right). \end{aligned}$$

The platform's profits are given by

$$\mathbb{E}[\Pi_p] = m\bar{b}f - k p_{in}.$$

Timing and equilibrium concept. The timing of the interaction is as follows:

1. the platform observes the total mass of users and sellers in the market (and, in Section 2.5, p_{out}) and commits to a more or less granular rating system and to a subscription fee f for users,
2. sellers observe the design choice of the platform (and, in Section 2.5, p_{out}) and make their entry decision,
3. users observe the number of sellers who join the platform, the design choices of the platform, and make their entry decision, determining $\bar{b}$,
4. profits realize.

Our equilibrium concept is Perfect Bayesian Equilibrium (PBE): sellers join the platform anticipating the number of users who will follow. Notice that, as in Gambato and Peitz (2023), the timing makes the sequentiality of the entry decisions of sellers and users immaterial. We choose to separate the two choices for clarity, but an equivalent simultaneous entry timing would generate the same PBE.

2.3.1 Equilibrium analysis

Users' entry decision. Users' outside option depends on their sophistication b. We assume that a user defined by some level of sophistication $b \in [0,1]$ would derive utility b from looking for, and consuming, content outside the platform.[13] Users with a high value b, therefore, are more difficult to convince to join the platform.

Consequently, there exists a threshold $\overline{b} \in (0,1)$, given by the solution to $U(z,\bar{b}) = U_{out}(\bar{b})$, such that:

13. We remain agnostic as to how such utility is actually generated to keep the outside option as reduced-form as possible. One can imagine that a sophisticated user has an easier time finding high quality content than an unsophisticated one, so that it would be costlier, for the latter, to consume any quality level.

- users with $b \in [0, \bar{b}]$ join the platform and use the recommendation system;
- users with $b \in (\bar{b}, 1]$ stay outside the platform.

Sellers' entry decision. Sellers join the platform as long as they are remunerated enough to be indifferent between joining the platform and staying outside. Given the price on the outside market, the expected profits that the platform has to match depend on the fraction of users and sellers that do not join the platform, that is, $(1-\bar{b})m$ and $K-k$ respectively.

Formally, each of the $K-k$ outside-the-platform sellers makes expected profits:

$$\mathbb{E}[\pi_{out}] = p_{out}\frac{(1-\bar{b})m}{K-k}.$$

The remaining k sellers who join the platform must be indifferent between joining and not joining; therefore, $p_{in} = \mathbb{E}[\pi_{out}]$.

Platform's design choice. The platform selects the granularity of the rating system, fee f, and, in Section 2.4, number of sellers k to maximize expected profits $\mathbb{E}[\pi_p]$. These depend on the equilibrium entry decisions that follow from these choices. In particular, platform revenue will be equal to $m\bar{b}f$, and operational costs will be equal to kp_{in}.

First, in equilibrium, it must be that $f = U(z^*, \bar{b})$, or $f = U(z_L^*, z_H^*, \bar{b})$, where both thresholds and $\bar{b}$ depend on the equilibrium k. When the platform can affect p_{out} through the number of users she manages to attract, the platform selects k to maximize profits knowing how this will affect entry behavior of both users and sellers, and then selects the granularity of the rating system accounting for this optimal k and the relative expected profits. In equilibrium, by construction, it must be that $\bar{b}$ is such that user $b < \bar{b}$ is strictly better off joining the platform, and user $b > \bar{b}$ strictly better off not joining. As $p_{in} = \mathbb{E}[\pi_{out}]$, platform profits are given by

$$\mathbb{E}[\Pi_p] \;=\; m\left(\bar{b}f - kp_{out}\frac{1-\bar{b}}{K-k}\right). \tag{2.3.1}$$

2.4 Monopolistic Platform

We first consider the case in which the platform is big enough to affect equilibrium prices on the outside market. We assume that the entry decisions of users and sellers determine the size of the outside market and the degree of competition sellers who do not join the platform face. Intuitively, the more users join the platform, the smaller is the demand outside, but because of the mechanical entry decision of users, the higher is the sophistication of the marginal users $\bar{b}$. On the

contrary, the more sellers join the platform, the lower is the degree of competition on the outside market.

This market structure generates several trade-offs. The platform wants to attract more sellers to make her recommendations more effective, which attracts users and reduces off-the-platform demand. The platform wants to limit the number of sellers hosted to reduce operational costs, both directly (through the number of products hosted) and indirectly (through the higher price sellers must receive to be indifferent between joining and not joining). Notice that, in this configuration, the total size of the demand does not affect strategic decisions because, as per equation (2.3.1), m does not affect the decisions of the platform at the margin.

We proceed as follows. First, we consider the baseline equilibrium in the absence of any recommendation system. Then, we solve separately for the equilibrium outcome under single-like and double-like rating systems. Finally, we compare the three and present the section's main results.

2.4.1 Different rating systems

No rating system. In the absence of ratings, the platform provides expected utility $\frac{1}{2}$ to its users. It follows immediately that any user $b \leq \bar{b} = \frac{1}{2}$ would prefer to join the platform. Because the marginal user outside of the platform is only willing to pay a price up to $\bar{b} = \frac{1}{2}$, it is immediate to obtain equilibrium prices

$$p_{out} = f = \frac{1}{2}, \qquad p_{in} = \mathbb{E}[\pi_{out}] = m\frac{1-\bar{b}}{K-k}p_{out} = \frac{m}{4(K-k)},$$

and platform profits

$$\Pi_p = \frac{m}{4}\left(\frac{K-2}{K-1}\right).$$

Single-like rating system. Under the binary like technology, we have that $\mathbb{E}[U(z)] = b\frac{1}{2}(1+z-z^k) + (1-b)\frac{1}{2}$ with $z^* = k^{1/(1-k)}$, and $\mathbb{E}[U_{out}] = b$ as obtained in Section 2.2. Then, $\bar{b} = \frac{1}{2-z+z^k}$ and $\bar{b} \geq 1/2$ as $z > z^k$.[14] Embedding the rating system in the current platform configuration allows us to draw some conclusions with regard to the attraction spiral affecting users and sellers, and how the strategic design choices of the platform affect equilibrium outcomes.

Lemma 2.4.1. *The share of users that join the platform, $\bar{b}(k)$, increases in the number of sellers, k.*

14. Notice that $\bar{b}$ is always well-defined because $z \in [0,1]$. Moreover, to lighten the notational burden, we will write z instead of $z(k)$ when the dependence on k does not need to be explicitly stated.

Proof. Using that $\bar{b} = \frac{1}{2-z+z^k}$ and the fact that $\frac{\partial z}{\partial k} > 0$, we have that:

$$\frac{\partial \bar{b}(k)}{\partial k} = -\frac{1}{(2-z+z^k)^2}\left(-\frac{\partial z}{\partial k} + z^k \log(z)\frac{\partial z}{\partial k}\right) > 0.$$

Note that, in order to derive this result, we have treated k as a real number. This is, however, without loss of generality, because if a function of a real variable is increasing, it will also be increasing when the variable is restricted to the naturals. □

It is intuitive that as more sellers join the platform, more users follow suite. Indeed, the more sellers are available on the platform, the more likely it is for an "acceptable" product to be available and, as per Proposition 1, the threshold defining what quality level is acceptable increases. This implies that the value of joining the platform increases and, with it, the marginal user $\bar{b}$ who finds it worthwhile to join.

On the other hand, more sellers generate an increase in operational costs that the added demand (and higher fee $\bar{b}$ that comes with it) cannot compensate indefinitely. Because attracting sellers reduces the competitive pressure outside the platform, sellers who join the platform require a higher royalty to be incentivized to join. At the same time, the marginal effect of adding one more seller decreases as k grows. These observations suggest that there is a limit to how many sellers the platform is interested in attracting.

Hence, we can ignore any corner solution:

$$\Pi_p(k) = m\left(\bar{b}^2 - \bar{b}k\frac{1-\bar{b}}{K-k}\right) = \frac{m\bar{b}}{K-k}(K\bar{b}-k)$$

is always well defined.

Double-like rating system. We consider now the effects of introducing a more granular rating system. Recall that, under the double-like rating system, users can select two thresholds (z_L and z_H) instead of one, but not all users will engage with the rating system due to the higher cognitive burden it entails.

Following the same logic as for the single-like rating system, we can use the equilibrium thresholds z_L^* and z_H^* to determine the marginal user who joins the platform as a decreasing function of α, that is, the number of users who do not rate what they consume, and k, the number of sellers the platform introduces. As before, the platform is able to provide a recommendation with a probability that depends on the number of ratings. Therefore, the expected utility of users under the double-like rating system can be written as

$$U(\alpha, z_L, z_H, \bar{b}) = \bar{b}(1-\alpha)u(z_L, z_H) + (1-\bar{b}+\alpha\bar{b})\frac{1}{2}.$$

In equilibrium, then, the share $\bar{b}$ of users who join the platform, and expected platform profits, are

$$\bar{b}(\alpha,k)=\frac{1}{2-(1-\alpha)(2u(z_L,z_H)-1)}=\frac{1}{2-(1-\alpha)(-z_H^k+z_H-z_Hz_L^k+z_H^kz_L)},$$

and

$$\Pi_p^{dl}=\frac{\bar{b}(\alpha,k)}{K-k}(K\bar{b}(\alpha,k)-k),$$

respectively.

2.4.2 Comparison of rating technologies

Unsurprisingly, the ability to manipulate the intensity of the network effects allows the platform to achieve higher profits, making the introduction of a rating technology always worthwhile:

Proposition 2. *The implementation of the single-like rating system always increases platform profits.*

Proof. Denoting by Π_p^{nr} the profits made under no ratings, $\Pi_p^l>\Pi_p^{nr}\Leftrightarrow\frac{\bar{b}}{K-k}(K\bar{b}-k)>\frac{1}{4}\left(\frac{K-2}{K-1}\right)\Leftrightarrow K>\frac{4\bar{b}-2}{4\bar{b}-1}$. As $\frac{4\bar{b}-2}{4\bar{b}-1}<1$, the result follows. □

The comparison between the more and less granular rating systems, leads to a few interesting observations. The double-like rating system is generally more effective than the single-like technology at attracting users. This follows from two separate but related effects. On one hand, the double-like rating system allows users to select as acceptable more quality levels ($z_L<z$ always), which reduces the probability of low-quality content to be recommended at all. On the other hand, it makes the return of adding one more seller decay faster then under the less granular rating system. Intuitively, because each product can be used more efficiently by the platform under the more granular rating system, the trade-off between more sellers attracting users and increasing the operational costs of the platform becomes unprofitable for a lower value of k. In other words, the platform needs less sellers to maximize profits.

In the absence of additional frictions, the more granular rating system is always preferred by the platform. As mentioned in Section 2.2, however, more granular rating systems are generally understood to increase the cognitive load of users. While we maintain the assumption that the cost of implementing a more granular rating technology is negligible, this additional cost is not, since it directly affects the effectiveness of the rating system. As we see next, this effect can make the more granular rating system worse than the single-like system even if relatively few users are affected:

Example 3. Suppose $K = 100$. For any $\alpha > \bar{\alpha}(100) = 0.042$, the platform's equilibrium profits are higher under the single-like rating system than under the double-like alternative. Similarly, for $K \in \{150, 200, 300, 1000\}$, the platform's equilibrium profits are higher under the single-like rating if α is such that $\alpha > \bar{\alpha}(150) = 0.0328$, $\alpha > \bar{\alpha}(200) = 0.0268$, $\alpha > \bar{\alpha}(300) = 0.02$, and $\alpha > \bar{\alpha}(1000) = 0.009$, respectively.

The double-like rating system is efficient when it comes to using the available products. As a result, users are more inclined to join the platform under this more granular system than if the platform introduced a less granular one. This implies that, all else equal, the marginal user that joins the platform is more sophisticated under the double-like rating system, which means that the equilibrium p_{out} is higher as well. Because not all users engage with the rating system, the platform cannot monetize its recommendation system as efficiently. If enough users fail to engage with the rating system, the platform is left having to compensate the fewer sellers more without being able to extract a high enough fee to compensate it fully. Notice that, as per Example 3, The larger K, the lower is the threshold of engagement above which the platform foregoes the more granular rating system. The result follows from the fact that, as K increases, so does the optimal number of sellers k that the platform wants to host. As k grows, the value of the more granular rating system decays, making users' engagement become relatively more valuable as a consequence.

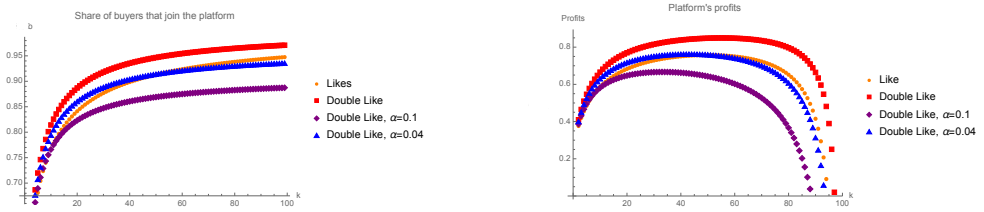


Figure 2.4.1. Like vs double like.

While the numerical results in Example 3 are just that, the mechanism behind it, illustrated in Figures 2.4.1, holds generally: if the platform is big enough, relative to the market, to be able to affect the degree of competition faced by those who do not join, a more granular rating system is only preferable if enough users engage with it.

The mechanism, however, depends crucially on the assumption of the platform being able to influence the equilibrium price on the outside market directly. In today's economic landscape, several competing platforms are active in the streaming market. We show next how the results above change when we consider a platform

with limited bargaining power—in particular, one that takes the equilibrium price of consuming on the outside channel as given.

2.5 Price-taker platform

We now assume that the platform is a price-taker when it comes to the price of streaming any piece of content. One can imagine, for example, multiple platforms all competing among themselves and with the outside channel when users are single-homing and captive of one of the available platforms. While we do not model platform competition explicitly, we proxy the effect that would arise for the strategic choice of the rating system in a world in which a single platform cannot affect the competitive conditions of the outside market by herself.

Formally, we assume p_{in} to be exogenously fixed; platform profits are now given by

$$\Pi_p(k) = m\bar{b}^2(k) - k p_{in}.$$

Unlike before, the size of the demand does matter now: while in Section 2.4 m scaled up both revenue and operational costs, now the former is completely unrelated to the mass of users in the economy. Henceforth, we apply the following normalization: we assume $p_{in} = 1$, and rewrite the problem as a function of the "exposure" that the platform can grant sellers, labelled as $\rho = \frac{m}{p_{in}}$. Intuitively, a higher ρ reflects a platform with higher reach or popularity and, therefore, proxies its bargaining power *vis-à-vis* sellers.[15] To lighten the notational burden, we henceforth refer to the share of users that join our representative platform under the single-like and double-like rating system as $\bar{b}(k)$ and $\tilde{b}(\alpha, k)$ respectively:

$$\Pi_p^l(k) \;=\; \bar{b}^2(k) - k\rho^{-1}; \quad \Pi_p^{dl}(k) \;=\; \tilde{b}^2(\alpha, k) - k\rho^{-1}. \tag{2.5.1}$$

The main result of the section emerges from a direct comparison of the above equations given strategic thresholds set by users and subsequent equilibrium platform demands $\bar{b}$ and $\tilde{b}$:

Proposition 4. *Assuming that $\tilde{b}^2(k, \alpha)$ is an increasing and concave function of k, if $\alpha < 0.16$, then the double-like rating system is preferred when the platform has relatively low bargaining power (i.e., ρ is small); otherwise, the binary rating system is chosen instead. If $\alpha > 0.16$, the binary rating system is always the preferred one.*

Proof. Here we provide a sketch of the proof, while the formal details can be found in Appendix 2.A. The proof is divided in four steps. (i) We apply Lemma

15. Alternatively, one can think of ρ as an inverse measure of how costly it is to host sellers relative to the number of buyers to profit off.

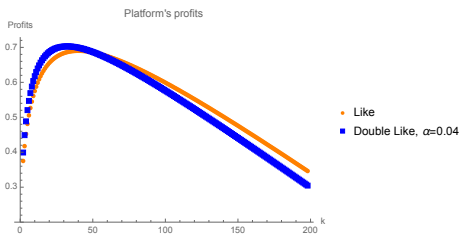
2.5.1 to show that profit functions are concave. (ii) We show that if $\alpha < 0.16$, there is single crossing and the crossing point is independent of ρ. (iii) We show that there is a range of values (ρ', ρ'') of ρ for which both $\Pi^{l}(k)$ and $\Pi^{dl}(k)$ have an interior maximum (i.e., at some $k > 2$). (iv) We show that there are some $\hat{\rho}_{dl} > \hat{\rho}_{l}$ such that if $\rho \in (\hat{\rho}_{dl}, \rho'')$ both the maximum of $\Pi_p^{l}(k)$ and $\Pi_p^{dl}(k, \alpha)$ occur to the right of the crossing point, and hence the like technology is preferred and, similarly, if $\rho \in (\rho', \hat{p}_{l})$, both maxima are reached to the left of the crossing point and hence the double-like technology is preferred. □

Proposition 4 is the result of several moving parts that interact in non-obvious ways. First, it should not be surprising that if platform engagement is low, the single-like rating system is preferable. It might be more surprising that the threshold at which the switch happens is much higher than it was in the previous section. This follows from the discussion above: when the platform does not affect the equilibrium price outside, the negative effect of lower engagement and more granular ratings is reduced. Unlike before, the granularity of the rating system does not affect p_{out} and, therefore, does not indirectly generates additional costs to the platform.

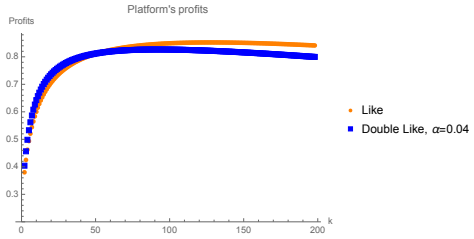
Far more interesting is the interplay between engagement and relative platform size. In words, Proposition 4 states that even if engagement with the rating system is high, a platform with enough bargaining power (that is, a platform for which attracting one more seller is relatively cheap) prefers to implement the least informative rating system. This might be counter-intuitive at first, since we are only capturing informational factors while any other strategic motive to reduce granularity (such as the platform pooling products of different qualities to promote exploration, as in Vellodi (2018)) is absent. However, there is a subtler trade-off between the double-like and the single-like rating systems that depends on the price-to-size ratio ρ^{-1}.

When the platform cannot afford to host a large number of sellers relative to the demand she can attract, the user optimally sets rather forgiving thresholds under both rating systems. In relative terms, then, this is when the significant difference in informativeness bites the hardest. Even though there might be less engagement, the marginal benefit of hiring an extra seller when few sellers are already hosted is higher if the platform selects the double-like rating system (Figure 2.5.1a). When the platform hosts a large number of sellers, instead, users strategically tighten their constraint: the informational gain of adopting the more informative system becomes less and less impactful. If ρ is very large, then, the loss of profit generated by any lack of engagement, which constrains the subscription fee users are willing to pay, dominates the trade-off (Figure 2.5.1b).

Formally, the result emerges as follows. Notice first that both $\Pi_p^{l}(k)$ and $\Pi_p^{dl}(k)$ are concave as functions of k. Next, it can be shown that $\Pi_p^{dl}(k = 2) > \Pi_p^{l}(k = 2)$



(a) Profits for $\rho^{-1} = 0.003$.



(b) Profits for $\rho^{-1} = 0.0005$.

Figure 2.5.1. Small platform size vs large platform size.

and $\Pi_p^{dl}(k \to \infty) < \Pi_p^{l}(k \to \infty)$. Hence, the two functions single-cross at some k that, crucially, does not depend on ρ. Let us denote this point by $\hat{k}$. With this in hand, we can show that, if ρ is small enough, both $\Pi_p^{l}(k)$ and $\Pi_p^{dl}(k)$ reach their maximum before $\hat{k}$ (and, therefore, the double-like system dominates). *Vice versa*, for high values of ρ, both function reach $\hat{k}$ before their maximum (so that the like technology yields higher profits). We notice that there is a parameter region in which ρ takes on medium values were the optimal rating system is ambiguous. We also notice that the result relies on the following:

Lemma 2.5.1. *The function $\bar{b}^2(k)$ is an increasing and concave function of k.*

Proof. See Appendix 2.A. □

A similar lemma for the double-like rating system is hard to state unambiguously. We can, however, conjecture that the function $\tilde{b}^2(k, \alpha)$ is an increasing and concave function of k, for all $\alpha \in (0, 1)$. While the mathematical complexity of the problem does not allow us to formally prove this result, which we had to assume in Proposition 4, we provide numerical simulations up to $k = 200$ in Appendix 2.B, and we show that $\tilde{b}^2(k)$ tends to $\frac{1}{1+\alpha}$ as k grows large.

2.6 Conclusion

In this chapter, we study the strategic role of granularity in online rating systems and its interaction with user engagement and platform size. We produce a tractable framework to account for users' strategic ratings as a means to signal their taste to a streaming platform, and the ramifications of their choices for a platform's optimal design. Users become pickier as the catalogue of content grows, which can help explain why platforms tend to select more granular, complex, and cognitively heavy rating systems when they have access to a limited number of sellers. In turn, the platform is shown to value engagement over informativeness as it grows in size and in popularity.

This chapter explains a somewhat odd dynamic in the evolution of streaming platforms' design philosophy. Namely, anecdotal evidence suggests that larger platforms seem to select simpler systems (YouTube being a clear example of this tendency), while platforms struggling to attract traffic seem to steer towards more informativeness. This would explain why Netflix introduced a more complex and granular system recently, when competition by alternative platforms (like Disney+ and HBO Max) became fiercer after a *de facto* monopoly in the second half of the 2010s.

The chapter is, so far, explicitly descriptive: we take several components of the complex environment under scrutiny as exogenous. However, our distributional assumptions and modeling choices would allow us to go further. In particular, the framework is designed with the intention of making content quality investment an endogenous, costly choice. We expect the rating system to positively affect the incentives to invest in quality, and generate a second, instrumental spiral between quality content and users' "pickiness". The platforms' choice to limit the informativeness of the rating system, then, might have a deterioration of the incentives to invest in quality as an unintended consequence.[16] We leave this additional dimension for future research.

16. In this context, one might see the long and widely covered strike of Hollywood writers of 2023 as evidence of a recent degradation of quality in content production. The use of AI powered tools has become more and more widespread in creative art.

Appendix 2.A Omitted proofs

Proof of Proposition 1

Proof. First, we prove the first part of the proposition, i.e., that it is always profitable to choose a non-trivial threshold. The utility a user derives under no rating system is $u_{nr} = u(z=0) = u(z=1) = \int_0^1 v(x)\tilde{f}(x)dx$. Hence,

$$
\begin{aligned}
u(z) \;>\; u_{nf} \;&\Leftrightarrow\; \frac{(1-\tilde{F}(z)^k)}{1-\tilde{F}(z)} \int_z^1 v(x)\tilde{f}(x)dx + \tilde{F}(z)^{k-1} \int_0^z v(x)\tilde{f}(x)dx \;>\; \int_0^1 v(x)\tilde{f}(x)dx \\
&\Leftrightarrow\; (1-\tilde{F}(z)^k) \int_z^1 v(x)\tilde{f}(x)dx + (\tilde{F}(z)^{k-1} - \tilde{F}(z)^k) \int_0^z v(x)\tilde{f}(x)dx \\
&>\; (1-\tilde{F}(z)) \int_0^1 v(x)\tilde{f}(x)dx \\
&\Leftrightarrow\; \tilde{F}(z) \int_0^1 v(x)\tilde{f}(x)dx + \tilde{F}(z)^k \int_0^z v(x)\tilde{f}(x)dx \\
&>\; \int_0^z v(x)\tilde{f}(x)dx + \tilde{F}(z)^k \int_0^1 v(x)\tilde{f}(x)dx \\
&\Leftrightarrow\; \tilde{F}(z)(1-\tilde{F}(z)^{k-1}) \int_0^1 v(x)\tilde{f}(x)dx \;>\; (1-\tilde{F}(z)^{k-1}) \int_0^z v(x)\tilde{f}(x)dx \\
&\Leftrightarrow\; \int_0^1 v(x)\tilde{f}(x)dx \;>\; \frac{1}{\tilde{F}(z)} \int_0^z v(x)\tilde{f}(x)dx \\
&\Leftrightarrow\; \mathbb{E}[v(x)] \;>\; \mathbb{E}[v(x)|x \;<\; z],
\end{aligned}
$$

which holds because v is increasing by assumption.

Now, we prove the second part of the proposition, i.e., that the optimal threshold increases with product quantity. First of all, we apply the Implicit Function Theorem to the function $g(k,z)$ from equation 2.2.1. We want to show that

$$\frac{\partial z}{\partial k} = -\frac{\frac{\partial g(k,z)}{\partial k}}{\frac{\partial g(k,z)}{\partial z}} > 0.$$

We have that the partial derivative with respect to k is:

$$
\begin{aligned}
\frac{\partial g(k,z)}{\partial k} \;=\; &\underbrace{\log(\tilde{F}(z))}_{<0} \underbrace{\left[(1-\tilde{F}(z))v(z) - \int_z^1 v(x)\tilde{f}(x)dx\right]}_{<0} \\
&\underbrace{+\,\tilde{F}(z)^k \mathbb{E}[v(x)] + (\tilde{F}(z)^{k-2} - 2\tilde{F}(z)^{k-1}) \int_0^z v(x)\tilde{f}(x)dx}_{>0} \;>\; 0
\end{aligned}
$$

The term over the second brace is negative:

$$(1-\tilde{F}(z))v(z)-\int_z^1 v(x)\tilde{f}(x)dx<0 \Leftrightarrow v(z)<\frac{1}{1-\tilde{F}(z)}\int_z^1 v(x)\tilde{f}(x)dx=\mathbb{E}[v(x)|x>z].$$

The term over the third brace is positive

$$\begin{aligned}&\tilde{F}(z)^k\mathbb{E}[v(x)]+(\tilde{F}(z)^{k-2}-2\tilde{F}(z)^{k-1})\int_0^z v(x)\tilde{f}(x)dx \;>\; 0\\ \Leftrightarrow\;& \tilde{F}(z)^2\mathbb{E}[v(x)]-2\tilde{F}(z)\int_0^z v(x)\tilde{f}(x)dx+\int_0^z v(x)\tilde{f}(x)dx \;>\; 0\end{aligned}$$

because the discriminant of the second degree equation in $\tilde{F}(z)$ is negative:

$$\Delta=4\left(\int_0^z v(x)\tilde{f}(x)dx\right)^2-4\mathbb{E}[v(x)]\int_0^z v(x)\tilde{f}(x)dx<0$$

for all $z\in(0,1)$.

On the other hand, the partial derivative with respect to z is given by:

$$\begin{aligned}\frac{\partial g(k,z)}{\partial z} \;=\;& (-1+\tilde{F}(z)-\tilde{F}(z)^k+\tilde{F}(z)^{k-1})v'(z)+(k-1)\tilde{F}(z)^{k-3}\tilde{f}(z)\Big(k\mathbb{E}[v(x)]\tilde{F}(z)(\tilde{F}(z)-1)\\ &+(k-2)(1-\tilde{F}(z))\int_0^z v(x)\tilde{f}(x)dx+v(z)\left(\tilde{F}(z)-2\tilde{F}(z)^2+\frac{k+1}{k-1}\tilde{F}(z)^3\right)\Big) \;<\; 0\end{aligned}$$

First, $v'(z)>0$ and $(-1+\tilde{F}(z)-\tilde{F}(z)^k+\tilde{F}(z)^{k-1})<0$ (note that $\tilde{F}(z)^{k-1}(1-\tilde{F}(z))<1-\tilde{F}(z)\Leftrightarrow\tilde{F}(z)^{k-1}<1$). Moreover, $(k-1)\tilde{F}(z)^{k-3}\tilde{f}(z)>0$ and it just remains to show that

$$\begin{aligned}&-k\mathbb{E}[v(x)]\tilde{F}(z)(1-\tilde{F}(z))\\ &+(k-2)(1-\tilde{F}(z))\int_0^z v(x)\tilde{f}(x)dx+v(z)\left(\tilde{F}(z)-2\tilde{F}(z)^2+\frac{k+1}{k-1}\tilde{F}(z)^3\right) \;<\; 0.\end{aligned}$$

Thus,

$$\begin{aligned}&(k-2)(1-\tilde{F}(z))\int_0^z v(x)\tilde{f}(x)dx+v(z)(\tilde{F}(z)-2\tilde{F}(z)^2+\frac{k+1}{k-1}\tilde{F}(z)^3)\\ &<\; (k-2)(1-\tilde{F}(z))\int_0^z v(x)\tilde{f}(x)dx+v(z)\tilde{F}(z)\\ &<\; (k-2)(1-\tilde{F}(z))\int_0^z v(x)\tilde{f}(x)dx+\tilde{F}(z)\int_z^1 v(x)\tilde{f}(x)dx\\ &=\; (k-2)(1-\tilde{F}(z))\tilde{F}(z)\mathbb{E}[v(x)|x\;<\;z]+\tilde{F}(z)\int_z^1 v(x)\tilde{f}(x)dx\\ &=\; (k-2)(1-\tilde{F}(z))\tilde{F}(z)(\mathbb{E}[v(x)]-\mathbb{E}[v(x)|x\;<\;z])+\tilde{F}(z)(1-\tilde{F}(z))\mathbb{E}[v(x)|x\;>\;z]\\ &<\; (k-2)(1-\tilde{F}(z))\tilde{F}(z)\mathbb{E}[v(x)]\\ &<\; k(1-\tilde{F}(z))\tilde{F}(z)\mathbb{E}[v(x)].\end{aligned}$$

And finally,

$$\frac{\partial z}{\partial k} = -\frac{\frac{\partial g(k,z)}{\partial k}}{\frac{\partial g(k,z)}{\partial z}} > 0.$$

□

Proof of Proposition 4

Proof. (i) $\Pi_p^l(k)$ is concave if and only if $\frac{\partial^2 \Pi_p^l(k)}{\partial k^2} < 0$. But by definition of the profit function in equation (2.5.1) and the fact that $n > 0$, this is equivalent to $\frac{\partial^2 \bar{b}^2(k)}{\partial k^2} < 0$, which is precisely the result in Lemma 2.5.1. The reasoning is completely analogous for $\Pi_p^{dl}(k,\alpha)$, given our assumption on the concavity of $\tilde{b}^2(\alpha,k)$.

(ii) Once we have established concavity, in order to show single-crossing we need to prove $\Pi_p^{dl}(k=2) > \Pi_p^l(k=2)$ and $\Pi_p^{dl}(k\to\infty) < \Pi_p^l(k\to\infty)$. On the one hand, $\Pi_p^l(k=2) = \frac{n}{6}(\sqrt{2}+3) - 2p$, and $\Pi_p^{dl}(k=2,\alpha) = \frac{3.4244n}{5.8489+\alpha} - 2p$. Hence, if $\alpha < 0.16$, $\Pi_p^{dl}(k=2,\alpha) > \Pi_p^l(k=2)$. On the other hand, when $k\to\infty$, both $u(z^l(k)) \to 1$ and $u(z^{dl}(k)) \to 1$. Hence, $\bar{b}^2(k) \to 1$ but $\tilde{b}^2(k,\alpha) \to \frac{1}{(1+\alpha)^2}$. Consequently, $\Pi_p^{dl}(k\to\infty) < \Pi_p^l(k\to\infty,\alpha)$ for all $\alpha > 0$. Thus, if $\alpha < 0.16$, there is single-crossing. Let us call $\hat{k}$ to the k such that $\Pi_p^{dl}(k) = \Pi_p^l(k,\alpha)$. By definition of the profit functions in equation (2.5.1), $\hat{k}$ is given by the solution to $\bar{b}(k) = \tilde{b}(k,\alpha)$, and it is independent of p.

(iii) By Lemma 2.5.1 and this Proposition's assumptions, we know that $\bar{b}^2(k)$ and $\tilde{b}^2(k,\alpha)$ are increasing functions of k. As $-kp$ is decreasing and linear in k, $\bar{b}^2(k) - kp$ has a maximum for all $p > p'$ for some $p' > 0$. However, if p is larger than some p'', the maximum will occur at $k = 2$. Thus, there is an interior maximum (i.e., at some $k > 2$), for $p \in (p', p")$.

(iv) Let us define[17]

$$\hat{p}_l := \frac{\partial \bar{b}^2(\hat{k})}{\partial k}; \quad \hat{p}_{dl} := \frac{\partial \tilde{b}^2(\hat{k},\alpha)}{\partial k}.$$

The partial derivatives are decreasing functions of k. Hence, for $p > \hat{p}_{dl}$, the maximum of $\Pi_p^{dl}(k,\alpha)$ takes place to the left of $\hat{k}$. Similarly, for $p > \hat{p}_l$, the maximum of $\Pi_p^l(k)$ also takes place to the left of $\hat{k}$. Hence, both take place to the left of $\hat{k}$ if $p \in (\max\{\hat{p}_{dl},\hat{p}_l\}, p'') = (\hat{p}_l, p'')$. In this case, the double like technology is preferred. Similarly, when $p \in (p', \min\{\hat{p}_{dl},\hat{p}_l\}) = (p', \hat{p}_{dl})$, the like technology is preferred. □

17. For the easiness of notation we drop the α term in $\hat{p}_{dl}$, as it is fixed along the whole reasoning.

Proof of Lemma 2.5.1

Proof. We have to show that $\bar{b}^2(k)$ is concave, which is equivalent to showing $\frac{\partial \bar{b}^2(k)}{\partial k^2} < 0$, where $\bar{b}(k) = \frac{1}{3-2u(z(k))}$. It holds that

$$\frac{\partial \bar{b}^2(k)}{\partial k^2} = 24\left(\frac{1}{3-2u(z(k))}\right)^4\left(\frac{\partial u(z(k))}{\partial k}\right)^2 + 4\left(\frac{1}{3-2u(z(k))}\right)^3\frac{\partial u(z(k))}{\partial k^2},$$

so then we need to show

$$6\left(\frac{1}{3-2u(z(k))}\right)\left(\frac{\partial u(z(k))}{\partial k}\right)^2 < -\frac{\partial u(z(k))}{\partial k^2}.$$

At this point we use the explicit expressions for the expected utility and its partial derivatives, in terms of k, which are given by:

$$\begin{aligned} u(z(k)) &= \frac{1}{2}\left(1 + k^{1/(1-k)} - (k^{1/(1-k)})^k\right); \\ \frac{\partial u(z(k))}{\partial k} &= -\frac{1}{2(1-k)}\mathrm{Log}(k)(k^{1/(1-k)})^k; \\ \frac{\partial^2 u(z(k))}{\partial k^2} &= \frac{\left(k^{1/(1-k)}\right)^k}{2(k-1)^3 k}\left((1-k+k\mathrm{Log}(k))^2 + (k-1)^2 k\mathrm{Log}^2(k)\frac{1}{1-k}\right). \end{aligned}$$

Substituting these expressions in the inequality above and simplifying, we obtain that $\bar{b}^2(k)$ is concave if and only if

$$0 < \overbrace{+2k(k-1)\mathrm{Log}(k) - (1-k)^2\left(2 - k^{1/(1-k)} + (k^{1/(1-k)})^k\right)}^{(*)} + \underbrace{k(k-1)\mathrm{Log}(k)\left(2 - k^{1/(1-k)} + (k^{1/(1-k)})^k\right) - k\mathrm{Log}^2(k)\left(2 - k^{1/(1-k)} + (3k-2)(k^{1/(1-k)})^k\right)}_{(**)}.$$

Showing that both $(*) > 0$ and $(**) > 0$ is enough to prove the desired result. On the one hand,

$$2k\mathrm{Log}(k)(k-1) > 2(k-1)^2\mathrm{Log}(k) > 2(k-1)^2 > 2(k-1)^2 - \left(k^{1/(1-k)} + (k^{1/(1-k)})^k\right)(k-1)^2,$$

which shows that $(*) > 0$. On the other hand, as $(k^{1/(1-k)})^k(3k-2) < 3$ for all $k > 2$, then for all $k > 3$ it holds that

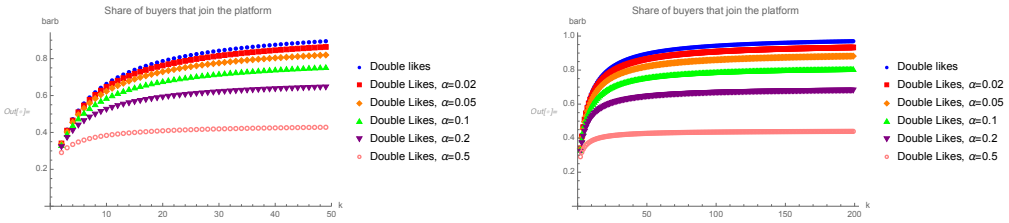
$$2(k-1) > 6\mathrm{Log}(k) > \mathrm{Log}(k)\left(2 - k^{1/(1-k)} + (k^{1/(1-k)})^k(3k-2)\right).$$

Checking that $(**)$ also holds for $k = 2$ is a mere numerical verification. □

Appendix 2.B Concavity of $\tilde{b}^2(k)$

In this paper, we conjecture that the continuous function $\tilde{b}^2(k)$ is a concave function of k for any $\alpha > 0$. These plots show that, up to $k = 200$, this is the case for some specific values of α. We present two different graphs to allow the reader to see in detail the tendency of the function. Once k is large enough, both $z_H(k)$ and $z_L(k)$ tend to $z(k)$ (or directly to 1), so that

$$\lim_{k\to\infty} \tilde{b}^2(k) = \frac{1}{1+\alpha}.$$



(a) k up to 50 **(b)** K up to 200

Figure 2.B.1. The concavity of $\tilde{b}^2(k)$.

References

Bedre-Defolie, Özlem, Bjorn Johansen, and Leonardo Madio. 2024. "Design and governance of quality on a digital platform." unpublished manuscript, [51, 52]

Belleflamme, Paul, and Martin Peitz. 2018. "Inside the engine room of digital platforms: Reviews, ratings, and recommendations." working paper, available at https://shs.hal.science/halshs-01714549/document, [55, 59]

Bolton, Gary, Ben Greiner, and Axel Ockenfels. 2013. "Engineering trust: Reciprocity in the production of reputation information." *Management science* 59(2): 265–85. [55]

Che, Yeon-Koo, and Johannes Hörner. 2018. "Recommender systems as mechanisms for social learning." *Quarterly Journal of Economics* 133(2): 871–925. [51]

Choi, Jay Pil, and Doh-Shin Jeon. 2023. "Platform design biases in ad-funded two-sided markets." *RAND Journal of Economics* 54(2): 240–67. [51]

Fradkin, Andrey, Elena Grewal, and David Holtz. 2021. "Reciprocity and unveiling in two-sided reputation systems: Evidence from an experiment on Airbnb." *Marketing Science* 40(6): 1013–29. [55]

Freimane, Marita. 2022. "Substituting away? The effect of platform bargaining regulation on content display." working paper, available at https://www.cresse.info/wp-content/uploads/2022/10/2022_ps7_pa1_Freimane.pdf, [52]

Gambato, Jacopo, and Martin Peitz. 2023. "Platform-Enabled Information Disclosure." working paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4596322, [50, 51, 61]

Gambato, Jacopo, and Luca Sandrini. 2023. "Not as good as it used to be: Do streaming platforms penalize quality?" *ZEW-Centre for European Economic Research Discussion Paper*, (23-032): [52]

Gómez-Uribe, Carlos A, and Neil Hunt. 2015. "The netflix recommender system: Algorithms, business value, and innovation." *ACM Transactions on Management Information Systems (TMIS)* 6(4): 1–19. [48]

Hagiu, Andrei, and Julian Wright. 2015. "Multi-sided platforms." *International Journal of Industrial Organization* 43: 162–74. [51]

Haraguchi, Junichi, and Yuta Yasui. 2023. "Effects of rating systems on platform markets: Monopolistic competition approach." working paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4343583, [52]

Hopenhayn, Hugo, and Maryam Saeedi. 2023. "Optimal information disclosure and market outcomes." *Theoretical Economics* 18(4): 1317–44. [52]

Hui, Xiang, Zekun Liu, and Weiqing Zhang. 2023. "From high bar to uneven bars: The impact of information granularity in quality certification." *Management Science* 69(10): 6109–27. [52]

Jeon, Doh-Shin, and Nikrooz Nasr. 2016. "News aggregators and competition among newspapers on the internet." *American Economic Journal: Microeconomics* 8(4): 91–114. [52]

Lafky, Jonathan. 2014. "Why do people rate? Theory and evidence on online ratings." *Games and Economic Behavior* 87: 554–70. [48]

Lizzeri, Alessandro. 1999. "Information revelation and certification intermediaries." *RAND Journal of Economics*, 214–31. [51]

Pommeranz, Alina, Joost Broekens, Pascal Wiggers, Willem-Paul Brinkman, and Catholijn M Jonker. 2012. "Designing interfaces for explicit preference elicitation: a user-centered investigation of preference representation and elicitation process." *User Modeling and User-Adapted Interaction* 22: 357–97. [48]

Tadelis, Steven. 2016. "Reputation and feedback systems in online platform markets." *Annual Review of Economics* 8: 321–40. [49, 51]

Teh, Tat-How. 2022. "Platform governance." *American Economic Journal: Microeconomics* 14 (3): 213–54. [50, 51]

Vatter, Benjamin. 2022. "Quality disclosure and regulation: Scoring design in medicare advantage." *Available at SSRN 4250361*, [52]

Vellodi, Nikhil. 2018. "Ratings design and barriers to entry." working paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3267061, [52, 68]

Chapter 3

Network Effects on Information Acquisition by DeGroot Updaters

3.1 Introduction

Information is key to making decisions. Nowadays, social networks have a significant impact on information processes. We discuss various issues with family, friends, and colleagues, affecting their opinions and shaping our own. Random conversations about an upcoming election, the job market, or stock market performances can influence our beliefs. Breakthrough news spreads rapidly, and individuals are constantly updating their opinions. Apart from this social supply of information, individuals can learn privately, such as by searching on the internet or consulting a book. Therefore, it is essential to understand how individuals behave when they seek to acquire information to make decisions in situations where both social communication and private learning take place. To what extent do people exert effort themselves, and to what extent do they rely on others?

In this chapter, we propose a model of information acquisition in networks in which individuals are boundedly rational, behaving as mechanical updaters when it comes to learning. With this in hand, they decide how much to invest in a

* Support by the German Research Foundation (DFG) through CRC TR 224 (Project B05) and funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 949465) are gratefully acknowledged. I thank Sven Rady, Francesc Dilmé, Alexander Frug, Alexander Winter, Simon Block, Justus Preusser, Axel Niemeyer, an anonymous referee and participants at the seminars at University of Bonn and Universitat Pompeu Fabra for their helpful comments and discussions. This article is published at Economic Theory and can be accessed at https://link.springer.com/article/10.1007/s00199-024-01568-7

costly information source to improve their knowledge of the state of the world. Mechanical updating here consists of agents merely taking weighted averages of the signals received—the so-called DeGroot updating rule from DeGroot (1974).

Despite considering boundedly-rational agents, we still apply the concept of Nash equilibrium at the stage where they determine their information acquisition. This is done in the spirit of an evolutionary concept of Nash equilibrium. An evolutionary model consists of a large population of boundedly-rational players playing some game repeatedly over time (Mailath, 1998). Evolutionary theory shows that such players eventually learn to play Nash equilibrium,[1] even in the absence of perfect rationality. The crucial assumption is that more successful behaviors become more prevalent due to a combination of imitation and the failure of unsuccessful behaviors.[2] In our model, boundedly-rational agents that update mechanically face the problem of provision of a local public good. The task of gathering information for subsequent decision-making recurs numerous times throughout an individual's life. In the spirit of evolutionary theory, we think of boundedly-rational agents who, despite their cognitive constraints, have learned to reach Nash equilibrium outcomes through their choices.

To provide an intuition for the formal model, consider an agent who wishes to become more informed about a particular issue. We assume that she has some prior knowledge, and that informative conversations take place in her neighborhood—for example, at the office. There, each colleague exerts a different and fixed influence that shapes the agent's final opinion. Anticipating this, she decides how much effort to spend on private learning. Given that learning tools are similar among neighbors, we assume a positive correlation when it comes to private learning signals. Hence, each agent faces a problem of information acquisition in which information is a local public good. Individuals have to decide how much to invest in private learning, knowing that free social learning will take place later. Depending on the substitutability between information acquired personally and information acquired by others, but also on the neighbors' choices, agents will raise or lower their learning efforts. Free-ride behaviors will arise.[3]

1. In particular, the central notion in evolutionary game theory is that of *evolutionary stable strategy*, and theory shows that any symmetric strict Nash equilibrium is indeed an evolutionary stable strategy. See Mailath (1998) or Samuelson (2002) for an overview.

2. This is also discussed by Aumann (1997), asserting that ordinary people, in their daily activities, do not consciously adhere to rationality but evolve "rules of thumb" through evolution. If these rules prove effective, they proliferate and multiply, eventually reaching the equilibrium that strict rationality would have predicted.

3. For an axiomatic characterization of costly information acquisition processes, see Duraj and Lin (2022).

This chapter provides three main contributions. First, we analyze and characterize the equilibria arising in the model. Depending on the network structure and their positions, agents will contribute with some learning or completely free-ride. In principle, there are multiple equilibria, and all of them can be calculated. A sufficient condition for equilibrium uniqueness is our second contribution. If this condition does not hold, the equilibria computations run in exponential time. Equilibrium uniqueness is controlled by the lowest eigenvalue of a matrix given by the network. This eigenvalue essentially captures how two-sided the corresponding network is, that is, whether agents can be divided into two sets with many links between them but just a few within. In a game of strategic substitutabilities, when an agent increases her effort, her neighbors decrease theirs in response, so that the neighbors' neighbors have to increase, and so on. When the network is two-sided, these direct effects accumulate and lead to several distinct equilibria. However, if the lowest eigenvalue is sufficiently large, the network will not be two-sided enough for the actions to rebound, and the equilibrium will be unique. Finally, we provide a welfare analysis. From a social (utilitarian) perspective, every agent would be more informed than she is in equilibrium. To satisfy this demand, at least the influential agents (those agents from which the others get the majority of information) have to increase their contribution. If the network is *too unbalanced*, this becomes a burden and the welfare of the influentials decreases. In general, the utilitarian optimum does not Pareto-dominate the equilibrium outcome.

The choice of the updating rule, i.e., how individuals process and incorporate the information received, is a relevant decision when trying to model social learning. One has to decide whether to employ the fully rational Bayesian focus or the naive, boundedly-rational approach, mainly represented by the already mentioned DeGroot rule. Quoting Acemoglu and Ozdaglar (2011), "which type of approach is appropriate is likely to depend on the specific question being investigated". We argue here that the DeGroot updating rule fits best within our context.

Bayesian updating requires an unrealistic cognitive demand for learning in large networks. However, the DeGroot rule provides a convenient alternative, given its simplicity and lack of restrictive requirements. In a simultaneous setting, Bayesian agents who receive Gaussian private signals behave like DeGroot updaters when subject to persuasion bias, as shown by DeMarzo, Vayanos, and Zwiebel (2003). In fact, if the game is one-shot (as it is in this chapter), persuasion bias is not even necessary for such a result to hold. Still, their model deviates slightly from standard rational assumptions, as neighbors' signal precision is unknown. The relationship between Bayesian and DeGroot rules, especially for one-shot games, supports our model and is further analyzed in the Appendix. Nonetheless, after the first period, a pure Bayesian (not suffering from persuasion bias) would adjust for the information buried in the network, while DeGrootian

agents would not.[4] The literature on networks has widely used the DeGroot rule in different settings. Golub and Jackson (2010) show that under some mild conditions on connectedness and influence, DeGroot agents converge to the belief that would result from the full aggregation of everyone's signal. Both Golub and Jackson (2012), devoted to study homophily, and Acemoglu, Ozdaglar, and ParandehGheibi (2010), which analyzes the tension between the spread of misinformation and information aggregation, also reflect how convenient DeGroot updating is for large networks analysis. However, the major drawback of the rule is that the choice of weights might seem arbitrary, particularly when communication lasts longer than one period. Furthermore, the assumption that everyone is informed at the outset may be too demanding. Banerjee, Breza, Chandrasekhar, and Mobius (2021) adapted the rule to sparse signals to address this issue.

This having been said, the empirical evidence heavily supports DeGroot updating. Various papers confront it against Bayesian learning in an experimental setting, concluding that it approximates better people's information aggregation rules (see Corazzini, Pavesi, Petrovich, and Stanca (2012), Mueller-Frank and Neri (2013), Grimm and Mengel (2020), Chandrasekhar, Larreguy, and Xandri (2020)). Although there is no definitive approach, many recent papers tend to use a boundedly rational model for both sequential and simultaneous settings. For example, Dasaratha and He (2020) assume that agents neglect redundancies of information and then aggregate heuristically, and Mueller-Frank and Neri (2021) consider a large class of boundedly rational or quasi-Bayesian rules, respectively.

Although modelling learning through a mechanical updating rule is overly simplistic, it allows us to isolate the network effects, which is the primary concern of this chapter. Furthermore, we argue that assuming exogenous and fixed weights reflects human behavior. The influence that our neighbors exert on a concrete topic is almost predetermined. A wide range of factors such as past interactions, trustworthiness, and expected level of knowledge defines an influence level before communication occurs. Similarly, it seems sensible that agents can endogenously set the influence of their own private learning on their views: the more time devoted to researching, the more reliable the agent perceives it to be. Thus, the expenditure of costly attention will reduce player-specific noise.

Galeotti and Goyal (2010) is a key paper in the literature on information acquisition in networks. In this paper, network-placed agents strategically select their links to access the information held by their neighbors. Every equilibrium displays the so-called "law of the few": the majority of individuals tend to get most

4. See Molavi, Tahbaz-Salehi, and Jadbabaie (2018) for an axiomatic foundation of the DeGroot rule under imperfect recall.

of their information from a tiny subset of the group, the influentials. Our model shows that this result holds true for networks where a subset of agents, the populars, has a significantly higher weight than the rest, such as the core-periphery network. In such networks, popular agents become influential and acquire most of the information, while the others free-ride. This finding contrasts with Banerjee et al. (2021), where the sparse-signals structure indicates that being popular alone is insufficient for being influential. However, two assumptions in Galeotti and Goyal (2010) differ from our model: links are endogenous, allowing an agent to reach any other individual in a potentially large network, and homogeneous, meaning perfect substitutability. Network effects on information acquisition have also been analyzed from a Bayesian perspective. In Myatt and Wallace (2019), rational agents acquire information about the state of the world from sources that provide noisy signals. Paying costly attention reduces noise, and signals are possibly correlated across players, similar to our model. However, incentives differ as agents not only want to match the state of the world but also care about coordination asymmetrically. Furthermore, there is no communication stage. The player's centrality (in the sense of Bonacich) and correlations determine information acquisition, but centrality entails less expenditure, in contrast to Galeotti and Goyal (2010) and this chapter. Finally, Denti (2017) introduces the concept of entropy reduction to study how players endogenously acquire costly information to decrease their uncertainty about fundamentals. Network effects induce externalities in information acquisition and are a source of multiple equilibria.

Regarding equilibrium analysis, our work closely follows that of Bramoullé, Kranton, and D'Amours (2014). Following previous attempts in the literature to characterize conditions for equilibrium in linear games of strategic complements (cf. Ballester, Calvó-Armengol, and Zenou (2006)) and strategic substitutes (especially in public goods, cf. Bramoullé, Kranton, et al. (2007)), the authors showed that equilibrium uniqueness and stability depend on the lowest eigenvalue of the network matrix.[5] This is dependent on the two-sidedness of the network. Although our chapter differs in setting and motivation, the best response function derived from our model is similar to that of Bramoullé, Kranton, and D'Amours (2014). Consequently, the result regarding the lowest eigenvalue characterizing equilibrium uniqueness is also similar. However, their model assumes that agents' contributions are reciprocal and weighted equally, which differs from our assumptions. This has two consequences. First, the potential theory introduced in Monderer and Shapley (1996), on which Bramoullé et al. base their results, cannot be applied here; second, a wider range of networks can be analyzed. Nevertheless, if

5. Using the lowest eigenvalue of the network matrix to determine the uniqueness of an outcome is a technique often employed in the social networks literature. See, for example, Melo (2022).

we restrict our setting to symmetric, homogeneous networks, an almost equivalent condition arises. Finally, our chapter is also related to Bramoullé, Kranton, et al. (2007) model of pure public goods in exogenous networks, where again all contributions are weighted equally and there is perfect substitutability. In that model, the authors find that multiple equilibria typically exist, and there is always one in which some individuals contribute while others free-ride. This equilibrium is typically unique.

The rest of the chapter is organized as follows. Section 3.2 describes and analyzes our model, Section 3.3 studies the equilibria, provides a uniqueness condition and presents some examples, and Section 3.4 analyzes the model from a social planner perspective. Section 3.5 briefly introduces a dynamic version of the model, and Section 3.6 concludes.

3.2 Model

We consider a finite set of n agents interacting via a social network represented by an $n \times n$ matrix $\mathbf{G} = (g_{ij})$, which is predetermined and stochastic: the entries in each row are non-negative and sum to one. Interactions need not be symmetric or two-sided, so in general $g_{ij} \neq g_{ji}$ and $g_{ij} > 0$ does not imply $g_{ji} > 0$.

Each agent holds a private signal s_i about a common underlying state of the world $\mu \in \mathbb{R}$, drawn independently from a normal distribution with mean μ and variance $\sigma^2 > 0$. There are two learning resources available to improve this signal, presented in the order in which they become accessible to the agent: active private learning from a more informative but costly source, and social learning from neighbors. The first resource involves drawing a signal $\mathscr{I}_i$ from a normal distribution with mean μ and variance $\tilde{\sigma}^2 < \sigma^2$, while the second resource involves aggregating the signals of the agent's direct neighbors in the network.

Both types of learning take the form of DeGroot updating of signals, following DeGroot (1974). Agents take a weighted average of their signals, i.e., they aggregate several indicators in just one. In the case of private learning, agent i decides the weights in the convex combination between s_i and $\mathscr{I}_i$. The costly signal $\mathscr{I}_i$ receives weight $x_i \in [0,1]$ at linear cost $x_i c$ with $c > 0$. Costly signals are positively correlated across agents, $\mathrm{Cov}(\mathscr{I}_i, \mathscr{I}_j) = \alpha > 0$ for all i, j. The original private signals are independent across agents and independent of all costly signals. Regarding social learning, agent i takes the weighted average of her neighbors' signals and her own. Weights are exogenously[6] given by the network matrix, and they represent influence or trust: agent i listens to agent j precisely at *intensity* g_{ij}.

6. Rational learners might adjust the weights based on neighbors' information levels, as discussed in Galeotti and Goyal (2010). However, in this case, we want to focus on situations

The mechanical updating process described can be viewed as active learning with attention costs for boundedly rational agents. In addition to normal signals, it can also be interpreted from a Bayesian perspective, as demonstrated in DeMarzo, Vayanos, and Zwiebel (2003). Agents assign subjective precisions π_{ij} to each other, attempting to estimate the true precision of their signals. If the signals are independent and normal, Bayesian updating is equivalent to DeGroot updating, with weights given by $\frac{\pi_{ij}}{\sum_{j=1}^{n}\pi_{ij}}$ for social learning.[7] A similar argument applies to the active learning process; see the Appendix for a motivation of the present framework in terms of quasi-Bayesian updating as defined in DeMarzo, Vayanos, and Zwiebel (2003).

In the following, we use the term "beliefs" to refer to the most recently updated signal an agent holds about μ. A precise description of the learning process is as follows: The agent receives $s_i \sim \mathcal{N}(\mu, \sigma^2)$ and decides how much to spend on learning $\mathscr{I}_i \sim \mathcal{N}(\mu, \tilde{\sigma}^2)$. Once x_i is selected, the belief becomes $p_i = (1-x_i)s_i + x_i\mathscr{I}_i$ at cost $x_i c$. Finally, the social communication stage yields beliefs

$$\sum_{j=1}^{n} g_{ij}p_j = \sum_{j=1}^{n} g_{ij}\left((1-x_j)s_j + x_j\mathscr{I}_j\right).$$

Note that if i and j are not neighbors, $g_{ij} = 0$, so summing over i's neighbors is equivalent to summing over all n individuals. At this point, only one communication stage is assumed, but considering t stages would imply the substitution of $\mathbf{G}$ by $\mathbf{G}^t$, as shown in Section 3.5.

Agent i aims to obtain the most precise belief about μ at minimum cost, as deviations are penalized through a quadratic loss function. This is specified in the payoff function

$$-\left(\mu - \sum_{j=1}^{n} g_{ij}p_j\right)^2 - x_i c = -\left(\mu - \sum_{j=1}^{n} g_{ij}((1-x_j)s_j + x_j\mathscr{I}_j)\right)^2 - x_i c.$$

Although agent i is a naive, mechanical learner, we assume, based on evolutionary theory, that she is capable of reaching Nash equilibrium outcomes. Specifically, deciding how much to contribute to a public good is a typical example of a process in which boundedly-rational agents evolve toward Nash equilibrium outcomes in the long run (Mailath, 1998). Hence, we allow agent i to form expectations and best respond to others' choices, *as if* she were rational at this stage. She chooses

---

where weights are pre-determined for a naïve learner due to past interactions, influence, or reputation, and cannot be modified.

7. If $g_{ij} = 0$, then $\pi_{ij} = 0$.

the amount x_i that maximizes her expected utility:

$$\max_{x_i\in[0,1]}\left\{\mathbb{E}\left[-\left(\mu-\sum_{j=1}^{n}g_{ij}((1-x_j)s_j+x_j\mathscr{I}_j)\right)^2\right]-x_ic\right\}. \tag{3.2.1}$$

3.3 Equilibrium

The equilibrium concept used in this model is Nash equilibrium, where each agent i chooses her information level x_i by best responding to others' equilibrium choices. It is important to note that $\mathbb{E}[s_i]=\mathbb{E}[\mathscr{I}_i]=\mu$ for all i. Additionally, every pair of signals is independent except for $\mathscr{I}_i$ and $\mathscr{I}_j$. As a result, $\mathbb{E}\left[\sum_{j=1}^{n}g_{ij}(x_j\mathscr{I}_j+(1-x_j)s_j)\right]=\mu$, while $\mathrm{Var}(x_j\mathscr{I}_j+(1-x_j)s_j)=x_j^2\tilde{\sigma}^2+(1-x_j)^2\sigma^2$ and $\mathrm{Cov}(x_j\mathscr{I}_j+(1-x_j)s_j,x_k\mathscr{I}_k+(1-x_k)s_k)=\alpha x_jx_k$. These equalities, along with the payoff equation, imply that only second moments matter. In fact,

$$\mathbb{E}\left[-\left(\mu-\sum_{j=1}^{n}g_{ij}(x_j\mathscr{I}_j+(1-x_j)s_j)\right)^2\right]=-\mathrm{Var}\left[\sum_{j=1}^{n}g_{ij}(x_j\mathscr{I}_j+(1-x_j)s_j)\right].$$

Using that for any sequence of random variables $\{\tilde{X}_j\}_{j=1}^n$ it holds that $\mathrm{Var}(\sum_{j=1}^n\tilde{X}_j)=\sum_{j=1}^n\mathrm{Var}(\tilde{X}_j)+2\sum_{j=1}^n\sum_{k=1}^{j-1}\mathrm{Cov}(\tilde{X}_j,\tilde{X}_k)$, the maximization problem for agent i can be simplified as follows:

$$\max_{x_i\in[0,1]}\left\{-\tilde{\sigma}^2\sum_{j=1}^{n}g_{ij}^2x_j^2-\sigma^2\sum_{j=1}^{n}g_{ij}^2(1-x_j)^2-2\alpha\sum_{j=1}^{n}\sum_{k=1}^{j-1}g_{ij}g_{ik}x_jx_k-cx_i\right\}. \tag{3.3.1}$$

The objective is strongly concave in the choice variable. The first order condition for an interior solution yields

$$x_i=\frac{2\sigma^2-c/g_{ii}^2}{2(\tilde{\sigma}^2+\sigma^2)}-\frac{\alpha}{g_{ii}(\tilde{\sigma}^2+\sigma^2)}\sum_{j\neq i}g_{ij}x_j.$$

Note that this expression is bounded above by 1 but could be negative. As $x_i\in[0,1]$ by assumption, the optimal choice of active learning for agent i given others' choices x_{-i} is

$$x_i^*=\max\left\{0,\frac{2\sigma^2-c/g_{ii}^2}{2(\tilde{\sigma}^2+\sigma^2)}-\frac{\alpha}{g_{ii}(\tilde{\sigma}^2+\sigma^2)}\sum_{j\neq i}g_{ij}x_j\right\}.$$

This best response function is similar to the one obtained when solving a maximization problem in a local public goods setting. Games of negative externalities or Cournot competition also yield similar forms. The only difference is that here,

δ is divided by g_{ii}, a parameter that varies across agents. In the other cases, the substitutability factor is the same for all agents.

Note that only the weighted out-degree matters for information acquisition, but not the weighted in-degree.[8] In other words, agents care about who they are listening to (the g_{ij}s), but not who listens to them (the g_{ji}s). Furthermore, if $g_{ii} = 0$, then $x_i^* = 0$ trivially, as active learning is a waste of resources for someone who does not assign positive weight to herself. Therefore, without loss of generality we can assume $g_{ii} > 0$. By setting

$$\bar{x}_i = \frac{2\sigma^2 - c/g_{ii}^2}{2(\tilde{\sigma}^2 + \sigma^2)},$$

and

$$\delta = \frac{\alpha}{(\tilde{\sigma}^2 + \sigma^2)},$$

we obtain

$$x_i^* = \max\left\{0, \bar{x}_i - \frac{\delta}{g_{ii}} \sum_{j \neq i} g_{ij} x_j\right\}.$$

Here, information refers to individuals' costly learnt signals and is a local public good. Each agent benefits from others' private learning via network communication. The quotient $\frac{\delta}{g_{ii}}$ scales the benefit and indicates the substitutability between an agent's and her neighbors' information. Agent i seeks to reach at least the *information target* $\bar{x}_i$ through a combination of her own information and her neighbors'. If the weighted contributions of the others are enough ($\frac{\delta}{g_{ii}} \sum_{j \neq i} g_{ij} x_j > \bar{x}_i$), then she will not spend on private learning, and $x_i^* = 0$. If not, she will make up the difference, and $x_i^* > 0$.

Let us analyze the scale factor $\frac{\delta}{g_{ii}}$, which measures the substitutability between the information purchased by an agent and her neighbors. On the one hand, $\frac{1}{g_{ii}}$ reflects how important others' contributions are to the particular agent i. If g_{ii} is small, almost all attention is paid to the neighbors, so their information matters considerably. In contrast, if g_{ii} is close to one, agent i essentially listens to herself. On the other hand, $\delta = \frac{\alpha}{\sigma^2 + \tilde{\sigma}^2}$ reflects the quality of the neighbors' information. The parameter α indicates how much information one can extract from others. Consequently, the higher α, the less information is purchased. The sum $\sigma^2 + \tilde{\sigma}^2$ expresses the overall level of uncertainty. If it grows, the incentives for an agent to

8. The out-degree of agent i is the total weight of links directed away from her. The in-degree is the total weight of links directed to her.

increase her information level also grow. Note also that $\delta \in \left[0, \frac{1}{2}\right]$ by the Cauchy-Schwarz inequality.[9] Therefore, for any $g_{ii} \geq 1/2$, the scale factor is always lower than one. This is not surprising: if an agent listens to herself more than to others, the information that she acquires matters more.

The information target $\bar{x}_i$ indicates how well-informed each agent aims to be. The more precise $\mathscr{I}_i$ is in expectation terms—the lower $\tilde{\sigma}^2$—, the higher $\bar{x}_i$ the agent wants to attain. Additionally, the degree of an agent's own attention, represented by g_{ii}, matters: acquiring information through costly learning is more profitable if the agent puts a higher weight on herself when updating. An increase in costs makes information acquisition less attractive. It is worth noting that $\bar{x}_i = 1$ only if $\tilde{\sigma}^2 = 0$ (i.e., $\mathscr{I}_i = \mu$ and it is a perfect signal) and $c = 0$. In all other cases, a convex combination of s_i and $\mathscr{I}_i$ is preferred. It is more convenient for the agent to have two independent signals, even if one is much more informative than the other, than just one. Therefore, she does not want to get rid of s_i entirely and sets $x_i^* < 1$.

It has been shown that $x_i^* \in [0,1]$. The combined best response function of the players maps $[0,1]^n$ to itself and is continuous. Brouwer's fixed point theorem guarantees the existence of an equilibrium.

Proposition 3.3.1. *The game of information acquisition by DeGroot updaters has at least one Nash equilibrium.*

We should pay special attention to the limiting case of uncorrelated costly signals, i.e., $\alpha = 0$. Since the correlation between signals is zero, agents cannot extract any information from each other. The equilibrium analysis is then trivial: as $\delta > 0$, each agent chooses the information level

$$x_i^* = \max\{0, \bar{x}_i\}.$$

Since $\bar{x}_i \leq 1$, x_i^* is well-defined. Moreover, $\bar{x}_i > 0$ if and only if $g_{ii} > \sqrt{\frac{c}{2\sigma^2}}$. This is the target for active learning: every individual that weighs themselves more than $\sqrt{\frac{c}{2\sigma^2}}$ will choose a positive x_i^*, independently of the network. Due to the lack of strategic substitutability, equilibrium is unique.

9. In fact, this inequality implies $0 \leq \alpha \leq \tilde{\sigma}^2$, and hence

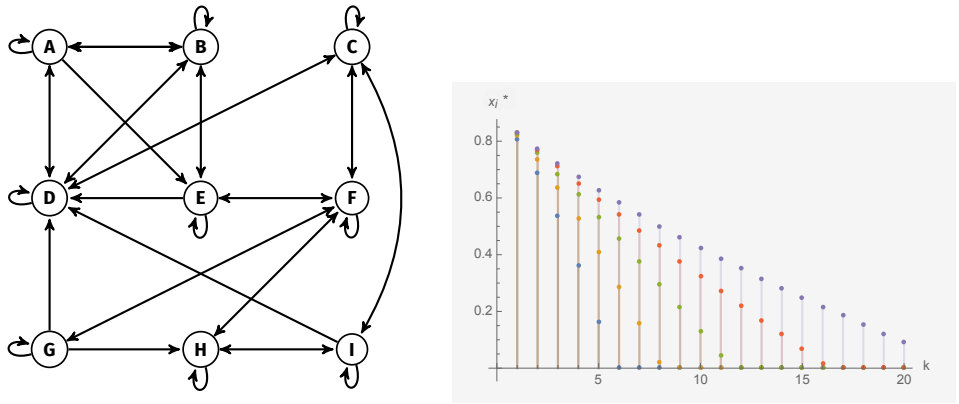
$$0 \leq \delta \leq \frac{\tilde{\sigma}^2}{\sigma^2 + \tilde{\sigma}^2} = \frac{1}{2}.$$

3.3.1 Examples

With the existence of equilibrium proved, and before further general analysis, some prominent networks and their equilibria are discussed as illustrations.

The first class we consider is the **k-regular graphs with homogeneous weights**. This class consists of structures with n agents, each having k neighbors. All connections have identical weight, such that $g_{ij} = \frac{1}{k}$ for all i and j. Such a network could represent a small community where each member listens to everyone else, and influences are homogeneous. Complete networks in which every agent has the same degree are a subset of regular graphs. Other examples of symmetric structures represented by regular graphs are big societies of n individuals who cluster in small k-neighborhoods and the circle. The unique equilibrium of the k-regular graph with homogeneous weights is given by

$$x_i^* = \frac{2\sigma^2 - ck^2}{2(\sigma^2 + \tilde{\sigma}^2)(-\delta + \delta k + 1)} \quad \text{if } c \leq \frac{2\sigma^2}{k^2}; \quad x_i^* = 0 \text{ otherwise.}$$



(a) 4-regular graph. **(b)** Information acquisition, k-regular graph.

Figure 3.3.1. Regular graphs.

As the number of neighbors grows, incentives to acquire information decrease. The more signals an agent listens to, the better, and active learning loses importance, as shown in Figure 3.3.1b. The parameters are set to $\sigma^2 = 5$, $\tilde{\sigma}^2 = 1$, and $\delta = 0.07$. Each color corresponds to a different cost of $\mathscr{I}_i$, ranging from $c = 0.3$ to $c = 0$. Note that $x_i^* = 0$ as soon as $c \geq \frac{10}{k^2}$. The non-symmetric example displayed in Figure 3.3.1a, which features nine agents of out-degree four, also has the above equilibrium. This is implied by the fact that only the acquisition choices of those whom the individual listens to matter. In other words, only the out-degree matters. Differences between networks with the same out-degree for all agents but different in-degrees will appear in the socially optimal allocation, where the in-degree

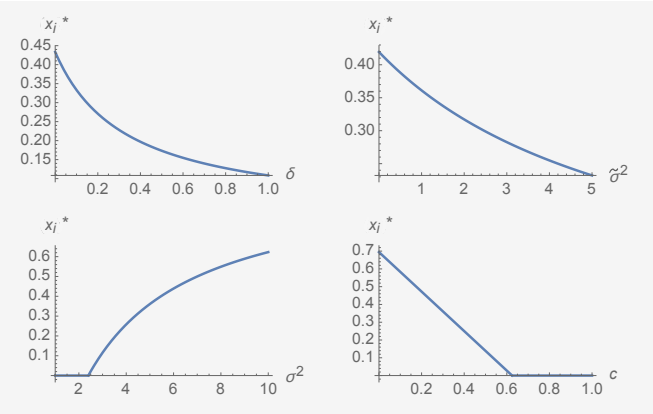


Figure 3.3.2. Comparative statics for the 4-regular graph.

also conditions behavior. This will be shown in Section 3.4. Comparative statics are presented in Figure 3.3.2 for a regular graph with four neighbors and weights $\frac{1}{4}$. The cost is $c = 0.3$ and the rest of the parameters are as above.

The second class is that of **stars**, where one prominent agent (the hub) is connected to every other agent (the spokes).[10] The spokes, in turn, are connected only to the hub and themselves. A department in a firm, with a supervisor and some employees, is a leading example. An auction with an auctioneer in the center could be another example. Assume that the hub puts the same weight on everyone, and the spokes put weight ε on the hub. Therefore, the information levels in equilibrium for a society of n agents are given by the following equations:

$$x_H^* = \frac{2\sigma^2(1-\varepsilon)^2((n-1)\delta-1) - c(\delta(n-1)-(1-\varepsilon)^2n^2)}{2(\sigma^2+\tilde{\sigma}^2)(1-\varepsilon)((n-1)\varepsilon\delta^2+\varepsilon-1)} \text{ if } c \leq \frac{2\sigma^2(1-\varepsilon)^2((n-1)\delta-1)}{(\delta(n-1)-(1-\varepsilon)^2n^2)};$$

$$x_H^* = 0 \text{ otherwise,}$$

$$x_S^* = \frac{2\sigma^2(1-\varepsilon)(1-\varepsilon-\delta\varepsilon) - c(1+\delta(\varepsilon-1)\varepsilon n^2)}{2(\sigma^2+\tilde{\sigma}^2)(1-\varepsilon)((n-1)\varepsilon\delta^2+\varepsilon-1)} \text{ if } c \leq \frac{2\sigma^2(1-\varepsilon)(1-\varepsilon-\delta\varepsilon)}{(1+\delta(\varepsilon-1)\varepsilon n^2)};$$

$$x_S^* = 0 \text{ otherwise.}$$

The star graph for five agents is shown in Figure 3.3.3, where node A represents the hub (H) and nodes B, C, D, and E are the spokes (S). The blue links have a weight of $\frac{1}{5}$, while the bold links have a weight of ε. The hub shares her attention equally, while the spokes put weight ε on her. The information acquisition

10. This particular network structure has been widely studied. For example, Jiménez-Martínez (2015) shows that in a dynamic model of communication, when the hub is not influenced by the spokes, even Bayesian agents are unable to reach correct limiting beliefs.

levels as a function of ε are also illustrated below, using the same parameter values as before: $c = 0.3$, $\sigma^2 = 5$, $\tilde{\sigma}^2 = 1$, and $\delta = 0.07$. When ε is small, the spokes need to invest significantly in learning. However, as ε grows, the hub's signal gains importance, and active learning becomes less valuable for the spokes. The hub responds to this behavior by adjusting her learning efforts. When ε is small, she can rely on the spokes to aggregate some signals, so only a modest amount of active learning is necessary. However, when ε grows large, the hub invests significantly more in learning to compensate for the drop in the spokes' contribution. Similarly, in a department of a firm, the supervisor reacts to employees' expertise, and the employees invest in learning only when it is useful. If they only follow the supervisor's orders, there are no incentives for them to learn independently.

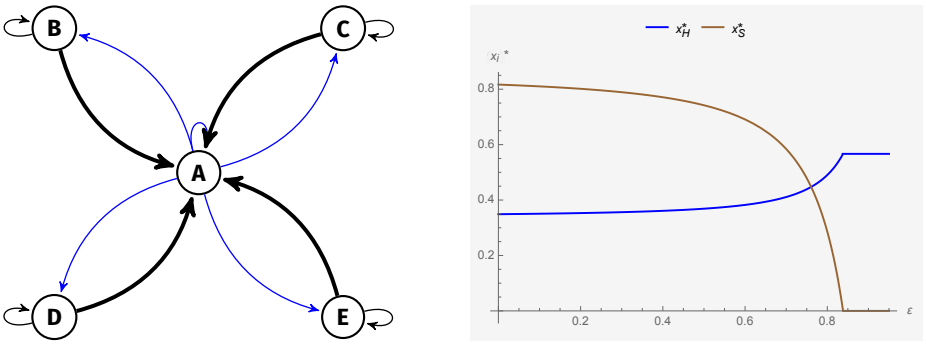


Figure 3.3.3. Star with five agents.

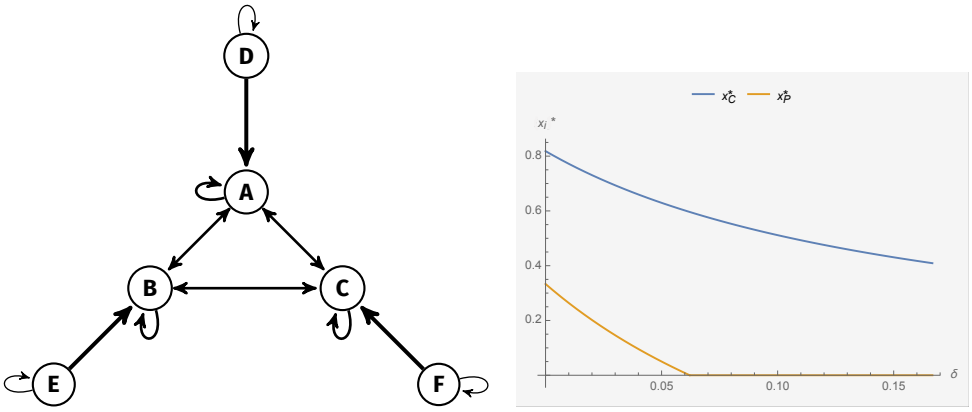
It is worth noting that the complete graph, in which each agent weights herself as ε and the other $n-1$ agents as $\frac{1-\varepsilon}{n-1}$, yields the same levels of information acquisition as the star with an own weight of ε.

We now move on to the class of **core-periphery** networks, which are characterized by a "dense, cohesive core and a sparse, unconnected periphery", as described in Borgatti and Everett (2000). Many relevant economic networks exhibit this structure, such as the lending behavior of banks (Fricke and Lux, 2015) or international trade networks (Fagiolo, Reyes, and Schiavo, 2010). Another example is the structure arising in Galeotti and Goyal (2010): the "few" constitute the core while the rest of the network (the periphery) free-rides on them. A particular case in which the core is formed by three individuals who share their attention equally and three periphery agents who listen to one core agent each is shown in Figure 3.3.4a. The periphery agents put almost all of their weight (in this case, $\frac{9}{10}$) on the core agents, which causes their acquisition levels to rapidly drop to zero as δ increases. As soon as there is a minimal amount of substitutability, the periphery agents stop purchasing. On the other hand, the core agents acquire abundant information, as the core acts as a k-regular network for them and it is independent from the periphery. The results for k-regular networks hold and hence, the larger

the core, the less information its agents acquire. The acquisition choices are given by:

$$x_C^* = \frac{2\sigma^2 - 9c}{2(\sigma^2 + \tilde{\sigma}^2)(2\delta + 1)} \text{ if } c \leq \frac{2\sigma^2}{9}; \quad x_C^* = 0 \text{ otherwise,}$$
$$x_P^* = \max\left\{0, \frac{2\sigma^2 - 100c}{2(\sigma^2 + \tilde{\sigma}^2)} - \frac{9\delta(2\sigma^2 - 9c)}{2(\sigma^2 + \tilde{\sigma}^2)(2\delta + 1)}\right\}.$$

The cost c is set to $c = 0.06$ in this example to illustrate the decay in the periphery agents' acquisition levels. A value of $c = 0.3$ as above would lead to no acquisition even if $\delta = 0$, as periphery agents barely weight themselves.



(a) Core-periphery for six agents. **(b)** Information acquisition as a function of δ.

Figure 3.3.4. Core-periphery networks.

Next we discuss the **criminal network** from Ballester, Calvó-Armengol, and Zenou (2006). The authors use the network from Figure 3.3.5 to highlight the fact that influence is not necessarily equivalent to the number of connections (degree). They identify the key criminal as agent A, who, when removed, leads to the highest aggregate reduction in crime, despite not having the highest degree. Agent A plays a crucial role in bridging two fully interconnected communities of five criminals each. In terms of information acquisition, Figure 3.3.6 shows the active learning choices for this network. Assuming equal shares of attention (i.e. each agent listens equally to every neighbor), there are three different kinds of individuals: agent A, agents B, F, G and K (referred to as the *B*-class) and agents C, D, E, I, H and J (*C*-class). Apart from listening to themselves, agent A listens to the four *B*-class agents, *B*-class agents listen to agent A, one *B*-class individual and three *C*-class individuals, and *C*-class agents listen to two *B*-class individuals and two *C*-class individuals. Interestingly, the most influential agent for Ballester et al. (agent A) is also the most informed. The *B*-class agents tend to free-ride on

the rest. The best reply functions are:

$$\begin{aligned} x_A^* &= \max\{0, \bar{x}_A - 4\delta x_B\}, \\ x_B^* &= \max\{0, \bar{x}_B - \delta(x_A + x_B + 3x_C)\}, \\ x_C^* &= \max\{0, \bar{x}_C - 2\delta(x_B + x_C)\}. \end{aligned}$$

Classes A and C respond to class B's choice, which is small in comparison. Information acquisition choices are shown in Figure 3.3.6 as a function of δ. The parameters are set as in the core-periphery example. Both A-class and C-class agents listen to five individuals each, so $\bar{x}_A = \bar{x}_C$. However, as B-class agents listen to six neighbors and weigh their own signal less, it holds that $\bar{x}_B < \bar{x}_A$. In the extreme case of $\delta = 0$, A-class agents and C-class agents purchase the same quantity, slightly more than B-class individuals. As soon as δ increases, B-class agents take advantage of substitutability and free-ride on A and the C-class agents. Agent A has only B-class neighbors, so although she extracts information from them, she has to make up the difference. In contrast, C-class agents have some C-class neighbors, so the free-riding behavior of B-class agents does not affect them as severely. Figure 3.3.6 shows that x_A^* is higher than x_C^* for all $\delta > 0$.

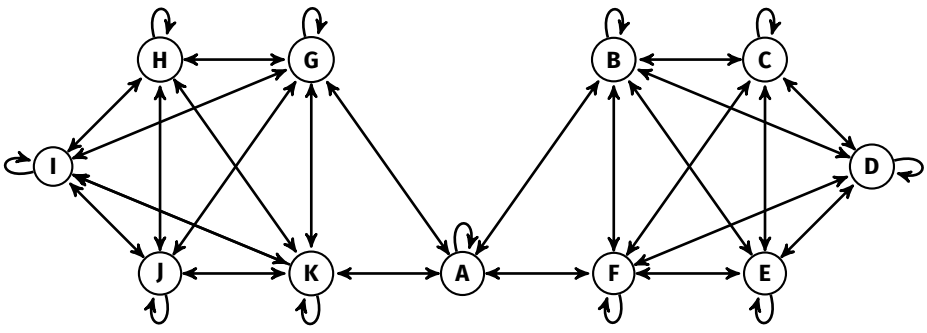


Figure 3.3.5. Criminal network.

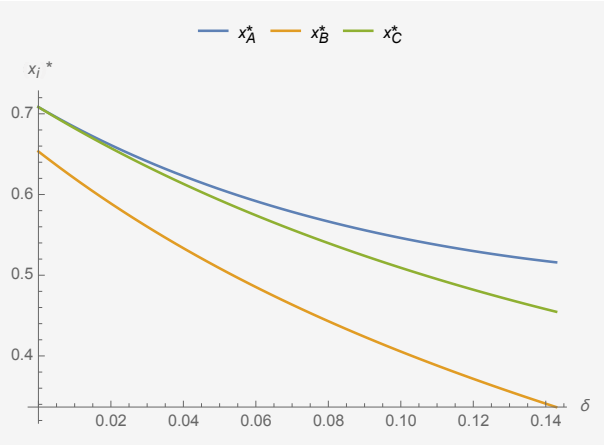


Figure 3.3.6. Information acquisition for the criminal network.

3.3.2 Equilibrium characterization

Let us divide the agents into two groups: *active* (A) agents, who are active learners ($x_i^* > 0$), and *passive* (P) agents. An equilibrium in which all agents belong to A is known as a *distributed equilibrium*, as effort is distributed among all agents. In contrast, a *specialized equilibrium* is such that only a few individuals (the specialists) learn, while the others free-ride.

This part mainly follows Bramoullé, Kranton, and D'Amours (2014). Without loss of generality, we can reorder the agents such that the first r are active and the last $n-r$ are passive. As $x_j = 0$ for all $j \in P$, for any individual i, we have $\sum_{j \neq i} g_{ij} x_j = \sum_{j \in A \setminus \{i\}} g_{ij} x_j$. Thus, for $i \in 1, ..., r$, an equilibrium requires that:

$$x_i^* = \bar{x}_i - \frac{\delta}{g_{ii}} \sum_{j \in A \setminus \{i\}} g_{ij} x_j^* > 0.$$

For $i \in \{r+1, ..., n\}$, an equilibrium requires that:

$$\bar{x}_i - \frac{\delta}{g_{ii}} \sum_{j \in A \setminus \{i\}} g_{ij} x_j^* \leq 0.$$

Let $\bar{\mathbf{x}}^A = (\bar{x}_1, ..., \bar{x}_r)$ and $\bar{\mathbf{x}}^P = (\bar{x}_{r+1}, ..., \bar{x}_n)$. The diagonal of a matrix A is denoted by d_A. Let $\mathbf{G}^A$ be the $r \times r$ minor corresponding to the active agents of the network, while $\mathbf{G}^P$ is the $(n-r-1) \times (n-r-1)$ minor of $\mathbf{G}$ corresponding to the passive agents. The $(n-r-1) \times r$ minor $\mathbf{G}^{P,A}$ of $\mathbf{G}$ is given by (g_{ij}) where $i \in P$ and $j \in A$. Rearranging the above expressions, we obtain the following result:

Proposition 3.3.2. *The profile of information levels* $\boldsymbol{x} = (x_1^*, ..., x_n^*) = (\boldsymbol{x}^A, 0)$ *with* $\boldsymbol{x}^A = (x_1^*, ..., x_r^*) \in (0,1]^r$ *constitutes an equilibrium if and only if*

$$\begin{cases} d_{\mathbf{G}^A} \bar{\boldsymbol{x}}^A &= \left[(1-\delta) d_{\mathbf{G}^A} + \delta \mathbf{G}^A\right] \boldsymbol{x}^A, \\ d_{\mathbf{G}^P} \bar{\boldsymbol{x}}^P &\leq \delta \mathbf{G}^{P,A} \boldsymbol{x}^A. \end{cases} \tag{3.3.2}$$

Note that given n agents, there are 2^n potential partitions. Obtaining all possible equilibria requires solving the system (3.3.2) for each partition. This can be done in two steps:

(i) First, solve for $\mathbf{x}^A$ in $d_{\mathbf{G}^A}\bar{\mathbf{x}}^A = \left[(1-\delta)d_{\mathbf{G}^A} + \delta\mathbf{G}^A\right]\mathbf{x}^A$. The solution is unique if and only if $\det[(1-\delta)d_{\mathbf{G}^A} + \delta\mathbf{G}^A] \neq 0$.

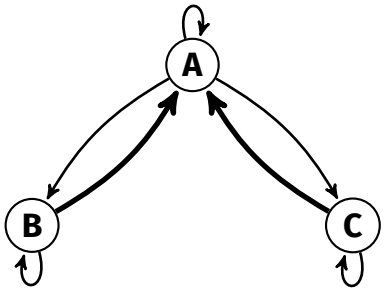
(ii) Then, check whether all components of x^A are strictly positive and $d_{\mathbf{G}^P}\bar{\mathbf{x}}^P \leq \delta\mathbf{G}^{P,A}\mathbf{x}^A$.

If the diagonal elements of $\mathbf{G}$ are identical, i.e., $g_{ii} = g_{jj}$ for all i,j, the number of equilibria is weakly lower than 2^n and can be computed in exponential time. In this case, the condition $\det[(1-\delta)d_{\mathbf{G}^A} + \delta\mathbf{G}^A] = 0$ simplifies to $\det\left[-\frac{(\delta-1)g_{ii}}{\delta}\mathbf{Id} + \mathbf{G}^A\right] = 0$, which holds if and only if $\mathbf{G}^A$ has an eigenvalue $\lambda = \frac{(\delta-1)g_{ii}}{\delta}$. Consequently, for almost all δ, the equation has a unique solution. While Bramoullé, Kranton, and D'Amours (2014) assume not only $d_{\mathbf{G}} = d_{\mathbf{Id}}$ but also matrix symmetry, we have shown that these assumptions are not necessary to obtain an explicit expression for equilibria.

3.3.3 Equilibrium uniqueness

In general, there might be multiple equilibria in this model. We present two examples.

The first example is a three-agent network that is incomplete and can also be visualized as a star. The weights of the connections between the agents are represented by the matrix $\mathbf{G}$. Figure 3.3.7 shows the graph corresponding to this network, with thicker arrows indicating larger weights.



$$\mathbf{G} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 2/3 & 1/3 & 0 \\ 2/3 & 0 & 1/3 \end{pmatrix}$$

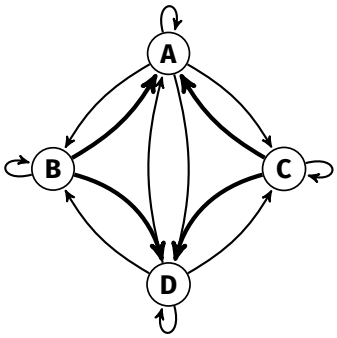
Figure 3.3.7. Incomplete three-agent network.

Since $g_{AA} = g_{BB} = g_{CC}$, it follows that $\bar{x}_A = \bar{x}_B = \bar{x}_C = \bar{x}$. Assuming $\delta = \frac{1}{2}$, the best reply functions become:

$$\begin{aligned} x_A &= \max\{0, \bar{x} - \frac{x_B + x_C}{2}\}, \\ x_B &= \max\{0, \bar{x} - x_A\}, \\ x_C &= \max\{0, \bar{x} - x_A\}. \end{aligned}$$

There are two distributed equilibria: $(x_A^*, x_B^*, x_C^*) = (\frac{\bar{x}}{2}, \frac{\bar{x}}{2}, \frac{\bar{x}}{2})$ and $(x_A^*, x_B^*, x_C^*) = (\frac{\bar{x}}{3}, \frac{2\bar{x}}{3}, \frac{2\bar{x}}{3})$. Additionally, there exist specialized equilibria where either agent A or both agents B and C purchase $\bar{x}$ while the others free-ride. Another similar example holds for $\delta = \frac{1}{k}$ and a star with k agents, demonstrating that the multiplicity of equilibria does not depend on the extreme assumption that $\delta = \frac{1}{2}$ (which is extreme in the sense that it implies $\tilde{\sigma}^2 = 0$).

The second example is a four-agents *eye*, shown in Figure 3.3.8. Weights are given by the matrix $\mathbf{G}'$.



$$\mathbf{G}' = \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 3/8 & 1/4 & 0 & 3/8 \\ 3/8 & 0 & 1/4 & 3/8 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$$

Figure 3.3.8. Four-agents *eye*.

Once again, we assume $\delta = \frac{1}{2}$. Since $\bar{x}_i = \bar{x}$ for all i, the best replies are as follows:

$$\begin{aligned} x_A &= \max\left\{0, \bar{x} - \frac{x_B + x_C + x_D}{2}\right\}, \\ x_B &= \max\left\{0, \bar{x} - \frac{3(x_A + x_D)}{4}\right\}, \\ x_C &= \max\left\{0, \bar{x} - \frac{3(x_A + x_D)}{4}\right\}, \\ x_D &= \max\left\{0, \bar{x} - \frac{x_A + x_B + x_C}{2}\right\}. \end{aligned}$$

There are two specialized equilibria: $\left(\frac{2}{3}\bar{x}, 0, 0, \frac{2}{3}\bar{x}\right)$ and $(0, \bar{x}, \bar{x}, 0)$. The rough idea behind multiplicity is that agents can be divided into two distinct groups so that active learning contributions vary between them. When one group learns more, the other decreases its effort, and vice versa. We will discuss this in detail in Subsection 3.3.3.2.

Next, we seek a structural condition on the network that guarantees uniqueness. It turns out that, given δ, the positive definiteness of a matrix that we denote $\mathbf{Q}$ ensures equilibrium uniqueness. This matrix $\mathbf{Q}$ can be determined from $\mathbf{G}$ in a one-to-one correspondence once δ is fixed.

Recall that agent i's expected payoffs are given by the following equation:

$$u_i(x_1,...,x_n) = \mathbb{E}\left[-\left(\mu - \sum_{j=1}^{n} g_{ij}((1-x_j)s_j + x_j\mathscr{g}_j)\right)^2\right] - x_i c.$$

Proposition 3.3.3. *The profile of active learning choices $\boldsymbol{x}^* = (x_1^*,...,x_n^*)$ is an equilibrium of the game if and only if*

$$(\boldsymbol{\theta} - \hat{\boldsymbol{Q}}\boldsymbol{x}^*)^T(\boldsymbol{x}^* - \boldsymbol{x}') \;\geq\; 0 \tag{3.3.3}$$

for any $\boldsymbol{x}' \in [0,1]^n$, with the matrix

$$\hat{\boldsymbol{Q}} = (\sigma^2 + \tilde{\sigma}^2)\begin{pmatrix} 2g_{11}^2 & 2\delta g_{11}g_{12} & ... & 2\delta g_{11}g_{1n} \\ 2\delta g_{22}g_{21} & 2g_{22}^2 & ... & 2\delta g_{22}g_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 2\delta g_{n1}g_{nn} & 2\delta g_{n2}g_{nn} & ... & 2g_{nn}^2 \end{pmatrix}$$

and the vector

$$\boldsymbol{\theta} = (2\sigma^2 g_{11}^2 - c, ..., 2\sigma^2 g_{nn}^2 - c).$$

Proof. First, the following equivalence is established: the profile $\boldsymbol{x}^*$ is an equilibrium if and only if

$$\frac{\partial}{\partial x_i}\left[u_i(x_i^*, x_{-i}^*)\right](x_i' - x_i^*) \leq 0$$

for all i and $x_i' \in [0,1]$.

Fixing a profile $\boldsymbol{x}^* \in [0,1]^n$ and an agent i, let us define

$$g(t) := u_i(x_i' + t(x_i^* - x_i'), x_{-i}^*)$$

for $0 \leq t \leq 1$ and $x_i' \in [0,1]$. The derivative with respect to t is given by $g'(t) = \frac{\partial}{\partial x_i}(u_i(x_i, x_{-i}^*))_{|x_i = x_i' + (x_i^* - x_i')}(x_i^* - x_i')$. If $\boldsymbol{x}^*$ is an equilibrium, $g(t)$ has a maximum at $t = 1$ and $g'(1) \geq 0$. Hence,

$$\frac{\partial}{\partial x_i}\left[u_i(x_i^*, x_{-i}^*)\right](x_i' - x_i^*) \leq 0.$$

Now, let us show the converse. Concavity of g follows from the concavity of u_i,[11] and then $g(t) \leq g(y) + g'(y)(t-y)$ for any $t, y \in [0,1]$. Choosing $t = 0$ and $y =$

11. The function u_i is clearly twice differentiable with respect to x_i and $\frac{\partial^2 u_i}{\partial x_i^2} = -2(-g_{ii}(\mathscr{g}_i - s_i))^2 < 0$.

1, we see that $g(0) \leq g(1) - g'(1)$. Moreover, $g'(1) \geq 0$ by assumption, so that $-g'(1) \leq 0$ and $g(0) \leq g(1)$. This inequality implies that

$$u_i(x_i^*, x_{-i}^*) \geq u_i(x_i', x_{-i}^*)$$

for all $x_i' \in [0,1]$, and $\boldsymbol{x}^*$ is an equilibrium.

Summing up with respect to all agent yields

$$\sum_{i=1}^{n} \left(\frac{\partial}{\partial x_i} \left[u_i(x_i^*, x_{-i}^*) \right] (x_i' - x_i^*) \right) \leq 0.$$

Denoting by $\left(\frac{\partial}{\partial x_i} u_i(x_i^*, x_{-i}^*) \right)_i$ the vector given by stacking up all $\frac{\partial u_i}{\partial x_i}$, the previous inequality can be rewritten as

$$\left(\frac{\partial}{\partial x_i} [u_i(x_i^*, x_{-i}^*)] \right)_i^T (\boldsymbol{x}' - \boldsymbol{x}^*) \leq 0.$$

The profile of active learning choices $\boldsymbol{x}^*$ is an equilibrium if and only if this inequality holds for any $\boldsymbol{x}' \in [0,1]^n$.[12] It just remains to explicitely derive the vector components, which are given by

$$\frac{\partial}{\partial x_i} [u_i(x_i^*, x_{-i}^*)] = 2g_{ii}^2 \sigma^2 - 2g_{ii}^2 x_i^* (\sigma^2 + \tilde{\sigma}^2) - 2\alpha g_{ii} \sum_{j \neq i} g_{ij} x_j^* - c.$$

Finally, it is a mere verification to check that defining $\hat{\boldsymbol{Q}}$ and $\boldsymbol{\theta}$ as above, $\boldsymbol{x}^*$ is an equilibrium if and only if

$$(\boldsymbol{\theta} - \hat{\boldsymbol{Q}} \mathbf{x}^*)^T (\boldsymbol{x}^* - \boldsymbol{x}') \geq 0.$$

□

If $\hat{\boldsymbol{Q}}$ is positive definite, there is just one vector of information levels $\mathbf{x}^*$ that satisfies (3.3.3). This is the sufficient condition for equilibrium uniqueness.

Proposition 3.3.4. *If the matrix $\hat{\boldsymbol{Q}}$ is positive definite, the equilibrium is unique.*

12. If $\hat{\mathbf{x}}$ is not an equilibrium, then there is some agent j such that

$$\frac{\partial}{\partial x_j} \left(u_j(\hat{x}_j, \hat{x}_{-j}) \right) (x_j' - \hat{x}_j) > 0$$

for some $x_j' \in [0,1]$. Hence, defining the profile $\tilde{\mathbf{x}}$ as $\tilde{x}_j = x_j'$ and $\tilde{x}_i = \hat{x}_i$ for $i \neq j$,

$$\sum_i \left(\frac{\partial}{\partial x_i} (u_i(\hat{x}_i, \hat{x}_{-i}) \right) (\tilde{x}_i - \hat{x}_i) = \frac{\partial}{\partial x_j} \left(u_j(\hat{x}_j, \hat{x}_{-j}) \right) (x_j' - \hat{x}_j) > 0.$$

Proof. Suppose $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$ are two different equilibria. Then, $(\boldsymbol{\theta} - \hat{\mathbf{Q}}\mathbf{x}_1^*)^T(\mathbf{x}_1^* - \mathbf{x}_2^*) \geq 0$ and $(\boldsymbol{\theta} - \hat{\mathbf{Q}}\mathbf{x}_2^*)^T(\mathbf{x}_2^* - \mathbf{x}_1^*) \geq 0$. Summing up both inequalities yields $(\boldsymbol{\theta} - \hat{\mathbf{Q}}\mathbf{x}_1^*)^T(\mathbf{x}_1^* - \mathbf{x}_2^*) + (\boldsymbol{\theta} - \hat{\mathbf{Q}}\mathbf{x}_2^*)^T(\mathbf{x}_2^* - \mathbf{x}_1^*) \geq 0$, which holds if and only if

$$(\mathbf{x}_2^* - \mathbf{x}_1^*)^T\hat{\mathbf{Q}}(\mathbf{x}_2^* - \mathbf{x}_1^*) \leq 0.$$

But $\hat{\mathbf{Q}}$ is positive definite, i.e. $\mathbf{x}^T\hat{\mathbf{Q}}\mathbf{x} > 0$ for all $x \neq 0$. Consequently, $\mathbf{x}_1^* = \mathbf{x}_2^*$ and the equilibrium is unique. □

Dividing $\hat{\mathbf{Q}}$ by $\sigma^2 + \tilde{\sigma}^2$ does not change its definiteness and simplifies the expression—recall that $\sigma^2 + \tilde{\sigma}^2 > 0$.[13] Thus, $\mathbf{Q}$ is given by

$$\mathbf{Q} = \frac{1}{\sigma^2 + \tilde{\sigma}^2}\hat{\mathbf{Q}} = \begin{pmatrix} 2g_{11}^2 & 2\delta g_{11}g_{12} & \dots & 2\delta g_{11}g_{1n} \\ 2\delta g_{22}g_{21} & 2g_{22}^2 & \dots & 2\delta g_{22}g_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 2\delta g_{n1}g_{nn} & 2\delta g_{n2}g_{nn} & \dots & 2g_{nn}^2 \end{pmatrix}. \tag{3.3.4}$$

The following result shows that $\mathbf{Q}$ is completely determined by $\mathbf{G}$, once δ is fixed. Consequently, equilibrium uniqueness for this model depends solely on the influence network $\mathbf{G}$.

Proposition 3.3.5. *Given δ, there is a one-to-one correspondence between $\mathbf{Q}$ and $\mathbf{G}$.*

Proof. Given δ, the matrix $\mathbf{Q}$ is defined element-wise from $\mathbf{G}$ as in (3.3.4). Assume δ is fixed and denote this transformation by ϕ_δ. Let us show that it is possible to recover $\mathbf{G}$ from $\mathbf{Q}$. Denoting by q_{ij} the elements in $\mathbf{Q}$, let us define (element-wise) the transformation τ_δ by $\tau_\delta(q_{ii}) = \sqrt{q_{ii}}$ for all i and $\tau_\delta(q_{ij}) = \frac{q_{ij}}{\delta\sqrt{q_{ii}}}$ for all $i \neq j$. It is trivial to check that $\tau_\delta(\phi_\delta(\mathbf{G})) = \mathbf{G}$ and $\phi_\delta(\tau_\delta(\mathbf{Q})) = \mathbf{Q}$. □

In general, the matrix $\mathbf{Q}$ is an $n \times n$ matrix that need not be symmetric. Note that $\mathbf{x}^T\mathbf{Q}\mathbf{x} = \frac{1}{2}\mathbf{x}^T(\mathbf{Q} + \mathbf{Q}^T)\mathbf{x}$, and $\mathbf{Q}$ is positive definite if and only if $\mathbf{A} := \frac{1}{2}(\mathbf{Q} + \mathbf{Q}^T)$ is positive definite. Since $\mathbf{A}$ is symmetric, we can use the characterization of positive definiteness in terms of eigenvalues: a symmetric matrix is positive definite if and only if all of its eigenvalues are positive. Let $\lambda_1(\mathbf{A})$ denote the lowest eigenvalue of $\mathbf{A}$.

Corollary 3.3.6. *If $\lambda_1(\mathbf{A}) > 0$, then the equilibrium is unique.*

13. It would be possible to divide by $2(\sigma^2 + \tilde{\sigma}^2)$ instead, but keeping the factor 2 simplifies the expression for the matrix $\mathbf{A}$ later.

The explicit expression for $\mathbf{A}$ is given by:

$$\mathbf{A} = \begin{pmatrix} 2g_{11}^2 & \delta(g_{12}g_{11} + g_{21}g_{22}) & \dots & \delta(g_{1n}g_{11} + g_{n1}g_{nn}) \\ \delta(g_{21}g_{22} + g_{12}g_{11}) & 2g_{22}^2 & \dots & \delta(g_{2n}g_{22} + g_{n2}g_{nn}) \\ \vdots & \vdots & \ddots & \vdots \\ \delta(g_{n1}g_{nn} + g_{1n}g_{11}) & \delta(g_{n2}g_{nn} + g_{2n}g_{22}) & \dots & 2g_{nn}^2 \end{pmatrix}.$$

Note that this sufficient condition is independent of the cost c of active learning but depends on the influences between agents and the substitutability of information acquisition.

The scope of this condition is the focus of our subsequent analysis. We will make more restrictive assumptions on the model to explore particular cases of interest, which will eventually lead to a result similar to that of Bramoullé, Kranton, and D'Amours (2014). Later, we will apply the sufficient condition to the examples in Section 3.3.1. We first prove an auxiliary lemma.

Lemma 3.3.7. *Let $\boldsymbol{A}$ be a symmetric matrix and $\beta, \delta > 0$. The matrix $\beta\boldsymbol{Id} + \delta\boldsymbol{A}$ is positive definite if and only if $\lambda_1(\boldsymbol{A}) \geq -\frac{\beta}{\delta}$.*

Proof. The matrix $\beta\,\mathbf{Id} + \delta\mathbf{A}$ is positive definite if and only if all the solutions λ to $\det[\lambda\mathbf{Id} - (\beta\,\mathbf{Id} + \delta\mathbf{A})] = 0$ are strictly positive. The equation is equivalent to $\det[\frac{\lambda-\beta}{\delta}\,\mathbf{Id} - \mathbf{A}] = 0$. Note that the eigenvalues of $\mathbf{A}$ are the solutions t to the equation $\det[t\,\mathbf{Id} - \mathbf{A}] = 0$. Consequently, as $t = \frac{\lambda-\beta}{\delta}$, the condition $\lambda > 0$ can be translated into all eigenvalues t of $\mathbf{A}$ verifying $t > -\frac{\beta}{\delta}$. This is precisely the condition $\lambda_1(\mathbf{A}) > -\frac{\beta}{\delta}$. □

Next, we consider two particular cases that are worth exploring. Assuming that all agents pay the same attention to themselves, i.e., $g_{ii} = g_{jj}$ for all i,j, we can denote the diagonal terms of $\mathbf{G}$ by $\beta := g_{ii} > 0$. We define $\bar{\mathbf{A}}$ as

$$\bar{\mathbf{A}} = \begin{pmatrix} 0 & \frac{g_{12}+g_{21}}{2} & \dots & \frac{g_{1n}+g_{n1}}{2} \\ \frac{g_{21}+g_{12}}{2} & \ddots & \dots & \frac{g_{2n}+g_{n2}}{2} \\ \vdots & \dots & \ddots & \vdots \\ \frac{g_{n1}+g_{1n}}{2} & \frac{g_{n2}+g_{2n}}{2} & \dots & 0 \end{pmatrix}.$$

Using Lemma 3.3.7, we see that $\lambda_1(\mathbf{A}) > 0$ if and only if $\lambda_1(\bar{\mathbf{A}}) > -\frac{\beta}{\delta}$. Note that $\bar{\mathbf{A}}$ is simply $\bar{\mathbf{A}} = \frac{1}{2}(\mathbf{G} + \mathbf{G}^T) - \beta\,\mathbf{Id}$. Here, $\bar{\mathbf{A}}$ reflects the average flow of information between a pair of networks, or the undirected network associated with $\mathbf{G}$.

Now, assume that the network displays reciprocal relations, i.e., $g_{ij} = g_{ji}$, in addition to same self-importance across agents. This means that the influence of agent i on agent j is the same as that of agent j on agent i, and so the matrix

$\mathbf{G}$ is symmetric and can be seen as undirected. Again, $\lambda_1(\mathbf{A}) > 0$ if and only if $\lambda_1(\bar{\mathbf{A}}) > -\frac{\beta}{\delta}$, but $\bar{\mathbf{A}}$ is now simply $\mathbf{G} - \beta\mathbf{Id}$. The matrix $\bar{\mathbf{A}} = \mathbf{G} - \beta\,\mathbf{Id}$ can be seen as a generalization of the matrix $\mathbf{G}$ in Bramoullé, Kranton, and D'Amours (2014), where $\bar{a}_{ij} \in [0,1]$ instead of $g_{ij} \in \{0,1\}$. The sufficient condition is equivalent to theirs. However, to derive such a result they use the potential theory developed by Monderer and Shapley (1996), which requires symmetry—this is why we cannot apply it to the general model.

Proposition 3.3.8. *The sufficient condition for the uniqueness of equilibrium can be specialized to two particular cases:*

- *If self-importance is equal across agents ($g_{ii} = g_{jj} = \beta$ for all i,j), the condition becomes $\lambda_1(\bar{\boldsymbol{A}}) > -\frac{\beta}{\delta}$ with $\bar{\boldsymbol{A}} = \frac{1}{2}(\boldsymbol{G} + \boldsymbol{G}^T) - \beta\boldsymbol{Id}$.*
- *If on top of that the influences are reciprocal ($g_{ij} = g_{ji}$ for all i,j), the condition becomes $\lambda_1(\bar{\boldsymbol{A}}) > -\frac{\beta}{\delta}$ with $\bar{\boldsymbol{A}} = \boldsymbol{G} - \beta\boldsymbol{Id}$.*

This proposition summarizes the results obtained so far, which show that the condition for the uniqueness of equilibrium can be specialized for two particular cases: when all agents have the same level of self-importance and when the network exhibits reciprocal relations between agents. In both cases, the condition involves the eigenvalue of a matrix $\bar{\mathbf{A}}$, which can be calculated based on the properties of the network. The precise definition of $\bar{\mathbf{A}}$ is given for each case.

3.3.3.1 Examples: Uniqueness

The networks analyzed in Section 3.3.1 are reviewed again to apply the equilibrium uniqueness condition.

First, we revisit the class of ***k*-regular graphs** with n agents that share their attention homogeneously. Proposition 3.3.8 applies, and the lowest eigenvalue of $\bar{\mathbf{A}}$ is $\lambda_1(\bar{\mathbf{A}}) = -\frac{1}{k}$. The equilibrium is unique if $\delta < 1$, which always holds. As an example of this class of networks, the matrix $\bar{\mathbf{A}}$ associated with the complete graph is given by

$$\bar{\mathbf{A}} = \begin{pmatrix} 0 & \frac{1}{n} & \dots & \frac{1}{n} \\ \frac{1}{n} & 0 & \dots & \vdots \\ \vdots & \dots & \ddots & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \dots & 0 \end{pmatrix}.$$

Next, we consider the class of **stars**. Due to the asymmetry of $\mathbf{Q}$ and the different terms in the diagonal (self-importance is not equal across agents), only Corollary 3.3.6 applies. The equilibrium is unique if $\lambda_1(\mathbf{A}) > 0$, which depends on both δ and ε. Matrix $\mathbf{A}$ is given here by:

$$\mathbf{A} = \begin{pmatrix} \frac{2}{n^2} & \delta((1-\varepsilon)\varepsilon + \frac{1}{n^2}) & \dots & \delta((1-\varepsilon)\varepsilon + \frac{1}{n^2}) \\ \delta((1-\varepsilon)\varepsilon + \frac{1}{n^2}) & 2(1-\varepsilon)^2 & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ \delta((1-\varepsilon)\varepsilon + \frac{1}{n^2}) & 0 & \dots & 2(1-\varepsilon)^2 \end{pmatrix}.$$

Figure 3.3.9 shows the values for which a unique equilibrium is ensured—every pair (δ, ε) such that the blue surface is above the orange plane.

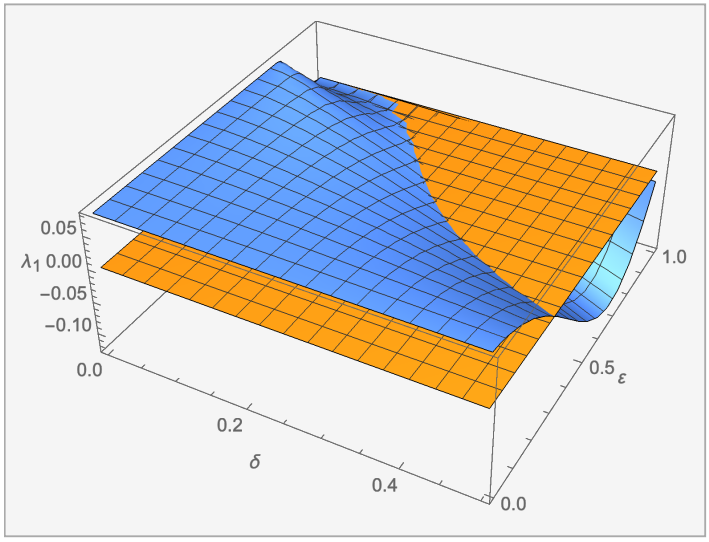


Figure 3.3.9. The lowest eigenvalue of the star.

A particular network structure belonging to the class of **core-periphery networks** was set in Figure 3.3.4a. Here,

$$\mathbf{A} = \begin{pmatrix} \frac{2}{9} & \delta\frac{2}{81} & \delta\frac{2}{81} & \delta\frac{9}{100} & 0 & 0 \\ \delta\frac{2}{81} & \frac{2}{9} & \delta\frac{2}{81} & 0 & \delta\frac{9}{100} & 0 \\ \delta\frac{2}{81} & \delta\frac{2}{81} & \frac{2}{9} & 0 & 0 & \delta\frac{9}{100} \\ \delta\frac{9}{100} & 0 & 0 & \frac{2}{100} & 0 & 0 \\ 0 & \delta\frac{9}{100} & 0 & 0 & \frac{2}{100} & 0 \\ 0 & 0 & \delta\frac{9}{100} & 0 & 0 & \frac{2}{100} \end{pmatrix}.$$

and we apply Corollary 3.3.6. It turns out that $\lambda_1(\mathbf{A}) > 0$ for all $\delta \in [0, \frac{1}{2}]$, so the equilibrium is always unique.[14]

14. The explicit expression for the lowest eigenvalue of $\mathbf{A}$ is $\lambda_1(\mathbf{A}) = \frac{981 - 100\delta - \sqrt{670761 - 163800\delta + 541441\delta^2}}{8100}$. We see that $\lambda_1(\mathbf{A})$ is a decreasing function of δ in $[0, 1/2]$. As it is strictly positive at $\delta = 1/2$, $\lambda_1(\mathbf{A}) > 0$ for all $\delta \in \left[0, \frac{1}{2}\right]$.

The **criminal network** from Ballester, Calvó-Armengol, and Zenou (2006) was represented in Figure 3.3.5. Proceeding as before, we calculate the lowest eigenvalue of $\mathbf{A}$.[15] We find that $\lambda_1(\mathbf{A}) > 0$ for all $\delta < 0.45011$, which guarantees a unique equilibrium for such values.

Finally, we consider the **incomplete network** depicted in Figure 3.3.7. Recall that $\delta = \frac{1}{2}$ and all diagonal terms are equal: $\beta = \frac{1}{3}$. To apply Proposition 3.3.8, we compute

$$\bar{\mathbf{A}} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{pmatrix}.$$

The uniqueness condition $\lambda_1(\bar{\mathbf{A}}) > -\frac{\beta}{\delta}$ is not satisfied because $\lambda_1(\bar{\mathbf{A}}) = -\frac{1}{\sqrt{2}} < -\frac{2}{3}$. This was expected, as we had already obtained two different equilibria for this particular network.

3.3.3.2 The lowest eigenvalue

The present subsection explores the meaning of the uniqueness condition and provides an intuition. A network is bipartite if agents can be divided into two sets, say R and S, such that if $i \in R$, i is not connected to any $j \in R$ except for herself. The network is completely bipartite if every $i \in R$ is connected to all $j \in S$. Bipartite networks represent disjoint or independent communities. An affiliation network is a classic example. Another bipartite network might be found when representing supervisor-candidate communication. A complete bipartite network represents one extreme of two-sidedness. The other extreme is the complete regular graph. In this subsection, we talk about two-sidedness as an intuitive measure of how close a network is to the complete bipartite graph.

First, let us briefly characterize $\lambda_1(\mathbf{A})$.[16] By definition, $\lambda_1(\mathbf{A}) = \min\{\lambda \in \mathbb{R} : \exists \epsilon \in \mathbb{R}^n \text{ satisfying } \lambda\epsilon = \mathbf{A}\epsilon\}$. Assuming $\epsilon \neq 0$, $\lambda\epsilon = \mathbf{A}\epsilon$ implies $\epsilon^T\lambda\epsilon = \epsilon^T\mathbf{A}\epsilon$, which leads to $\lambda\epsilon^T\epsilon = \epsilon^T\mathbf{A}\epsilon$, and finally to $\lambda||\epsilon||^2 = \epsilon^T\mathbf{A}\epsilon$. So, if $||\epsilon|| = 1$, then $\lambda = \epsilon^T\mathbf{A}\epsilon$. Hence, $\lambda_1(\mathbf{A}) = \min\left\{\lambda \in \mathbb{R} : \lambda = \epsilon^T\mathbf{A}\epsilon \text{ and } ||\epsilon|| = 1\right\}$.

Following Bramoullé, Kranton, and D'Amours (2014), we can use an eigenvector ϵ associated to $\lambda_1(\mathbf{A})$ to separate the agents into two groups. If $\epsilon_i \geq 0$, agent i

15. Let $y_1(\delta)$, $y_2(\delta)$ and $y_3(\delta)$ be the three roots of $-32400 - 97200\delta + 259200\delta^3 + (3096 + 6192\delta - 7200\delta^2)y + (-97 - 97\delta)y^2 + y^3$. Let $y_1(\delta)$ be the smallest root in $\delta \in \left[0, \frac{1}{2}\right]$. Then, $\lambda_1(\mathbf{A}) = \frac{1}{450}y_1(\delta)$ and $\lambda_1(\mathbf{A}) > 0 \Leftrightarrow \delta < 0.45011$.

16. Remember that when $\mathbf{Q}$ is symmetric, then $\mathbf{A} = \mathbf{Q}$.

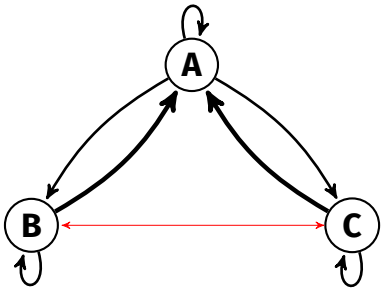
belongs to set R. Otherwise, she belongs to set S. This leads to the decomposition

$$\lambda_1(\mathbf{A}) = \varepsilon^T \mathbf{A} \varepsilon = \overbrace{\sum_{i,j \in R} \epsilon_i \epsilon_j q_{ij}}^{>0} + \overbrace{\sum_{i,j \in S} \epsilon_i \epsilon_j q_{ij}}^{>0} + 2 \overbrace{\sum_{i \in R, j \in S} \epsilon_i \epsilon_j q_{ij}}^{<0}.$$

The greater the lowest eigenvalue, the more weight the network puts within sets and the less it puts between sets. Hence, the size of $\lambda_1(\mathbf{A})$ is related to the two-sidedness of the graph $\mathbf{A}$. The closer the network is to the complete bipartite graph, the lower $\lambda_1(\mathbf{A})$. This is because transferring weight from links within R or S to links between both sets decreases $\lambda_1(\mathbf{A})$. Creating new links between sets or removing links within R or S belong to that kind of weight transfer. Thus, making the graph more two-sided decreases the lowest eigenvalue.

Let us show how the division of agents into the two groups is induced by agents' listening structures. We have $\lambda_1(\mathbf{A}) = \epsilon^t \mathbf{A} \epsilon = \sum_{i,j} q_{ij} \epsilon_i \epsilon_j$ with $||\epsilon|| = 1$. Without loss of generality, let us assume that $\lambda_1(\mathbf{A}) > 0$ (if not, a similar reasoning holds). Then, agent i belongs to R if and only if $\lambda_1(\mathbf{A})\epsilon_i > 0$. Since $\lambda_1(\mathbf{A})\epsilon_i = \mathbf{A}\epsilon_i$, we see that $i \in R$ if $\sum_{j=1}^n q_{ij}\epsilon_j \geq 0$. Consequently, if the listening structures of two agents are similar, they will belong to the same set. For example, if $(q_{ij})_j$ and $(q_{kj})_j$ are similar, then $\sum_{j=1}^n q_{ij}\epsilon_j > 0$, $\sum_{j=1}^n q_{kj}\epsilon_j > 0$, and both i and k belong to R. Hence, the division of agents into two groups induced by $\lambda_1(\mathbf{A})$ responds to their listening structures.

Recall that $\lambda_1(\mathbf{A}) > 0$ ensures uniqueness. The less two-sided the network is, the higher the chances of a unique equilibrium. Roughly, two-sided networks allow the agents from R and S to switch contributions in different equilibria. This occurs because the effects of substitutability (namely, the fact that if an agent contributes more, her neighbors contribute less and so on) accumulate and lead to several equilibrium configurations. When the network is not two-sided, this rebounding effect collapses, and there is only one equilibrium.



$$\mathbf{G}' = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 2/3 - \varepsilon & 1/3 & \varepsilon \\ 2/3 - \varepsilon & \varepsilon & 1/3 \end{pmatrix}$$

Figure 3.3.10. The extra link makes the network complete.

As an example, consider the incomplete network with three agents from Figure 3.3.7. Recall that for $\delta = \frac{1}{2}$ the network features multiple equilibria. The lowest

eigenvalue is $\lambda_1(\mathbf{Q}) = \frac{2-3\sqrt{2}\delta}{9}$, and an associated eigenvector is $(-\sqrt{2}, 1, 1)$. Thus, the partition is given by $R = \{\text{A}\}$ and $S = \{\text{B,C}\}$.[17] The network is considerably two-sided. What would happen if we add a link between agents B and C, slightly decreasing the two-sidedness of the network according to ε? The resulting network, shown in Figure 3.3.10, would be less similar to the bipartite network of three agents. In fact, it turns out that for all $\varepsilon > 0.057$, $\lambda_1(\mathbf{Q}') > 0$ and the equilibrium is unique, where $\mathbf{Q}'$ is the matrix induced by $\mathbf{G}'$. The network is less two-sided, and multiplicity disappears.

Furthermore, it is worth noting that $g_{ii} > 0$ for all i and self-importance (which represent the influence an agent exert on herself) contribute to the positivity of $\lambda_1(\mathbf{Q})$. For example, consider the case $g_{ii} = g_{jj} = \beta$ for all i,j. In this case, Proposition 3.3.8 simplifies the uniqueness condition to $\lambda_1(\bar{\mathbf{A}}) > -\frac{\beta}{\delta}$, where $\bar{\mathbf{A}} = \frac{1}{2}(\mathbf{G} + \mathbf{G}^T) - \beta\mathbf{Id}$. Then, $\lambda_1(\bar{\mathbf{A}}) = \epsilon^T\bar{\mathbf{A}}\epsilon = \sum_{i\neq j}\left(\frac{g_{ij}+g_{ji}}{2}\right)\epsilon_i\epsilon_j$. The equilibrium is unique if

$$\sum_{i\neq j\in R}(g_{ij}+g_{ji})\epsilon_i\epsilon_j + \sum_{i\neq j\in S}(g_{ij}+g_{ji})\epsilon_i\epsilon_j + \sum_{i\in S, j\in S}(g_{ij}+g_{ji})\epsilon_i\epsilon_j > -\frac{\beta}{\delta}.$$

As agents put more weight on their own signals (i.e., as β grows), the network becomes less bipartite, which contributes to potential equilibrium uniqueness.

3.4 Social Welfare

So far, agents have behaved individually. Now, the focus is shifted to a social perspective that maximizes aggregated welfare in the network. We can think of a utilitarian social planner who decides on the levels of active learning to pursue this goal.

The vector of learning levels $\mathbf{x}^{UO} = (x_1^{UO}, ..., x_n^{UO})$ that maximizes the sum of agents' utilities (the utilitarian optimum) is given by

$$x_i^{UO} = \max\left\{0, \frac{2\sigma^2 - c/(\sum_{j=1}^n g_{ji}^2)}{2(\sigma^2+\tilde{\sigma}^2)} - \delta\frac{\sum_{j=1}^n g_{ji}}{\sum_{j=1}^n g_{ji}^2}\sum_{j\neq i} g_{ij}x_j\right\}. \tag{3.4.1}$$

Agent i's learning target in the utilitarian optimum is

$$\tilde{x}_i = \frac{2\sigma^2 - c/(\sum_{j=1}^n g_{ji}^2)}{2(\sigma^2+\tilde{\sigma}^2)}.$$

17. Note that $(\sqrt{2}, -1, -1)$ is also an eigenvector associated to the eigenvalue $\frac{2-3\sqrt{2}\delta}{9}$. The partition induced by it is the equivalent to the one above: $R = \{B, C\}$ and $S = \{A\}$.

Now, we compare the target $\tilde{x}_i$ to the target in equilibrium, $\bar{x}_i$. The sum of the squares of i's influences is greater than the square of her self-influence: $\sum_{j=1}^{n} g_{ji}^2 \geq g_{ii}^2$. Then, $\frac{c}{\sum_{j=1}^{n} g_{ji}^2} \leq \frac{c}{g_{ii}^2}$, and directly from the definitions of targets, we get

$$\tilde{x}_i \geq \bar{x}_i.$$

Thus, in the utilitarian optimum, each agent would like to learn strictly more, except in the trivial case where she is isolated. This effect is due to $\sum_{j=1}^{n} g_{ji}^2$, which substitutes g_{ii}^2 in the target expression. Before, each agent just cared about self-benefit: the more she listened to her signal, the more information she needed. Now, the goal is shifted, and individuals must care about the influence they have on others. The term $\sum_{j=1}^{n} g_{ji}^2$ is a measure of i's total impact on the network. The larger the influence, the higher the target $\tilde{x}_i$.

However, the utilitarian level of active learning x_i^{UO} need not be higher than the equilibrium choice x_i^*. The last term in (3.4.1) indicates the amount of information agent i does not need to purchase because of the substitutability effect. Substitutability is driven here by $\delta \frac{\sum_{j=1}^{n} g_{ji}}{\sum_{j=1}^{n} g_{ji}^2}$, whereas it was driven by the factor $\delta \frac{1}{g_{ii}}$ in equilibrium. Hence, it might be the case that an agent who engages in high levels of active learning in equilibrium is not influential at all (i.e., $\sum_{j=1}^{n} g_{ji}$ is small), and the planner asks her to decrease her effort. Even though every agent desires to become more informed (the target is higher), utilitarian maximization implies a more efficient share of effort in global terms. Thus, in general, there is no ranking regarding acquisition decisions. Formally,

$$x_i^{UO} \geq x_i^* \Leftrightarrow \frac{c}{2(\sigma^2 + \tilde{\sigma}^2)} \left(\frac{1}{g_{ii}^2} - \frac{1}{\sum_{j=1}^{n} g_{ji}^2} \right) \geq \delta \left(\frac{\sum_{j=1}^{n} g_{ji}}{\sum_{j=1}^{n} g_{ji}^2} \sum_{j=1}^{n} g_{ji} x_j^{UO} - \frac{\sum_{j=1}^{n} g_{ij} x_j^*}{g_{ii}} \right).$$

On the one hand, we observe that the inequality would hold for networks in which attention is homogenously shared. On the other hand, it would also hold for networks with low levels of substitutability. This condition can be formalized.

Proposition 3.4.1. *For every network structure* $\boldsymbol{G}$ *there exists some* $\bar{\delta} \in (0,1)$ *such that if* $\delta \leq \bar{\delta}$, *then* $x_i^{UO} \geq x_i^*$ *for every agent* i.

The intuition behind this result is simple: for low levels of substitutability, every agent relies on her information target, which is always higher under the utilitarian planner. To illustrate the relation between network balance and the ranking in acquisition choices we provide an example. Suppose we have the network shown in Figure 3.3.7, and let the parameter values be $\delta = 0.2$, $c = 0.1$, $\sigma^2 = 3$, and $\tilde{\sigma}^2 = 1$. The network matrix and the equilibrium and utilitarian optimal choices are

$$\mathbf{G} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}, \qquad \begin{aligned} \mathbf{x}^* &= (0.23, 0.23, 0.23), \\ \mathbf{x}^{UO} &= (0.50, 0.51, 0.51). \end{aligned}$$

Here, the network is balanced, meaning weights are shared similarly among agents, and the utilitarian optimal choices are larger. However, if agent 1 becomes stubborn (i.e., g_{11} is close to 1), the network becomes unbalanced:

$$\mathbf{G}' = \begin{pmatrix} 8/10 & 1/10 & 1/10 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}, \qquad \begin{aligned} (\mathbf{x}^*)' &= (0.69, 0, 0), \\ (\mathbf{x}^{UO})' &= (0.57, 0.39, 0.39). \end{aligned}$$

In this case, agent 1 is the only one exerting effort in equilibrium, while the others free-ride. From a social point of view, this is not efficient, and agent 1 has to decrease her contribution while agents 2 and 3 increase theirs significantly.

Finally, we show by example that, in general, there is no Pareto dominance between the utilitarian optimum and equilibria—not even for low values of δ. Consider the above networks $\mathbf{G}$ and $\mathbf{G}'$ for the same parameter values again, focusing on agent 1. In $\mathbf{G}$, agent 1's utilities in the utilitarian optimum and the unique equilibrium are

$$\begin{aligned} U_1(\mathbf{x}^*) &= -2.322, \\ U_1(\mathbf{x}^{UO}) &= 0.017, \end{aligned}$$

while in $\mathbf{G}'$, agent 1 utilities are given by

$$\begin{aligned} U_1((\mathbf{x}^*)') &= 0.025, \\ U_1((\mathbf{x}^{UO})') &= 0.008. \end{aligned}$$

In such an unbalanced network, agent 1 strictly prefers the equilibrium allocation: as her self-importance is large, the increase in agent 2 and 3 information purchases does not make up for the decrease in hers under the utilitarian optimum. Thus, $\mathbf{x}^{UO}$ maximizes the sum of utilities, but in general it does not improve the well-being of every agent.

3.5 Extension to multiple periods

So far, we have considered a scenario where agents communicate only once. However, if we introduce multiple communication periods, agents can obtain information not only from their immediate neighbors but also from neighbors' neighbors. As time progresses, the DeGrootian posterior signal incorporates signals from individuals located at increasing distances. After t periods, each agent holds a belief containing signals from all individuals who live within t degrees of separation. One significant advantage of DeGroot updating is that the weights of period t are

simply given by the stochastic matrix $\mathbf{G}^t$. This implies that the information acquisition problem for t periods is identical to the one considered so far, except that the matrix $\mathbf{G}$ is now replaced by $\mathbf{G}^t$.

The limiting case $t \to \infty$ corresponds to long-run communication. There, each agent's posterior signal aggregates information from everyone in the network. Under very mild conditions there is *convergence*, meaning that different agents' posterior signals coincide. The $n \times n$ stochastic matrix $\mathbf{G}$ is said to be convergent if $\lim_{t\to\infty} \mathbf{G}^t\mathbf{v}$ exists for all $\mathbf{v} \in \mathbb{R}^n$. In this case, there exists a unique left eigenvector $\boldsymbol{\pi} = (\pi_1, ..., \pi_n)$ of $\mathbf{G}$ whose entries sum to 1 such that $(\lim_{t\to\infty} \mathbf{G}^t\mathbf{v})_i = \boldsymbol{\pi}^t\mathbf{v}$ for every i and all $\mathbf{v} \in \mathbb{R}^n$,[18] that is:

$$\lim_{t\to\infty} \mathbf{G}^t = \begin{pmatrix} \boldsymbol{\pi}^t \\ \vdots \\ \boldsymbol{\pi}^t \end{pmatrix}.$$

The components of $\boldsymbol{\pi}$ indicate how much each agent is listened to in the long-run. Again, this is equivalent to a public goods game, as there are n agents privately deciding how much to collaborate towards a common payoff. Thus, an influential individual, i.e., an individual with a large π_i, will purchase a significant amount of information, while another whose influence vanishes will just free-ride.

Requiring one agent to put positive weight on her belief (i.e., at least one $g_{ii} > 0$) is enough to ensure convergence for a stochastic matrix.[19] Hence, every network matrix analyzed in this chapter is convergent. Equilibrium efforts for $t \to \infty$ are given by

$$x_i^* = \max\left\{0, \frac{2\sigma^2 - c/\pi_i^2}{2(\sigma^2 + \tilde{\sigma}^2)} - \frac{\delta}{\pi_i}\sum_{j\neq i} \pi_j x_j\right\}.$$

All results shown so far hold for the long run with the corresponding matrix $\lim_{t\to\infty} \mathbf{G}^t$.

The utilitarian optimum is given by

$$x_i^{UO} = \max\left\{0, \frac{2\sigma^2 - c/(n\pi_i^2)}{2(\sigma^2 + \tilde{\sigma}^2)} - \frac{\delta}{\pi_i}\sum_{j\neq i} \pi_j x_j\right\}.$$

18. This result is taken from Golub and Jackson (2010).

19. If there is one agent i such that $g_{ii} > 0$, then the matrix is aperiodic. For strongly connected matrices, aperiodicity is necessary and sufficient for convergence; see Golub and Jackson (2010).

It is worth noting that $x_i^* \leq x_i^{UO}$ for all agents i and every network, in stark contrast to the one-shot game. In the limit, neighborhoods disappear and each agent i influences every other agent, including herself, in the same manner: π_i. Hence, the substitutability of information is identical under both the utilitarian optimum and the equilibrium allocation. However, as the information target is always higher under the utilitarian optimum, the levels of information acquisition are also higher. However, the utilitarian optimum is not always a Pareto improvement. In a setting with very low (almost negligible) substitutability levels, for example, agents would prefer the equilibrium allocation to the utilitarian optimum.

3.5.1 Examples: Infinitely many communication periods

The networks analyzed in Section 3.3.1 are reviewed again assuming that agents communicate for infinitely many periods before acquisition decisions are made. Here, the network matrix $\mathbf{G}$ is substituted with the matrix $\lim_{t\to\infty} \mathbf{G}^t$, which is well-defined since every network matrix analyzed in this chapter converges.

First, let us revisit the case of k-**regular graphs** with n agents that share their attention homogeneously. Given a fixed n, the specific limiting matrix $\lim_{t\to\infty} \mathbf{G}^t$ depends not only on $k \leq n$ but also on the network configuration. Agents not belonging to a cycle in the graph (excluding loops) will not be listened to in the long run, resulting in $\pi_j = 0$ for such agents j. The remaining agents (whose number we denote by $\tilde{k}$) share attention homogeneously. It holds that $\tilde{k} \geq k$. Then,

$$\pi_j = \begin{cases} 0 & \text{if there is no cycle to which } j \text{ belongs,} \\ \frac{1}{\tilde{k}} & \text{otherwise.} \end{cases}$$

And hence,

$$x_j^* = \begin{cases} 0 & \text{if there is no cycle to which } j \text{ belongs,} \\ \frac{2\sigma^2 - \tilde{k}^2 c}{2(\sigma^2+\tilde{\sigma}^2)(-\delta+\delta\tilde{k}+1)} & \text{if } j \text{ belongs to a cycle and } 2\sigma^2 - \tilde{k}c \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Comparing this with the one-shot communication version, we observe that fewer agents acquire information in the long run. Additionally, as $\tilde{k} \geq k$, the information acquisition levels decrease. This happens because, over time, all agents become connected to those who acquire information, and thus there is no need to acquire as much as before.

Now, we move to the class of **stars**. There, the hub pays homogeneous attention to the spokes, who, in turn, pay her attention ε. In the limit,

$$\boldsymbol{\pi} = \left(\frac{n\varepsilon}{(n-1)+n\varepsilon}, \frac{1}{(n-1)+n\varepsilon}, \ldots, \frac{1}{(n-1)+n\varepsilon} \right).$$

As the hub pays attention to the spokes, the star maintains the importance of all its members in the long run. Therefore, everyone is listened to, and, in principle, everyone acquires information—although the extent of information acquisition will also depend on the specific parameters involved. The hub is still the most influencial if ε is not too small. The acquisition levels in equilibrium are given by:

$$x_H^* = \frac{2\sigma^2 + \frac{1}{\varepsilon^2 n^2}(2\delta\varepsilon n(1+\varepsilon(n-2))n)\sigma^2 - c(\varepsilon n + n - 1)^2(1-\delta(2+(-1+\varepsilon(n-1))n))}{2(\sigma^2+\tilde{\sigma}^2)(\varepsilon-1)((n-1)\varepsilon\delta^2+\varepsilon-1)}$$

$$\text{if } c \leq \frac{(2\delta\varepsilon n(1+\varepsilon(n-2))n)\sigma^2}{(\varepsilon n + n - 1)^2(1-\delta(2+(-1+\varepsilon(n-1))n))}; \quad x_H^* = 0 \text{ otherwise,}$$

$$x_S^* = \frac{2\sigma^2\varepsilon n(\delta\varepsilon n - 1) - c(\delta - \varepsilon n)(\varepsilon n + n - 1)^2}{2(\sigma^2+\tilde{\sigma}^2)(\varepsilon-1)((n-1)\varepsilon\delta^2+\varepsilon-1)\varepsilon n} \text{ if } c \leq \frac{2\sigma^2\varepsilon n(\delta\varepsilon n - 1)}{(\delta-\varepsilon n)(\varepsilon n + n - 1)^2};$$

$$x_S^* = 0 \text{ otherwise.}$$

Now, revisiting the **core-periphery** network, it is important to note that agents in the core do not listen to peripheral agents, rendering the latter with no weight in the limit vector $\boldsymbol{\pi}$. If the core is composed by k agents, $\pi_j = \frac{1}{k}$ if j belongs to the core and $\pi_j = 0$ if j is peripheral. For the specific configuration from Section 3.3.1, acquisition levels are:

$$x_C^* = \frac{2\sigma^2 - 9c}{2(\sigma^2+\tilde{\sigma}^2)(2\delta+1)} \text{ if } c \leq \frac{2\sigma^2}{9}; \quad x_C^* = 0 \text{ otherwise,}$$

$$x_P^* = 0.$$

Core agents acquire the same information, but peripheral agents do not. Long-run communication does not affect core agents because they were already sharing information homogeneously (the restriction of $\mathbf{G}$ to the core is invariant under exponentiation to the power of t). Peripheral agents, in contrast, take into account their information in the one-shot game, but, with multiple communication rounds, their weight vanishes. Thus, they find it unprofitable to privately acquire information.

Finally, let us reexamine the **criminal network** from Ballester, Calvó-Armengol, and Zenou (2006). The limiting vector $\boldsymbol{\pi}$ is given by

$$\boldsymbol{\pi} = \left(\frac{5}{59}, \frac{6}{59}, \frac{5}{59}, \frac{5}{59}, \frac{5}{59}, \frac{6}{59}, \frac{6}{59}, \frac{5}{59}, \frac{5}{59}, \frac{5}{59}, \frac{6}{59}\right).$$

In the long run, only in-degree matters, but not network position. Hence, agent A no longer has a distinct role and there are just two classes of agents: B-class and C-class (to which agent A belongs now). B-class agents have in-degree 6, and

$\pi_j = \frac{6}{59}$, and C-class agents have in-degree five, so $\pi_j = \frac{5}{59}$. Best reply functions are given by:

$$x_B^* = \max\left\{0, \bar{x}_B - \delta\frac{59}{6}(3x_B + 7x_C)\right\},$$
$$x_C^* = \max\left\{0, \bar{x}_C - \delta\frac{59}{5}(4x_B + 6x_C)\right\}.$$

Similar to the one-period communication game, B-class agents acquire less information. In particular, for the configuration of parameters used in Section 3.3.1, we can observe in Figure 3.5.1 that B-class agents completely free-ride on C-class agents. This happens because B-class agents consider acquiring private information too costly, relying instead on the information obtained from the seven C-class agents.

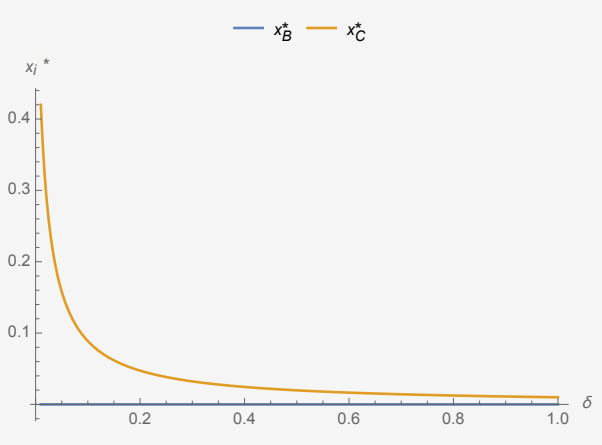


Figure 3.5.1. Acquisition for the criminal network in the long-run.

Finally, the examples from Section 3.3.3 are trivial in the long run. The incomplete three-agent network converges to a matrix characterized by the limit vector $\pi = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$, while the four-agents *eye* converges to a matrix characterized by the limit vector $\pi = \left(\frac{3}{10}, \frac{1}{5}, \frac{1}{5}, \frac{3}{10}\right)$. Both cases lead to unique equilibrium configurations.

3.6 Conclusion

This chapter has analyzed the behavior of DeGroot updaters in a networked environment and studied the impact of substitutability and network structure on information acquisition and welfare. We have shown that the substitutability of agents' active learning efforts induces free-riding behavior and can lead to multiple equilibria. We have also provided a sufficient condition for equilibrium uniqueness in terms of the lowest eigenvalue of the matrix $\mathbf{A}$, which is determined by $\mathbf{G}$ and the

parameter of substitutability δ. When this eigenvalue is positive, the equilibrium is unique. Even if there are multiple equilibria, we have proposed a procedure for calculating them.

In terms of welfare, we have found that the information target is lower in equilibria than under the utilitarian paradigm. This is significant since the target is precisely the level of information an agent will have at the end of the game. We have shown that it is socially desirable to increase the information level of every agent. While increasing agents' active learning may seem like a solution, we show that in the one-shot game it is not. Not only the ranking in targets does not imply a ranking in acquisition levels, but the utilitarian optimum does not Pareto dominate the equilibrium allocation. Nevertheless, over the long run, neighborhood frictions are eliminated and the utilitarian allocation always exceeds the equilibrium allocation.

An interesting avenue for further research would be the implementation problem of a planner trying to incentivize DeGroot updaters to move from equilibrium levels of active learning to the utilitarian optimum. Public information policies, such as subsidizing external information sources, rewarding learning contributions, or creating new links to foster communication, could also be explored.

Appendix 3.A Quasi-Bayesian Foundation

Regarding agents' cognitive sophistication, this paper follows the boundedly rational approach, which assumes that agents have limited cognitive resources and do not possess precise knowledge of their environment. Nonetheless, it is useful to connect the assumptions of this paper to the standard Bayesian framework. In this appendix, we provide a pure theoretical motivation for DeGroot updating in networks, following DeMarzo, Vayanos, and Zwiebel (2003). DeGroot updating can be viewed as a Bayesian updating process for agents that receive normally distributed signals but do not know the true variances of their neighbors' signals.

Consider n agents who want to estimate some unknown parameter $\mu \in \mathbb{R}$. Agent i receives an independent signal $x_i^0 \sim \mathcal{N}(\mu, \sigma_i^2)$, and she assigns some precision $\pi_{ij} = \frac{1}{\mathrm{Var}_i(x_j^0)}$ to agent j's signal, which may or may not be the true precision. Note that this assumption does not align with the standard Bayesian approach, which assumes that agents have precise knowledge of the signal structure. Agents communicate according to a social network $\tilde{\mathbf{G}}$, which is a directed graph that indicates whether agent i listens to agent j; $\tilde{g}_{ij} = 1$ if agent i listens to agent j, and $\tilde{g}_{ij} = 0$ otherwise. Each agent knows her own information, so $\tilde{g}_{ii} = 1$. Truthful reporting is assumed. Given normality and the assigned precisions, a sufficient statistic for the signals is their weighted average, with weights given by the precisions. DeMarzo, Vayanos, and Zwiebel (2003) denote such a statistic by x_i^1, and refer to it as agent i's *belief* after communication:

$$x_i^1 = \sum_{j=1}^{n} \frac{\tilde{g}_{ij}\pi_{ij}}{\sum_{j=1}^{n} \tilde{g}_{ij}\pi_{ij}} x_j^0.$$

The sufficiency of the statistic x_i^1 comes from the application of the Fisher-Neyman factorization theorem. Defining $g_{ij} := \sum_{j=1}^{n} \frac{\tilde{g}_{ij}\pi_{ij}}{\sum_{j=1}^{n} \tilde{g}_{ij}\pi_{ij}}$, we obtain the stochastic matrix $\mathbf{G} = (g_{ij})$. A DeGrootian population communicating according to $\mathbf{G}$ holds the same beliefs as the quasi-Bayesian population from DeMarzo, Vayanos, and Zwiebel (2003).[20] This insight provides additional motivation for the model described in this paper.

20. We say quasi-Bayesian because the critical assumption of potentially misperceived variances is not standard Bayesian. In a fully Bayesian world, agents would know the true precisions, and hence $\pi_{ij} = \pi_{kj} = \frac{1}{\mathrm{Var}(x_j^0)}$ for all i, k. This would imply $g_{ij} = g_{kj}$ in the equivalent DeGrootian network.

References

Acemoglu, Daron, and Asuman Ozdaglar. 2011. "Opinion dynamics and learning in social networks." *Dynamic Games and Applications* 1(1): 3–49. [81]

Acemoglu, Daron, Asuman Ozdaglar, and Ali ParandehGheibi. 2010. "Spread of (mis) information in social networks." *Games and Economic Behavior* 70(2): 194–227. [82]

Aumann, Robert J. 1997. "Rationality and bounded rationality." *Games and Economic behavior* 21(1-2): 2–14. [80]

Ballester, Coralio, Antoni Calvó-Armengol, and Yves Zenou. 2006. "Who's who in networks. Wanted: The key player." *Econometrica* 74(5): 1403–17. [83, 92, 103, 110]

Banerjee, Abhijit, Emily Breza, Arun G. Chandrasekhar, and Markus Mobius. 2021. "Naïve Learning with Uninformed Agents." *American Economic Review* 111(11): 3540–74. [82, 83]

Borgatti, Stephen P, and Martin G Everett. 2000. "Models of core/periphery structures." *Social Networks* 21(4): 375–95. [91]

Bramoullé, Yann, Rachel Kranton, et al. 2007. "Public goods in networks." *Journal of Economic Theory* 135(1): 478–94. [83, 84]

Bramoullé, Yann, Rachel Kranton, and Martin D'Amours. 2014. "Strategic interaction and networks." *American Economic Review* 104(3,b): 898–930. [83, 94, 95, 100, 101, 103]

Chandrasekhar, Arun G, Horacio Larreguy, and Juan Pablo Xandri. 2020. "Testing models of social learning on networks: Evidence from two experiments." *Econometrica* 88(1): 1–32. [82]

Corazzini, Luca, Filippo Pavesi, Beatrice Petrovich, and Luca Stanca. 2012. "Influential listeners: An experiment on persuasion bias in social networks." *European Economic Review* 56(6): 1276–88. [82]

Dasaratha, Krishna, and Kevin He. 2020. "Network structure and naive sequential learning." *Theoretical Economics* 15(2): 415–44. [82]

DeGroot, Morris H. 1974. "Reaching a Consensus." *Journal of the American Statistical Association* 69(345): 118–21. [80, 84]

DeMarzo, Peter M, Dimitri Vayanos, and Jeffrey Zwiebel. 2003. "Persuasion bias, social influence, and unidimensional opinions." *Quarterly Journal of Economics* 118(3): 909–68. [81, 85, 113]

Denti, Tommaso. 2017. "Network effects in information acquisition." working paper, available at *https://drive.google.com/file/d/1iAEILmFa5F_ihdaZSVQMnIWX235b3Nfs/view?usp=sharing*, [83]

Duraj, Jetlir, and Yi-Hsuan Lin. 2022. "Costly information and random choice." *Economic Theory* 74(1): 135–59. [80]

Fagiolo, Giorgio, Javier Reyes, and Stefano Schiavo. 2010. "The evolution of the world trade web: a weighted-network analysis." *Journal of Evolutionary Economics* 20(4): 479–514. [91]

Fricke, Daniel, and Thomas Lux. 2015. "Core–periphery structure in the overnight money market: evidence from the e-mid trading platform." *Computational Economics* 45 (3): 359–95. [91]

Galeotti, Andrea, and Sanjeev Goyal. 2010. "The law of the few." *American Economic Review* 100 (4): 1468–92. [82–84, 91]

Golub, Benjamin, and Matthew O. Jackson. 2010. "Naive learning in social networks and the wisdom of crowds." *American Economic Journal: Microeconomics* 2 (1): 112–49. [82, 108]

Golub, Benjamin, and Matthew Oscar Jackson. 2012. "How homophily affects the speed of learning and best-response dynamics." *Quarterly Journal of Economics* 127 (3): 1287–338. [82]

Grimm, Veronika, and Friederike Mengel. 2020. "Experiments on belief formation in networks." *Journal of the European Economic Association* 18 (1): 49–82. [82]

Jiménez-Martínez, Antonio. 2015. "A model of belief influence in large social networks." *Economic theory* 59: 21–59. [90]

Mailath, George J. 1998. "Do people play Nash equilibrium? Lessons from evolutionary game theory." *Journal of Economic Literature* 36 (3): 1347–74. [80, 85]

Melo, Emerson. 2022. "On the uniqueness of quantal response equilibria and its application to network games." *Economic Theory* 74 (3): 681–725. [83]

Molavi, Pooya, Alireza Tahbaz-Salehi, and Ali Jadbabaie. 2018. "A theory of non-Bayesian social learning." *Econometrica* 86 (2): 445–90. [82]

Monderer, Dov, and Lloyd S Shapley. 1996. "Potential games." *Games and Economic Behavior* 14 (1): 124–43. [83, 101]

Mueller-Frank, Manuel, and Claudia Neri. 2013. "Social learning in networks: Theory and experiments." working paper, available at SSRN: *https://ssrn.com/abstract=2328281*, [82]

Mueller-Frank, Manuel, and Claudia Neri. 2021. "A general analysis of boundedly rational learning in social networks." *Theoretical Economics* 16 (1): 317–57. [82]

Myatt, David P, and Chris Wallace. 2019. "Information acquisition and use by networked players." *Journal of Economic Theory* 182: 360–401. [83]

Samuelson, Larry. 2002. "Evolution and game theory." *Journal of Economic Perspectives* 16 (2): 47–66. [80]