Information Sequencing with Naive Social Learning*

Pëllumb Reshidi †

November 2020

[Click here for the latest version]

Abstract

Does the order and timing of information arrival affect beliefs formed within a group? We address this question by extending the DeGroot social learning model to allow for sequential information arrival. We find that the final beliefs can be altered by varying only the sequencing of information arrival, keeping the information content unchanged. We identify the optimal and pessimal information release sequences that yield the highest and lowest attainable consensus, respectively. In doing so, we bound the variation in final beliefs that can be attributed to the variation in the sequencing of information. We show that groups in which all members are equally influential are those most susceptible to information sequencing. Finally, with regard to information aggregation, as the number of group members grows, the sequential arrival of information compromises the group's beliefs: in all but particular cases, beliefs converge away from the truth.

*I am indebted to Leeat Yariv, Roland Bénabou and Alessandro Lizzeri for their continued guidance. I have greatly benefited from discussions with Ben Golub, Pietro Ortoleva, Can Urgun, Xiaosheng Mu, Joao Thereze, Mu Zhang, and Evgenii Safonov. I am also grateful to audiences at Princeton University and "The 30th International Conference on Game Theory - Stony Brook Center for Game Theory" for their helpful comments and discussions.

†Department of Economics, Princeton University preshidi@princeton.edu

1 Introduction

1.1 Overview

In a wide range of settings, social learning is a vital channel of information transmission. Examples abound: the members of hiring committees rely on colleagues' impressions of candidates, individuals consider others' choices when deciding which car to purchase, to which schools to send their children, or which political candidates to support. In learning environments like these, information often arrives over time. Consider a committee evaluating a candidate, where, in addition to common information—the candidate's CV, letters of recommendation, etc.—each member receives additional private information—e-mails from advisors, conference interactions, and so on. Some of the information might work in the candidate's favor while some might not. Does it matter if information arrives all at once or in sequence? Does it matter if good news arrives first, followed by bad, or vice versa? Do features of the committee, such as the social influence of various members, interact with the timing of information? These questions are at the heart of this project.

Under Bayesian learning, the order by which agents receive information has no impact in and of itself. Nonetheless, Bayesian updating is arguably demanding in such settings in terms of both calculation complexity and the information required about the underlying network of connections. Consequently, going back to DeGroot (1974), a large literature has focused on naive social learning, where agents use simple heuristics to update their beliefs. In this paper, we study the impact of information timing when agents learn naively. We show that the sequencing of information can have a large impact. This impact is particularly pronounced in "equal" societies, where each agent has the same social influence. Furthermore, information order effects can be important even in the limit: even when the information available to society leaves little aggregate uncertainty, for many information arrival sequences, naive learning leads to biased consensus beliefs.

In settings in which agents act sequentially but take actions only once, strategic concerns are limited. Therefore, it is typically assumed that agents update their beliefs via Bayes' rule. On the other hand, when agents act repeatedly, despite its normative appeal, the requirement of Bayesian updating is arguably strong. A growing body of literature, both theoretical and experimental, casts doubt on the ability of individuals to meet the demands of Bayesian updating, especially in complex environments, where doing so is cognitively demanding.[1] For proper Bayesian updating, in a network setting with repeated actions, agents would have to know the full network structure, the precision of each agent's information, and the full timing of information arrival, and keep track of all these variables simultaneously. Alternatively, in addition to their beliefs, agents would have to reliably communicate the sources of information, the sources of their sources, and so on. In such cases, the state-space expands rapidly with the number of rounds.[2] The cognitive demands required to carry out such procedures are substantial. Consequently, the literature has identified simpler, less demanding updating heuristics. Departing from Bayesian learning in such environments opens up a new dimension of interest: it allows for the timing and order of information arrival to affect the final beliefs.

To study the effect that the timing and order of information have on beliefs formed by a group, we analyze a setting in which a group of agents repeatedly guess a state of interest while observing previous guesses of other group members they are connected to. Concretely, we extend the classic DeGroot (1974) naive social-learning model to allow for sequential information arrival. Under DeGroot learning, agents update their beliefs by taking weighted averages of their past beliefs and the past beliefs of other group members they pay attention to. Importantly, the weight each agent places on others is assumed to be fixed. The reduced complexity requirement, coupled with the tractability of DeGroot learning, has led to the model's proliferation in the study of social learning with repeated updating.[3] In the model we analyze, each agent is associated with a private signal, a time at which the signal arrives, and a fixed weight the agent places on the signal once it is received. Information arrives at specific rounds. In particular, a single agent or a group of agents may receive their private signals at different times. After each information release, communication takes place, and the new information disseminates in the network, leading to a new consensus. Subsequently, other agents receive their private signals, followed by communication, and so on; until all agents receive their private signals. From this point onward, communication takes place as in the DeGroot

[1] See Section 1.2 for the relevant work.

[2] For further discussion see DeMarzo et al. (2003). Alternatively, we can consider a setup in which agents do not have complete information of the network structure or the quality and timing of the signals. In such cases, based on the observable information, agents would have to update their beliefs about the network structure and features of information quality and timing in each period. Doing so is at least as cognitively demanding as the setup discussed above.

[3] Several experimental papers find that the DeGroot model explains observed data well, while recently, an axiomatic foundation of the model has also been established. See Section 1.2 for the relevant papers.

model.

Our first finding is that the sequence of information arrival affects the group's final beliefs, even when the network structure and the signals associated with each agent are fixed. In other words, although the environment and the objective evidence are unchanged, the order by which this evidence is presented leads to different final beliefs. The influence a signal has on the group's final beliefs depends on the weight that the agent receiving it places on the signal, the network influence of this agent, and the time at which the signal was received. As it turns out, the earlier a signal is released, the lower its weight will be on the final beliefs formed by the group. This observation allows us to identify the information release sequences that yield the highest and lowest attainable consensus. In particular, these sequences that attain the extreme consensus values release information in a monotonic order, from lowest to highest, or vice versa, with a possible joint release of information in the last round. By identifying these sequences, we bound the variation in the final beliefs that can be attributed to the timing and order of information release. Thus, within our framework, whatever the timing and order of information turn out to be, the final beliefs must fall within the identified bounds.

Next we analyze features of the underlying network that affect its susceptibility to the sequencing of information. In particular, we take an ex-ante approach and assess how the expected gap between the highest and lowest attainable consensus beliefs differ as the *in uence* of each group member changes. Formally, *in uence* corresponds to the eigenvector centrality of an agent. This measure is commonly used in the social-network literature as a proxy for agents' "centrality". An agent's influence depends on the network structure, how many other members listen to the particular group member, and what weights they place on her opinion. We find that the expected gap is maximized when the influence is uniformly distributed across all group members. That is, groups in which each member has equal voice are groups most susceptible to manipulation through information timing.

Our final set of results relates to the group's ability to adequately aggregate information as the number of group members grows. The *wisdom of crowds*, first analyzed in Golub and Jackson (2010), states that under DeGroot learning, in large societies where no agent is disproportionately influential, beliefs converge to the realized true state, as would beliefs under optimal updating. With sequential information arrival, however, we find that *wisdom* generally fails, and beliefs converge away from the truth even as the number of agents grows large. We emphasize that *wisdom* typically fails if agents have a prior with regard to the state they aim to learn. In cases with limited information, the prior helps create a more efficient estimate of the realized state. However, when information is abundant and sufficient for the realized state to be fully revealed, the impact of the prior optimally washes away. In the current setting, although the influence of the prior diminishes with each round of information release, this influence does not reduce all the way to zero, and consequently, the prior affects the final consensus.

We show that even in cases in which agents have an uninformative prior—think of learning about a new product, a new political candidate, etc.—their beliefs may nonetheless converge away from the truth if information is released sequentially. What causes *wisdom* to fail in this case is the distorted weights placed on signals arriving at different times. Because these weights depend on the timing of the signals' release, if there is any correlation between the value of the signals and the round at which they are released, the final beliefs converge away from the truth. However, *wisdom* persists if agents have an uninformative prior and there is no relation between the value of the signal and the round at which it is released. This set of findings reveal that once we consider an environment with sequential information arrival, the ability to adequately aggregate information is greatly compromised, in all but very specific cases *wisdom* fails.

Taken together, this work emphasizes that not only the information content but the timing and order by which this information arrives play a role in determining the final beliefs formed via naive social learning.

1.2 Related Literature

In the sequential learning literature, where agents act only once, the Bayesian learning approach has been quite standard, see Banerjee (1992), Bikhchandani et al. (1992), Smith and Sørensen (2000), Acemoglu et al. (2011), Lobel and Sadler (2015). By acting only once, agent's strategic concerns are rather limited, making Bayesian learning seemingly a reasonable assumption. That is not to say that non-Bayesian models are not present in this setting, see Eyster and Rabin (2010), Bohren (2016), Dasaratha and He (2020), to name a few.

The Bayesian paradigm has been applied on settings with repeated actions in paper such as Gale and Kariv (2003), Rosenberg et al. (2009), Mueller-Frank (2013), Mossel et al. (2015), Mossel et al. (2020). These papers either limit the strategic interactions, for example, Gale and Kariv

(2003) assume that agents observe only the distribution of actions, and that each agent is a small enough part of society for their actions to not affect aggregate outcomes, and consequently, greatly reduce strategic concerns; or study aspects such as whether there will be agreement and proper information aggregation at the asymptotic steady state. While understanding features of Bayesian learning in this setting is of great importance, the complexity required to carry out such learning in practice is immense. Hazla et al. (2019) study the complexity required for Bayesian learning in repeated action settings, and in line with what has anecdotally been accepted for a long time, find that it is rather extreme.[4]

That carrying out proper Bayesian updating is so complex presents an issue since experimental works such as Kübler and Weizsäcker (2004), Choi et al. (2008), Choi et al. (2012), Corazzini et al. (2012), Eyster et al. (2015), Enke and Zimmermann (2017), Brandts et al. (2015), Chandrasekhar et al. (2020) show that people tend to make far from optimal choices in rather straightforward environments requiring basic inference. To avoid the requirements for complexity that seems far beyond what humans can reasonably achieve, and to gain additional tractability, there is a large body of literature studying social learning with repeated actions under non-Bayesian learning. Papers such as Bala and Goyal (1998), Levy and Razin (2018), Mueller-Frank and Neri (2020), fall within this category. Non-Bayesian papers closest to this work are DeGroot (1974), DeMarzo et al. (2003), Golub and Jackson (2010) and Banerjee et al. (2019).

To a large extend, the DeGroot model, first analyzed in DeGroot (1974), and further studied and motivated in DeMarzo et al. (2003) in a framework relevant to this paper, has emerged as the canonical social-learning model in settings with repeated actions. This model, with it's rather straightforward heuristics, provides a simple way of updating beliefs, and thus, does not suffer from the overwhelming complexity as would be the case with proper Bayesian updating. The model is also very tractable, and calculating where beliefs ultimately converge is straightforward. In addition, experimental studies such as Chandrasekhar et al. (2020), Grimm and Mengel (2020), Brandts et al. (2015) suggest that the model approximates observed behavior often better than the benchmark Bayesian model.[5] In a companion paper, Reshidi (2020) we reach similar conclusions

[4]In particular Hazla et al. (2019) show that the problem is PSPACE-hard.

[5]In a village field experiment and in a student experiment, Chandrasekhar et al. (2020) find the share of Bayesian agents to be 10% and 50% respectively. Grimm and Mengel (2020) find that in explaining individual level decisions the Bayesian model is outperformed by the DeGroot model. Brandts et al. (2015) find that, in line with the DeGroot model, agents who have more outgoing degrees have a greater influence on the final consensus.

in a setting where information arrives sequentially. Furthermore, Molavi et al. (2018) develop axiomatic foundations of the DeGroot model.[6]

Finally, of relevance for this project is the work of Golub and Jackson (2010) and Banerjee et al. (2019). The former studies the information aggregation properties of the DeGroot model, and identifies conditions under which, as the number of agents grows, the final beliefs of the group converge to the truth; while the latter extends the model to allow for uninformed agents.

2 Model

2.1 Model Description

A set of agents, connected via a social network, seek to figure out a fundamental truth. Each individual receives a noisy private signal with regard to the truth. Not all individuals receive information at the same time. For some, the private signals may arrive early on, while others might have to wait longer to receive their private signal. Since each agent's signal is noisy, agents benefit from communicating with one another. Importantly, the network serves as a channel of information dissemination. The network structure is represented by a directed graph illustrating to whom each agent listens to and how much they "trust" each of these sources. This network structure is assumed to be exogenous. It may represent social networks, organizational relations, geographical proximity, etc.

Agents communicate with one another between each round of information arrival, allowing information to disseminate across the network. In such a setup, Bayesian agents would adjust the "weight" they place on their neighbors' opinions multiple times. In practice, however, there are several reasons why agents may fail to adjust these weights perfectly. To adequately accommodate these weights agents would have to not only perfectly know the network structure, but also know the exact timing of information arrival for each other agent in the network. Furthermore, they would need to have common knowledge of rationality, common knowledge that all other agents know the whole network structure, and know the exact timing of information arrival, so that they know how to weigh each opinion of their neighbors adequately. Even with this knowledge, agents

[6] Molavi et al. (2018) find that DeGroot learning it is the unique learning rule to satisfy *imperfect recall*, *label neutrality*, *monotonicity* and *seperability*.

would have to keep track of a state-space that grows rapidly after each round, as they would have to keep track of the whole network structure and how each of their neighbors' opinions is affected. The same demand for knowledge and complexity is related to the weights agents place on their private signal once it arrives. The central assumption made in this paper is that instead of adjusting the weight that agents place on their neighbors and their own signal after each round, agents continue to "value" their neighbors' beliefs with the same weight and set the same weight on their signal once it arrives, regardless of when it arrives.

In this section, we assume that there are enough rounds of communication between each information release round to allow for the network beliefs to converge. This assumption is made for tractability purposes, and we relax it later on. We identify where beliefs finally converge after all information has been released and disseminated in the network. The final consensus turns out to be a weighted average of the initial consensus, and each agent's private signals. The weight that a signal attains depends on the importance that the agent who received it initially places on it, the social influence of that agent, and the time when that signal was received.

2.2 Formal Model

A finite set $N = \{1, 2, \ldots, n\}$ of agents interact with each other through a social network. Each agent is represented by a node while their interactions are captured by a $n \times n$ nonnegative matrix M. The matrix M is a stochastic matrix, implying that the sum of each row is normalized to add up to one. The matrix may, but is not required to be symmetric, that is m_{ij} is not necessarily equal to m_{ji}, where m_{ij} represents the entry on row i and column j of the matrix M. Entry m_{ij} captures the attention, or the weight that agent i places on agent j.

Agent i in time $t \in \{0, 1, \ldots, \}$ has belief $b_{i,t} \in \mathbb{R}$. We let b_t represent the vector of beliefs of all agents in time t. Agents start with a common belief c_0, thus, all entries of b_1 are equal to c_0. Associated with agent i is a signal s_i which the agent receives at time $\hat{t}_i$. Let λ_i represent the weight agent i places on their own signal once it arrives. Let $\gamma(k)$ represent the agents for whom information arrives in time k, that is, $i \in \gamma(k)$ if $\hat{t}_i = k$. The signals arrive after communication has taken place in that round; hence, the vector of beliefs evolves as follows[7]

$$b_t = (I - \Gamma_t) \circ M b_{t-1} + \Gamma_t \circ ((I - \Lambda) \circ M b_{t-1} + \Lambda s) \quad (1)$$

Where $\circ$ represents the element-wise product or the Hadamard product, Γ_t is a diagonal matrix with $\gamma_{ii} = 1$ if $i \in \gamma(t)$ and $\gamma_{ii} = 0$ otherwise, and Λ is also a diagonal matrix with λ_i being the diagonal entry in row i and column i. Thus, the beliefs of agent i are as follows

$$b_{i,t} = \begin{cases} (1 - \lambda_i) \left(\sum_{j=1}^{N} m_{ij} b_{j,t-1} \right) + \lambda_i s_i & if \quad t = \hat{t}_i \\ \sum_{j=1}^{N} m_{ij} b_{j,t-1} & if \quad t \neq \hat{t}_i \end{cases}$$

Hence, if agent i does not receive their private signal in round t, $t \neq \hat{t}_i$, they form their new beliefs by simply taking a weighted average of their own previous period beliefs, as well as the beliefs of the agents they pay attention to, or are connected via a social network with, all agents $j \neq i$ such that $m_{ij} > 0$. When, however, agent i receives a signal, beliefs are updated in the same manner and afterwards the signal is incorporated with weight λ_i.

Note that in all rounds in which participants do not receive a private signal updating is carried out as in the classic DeGroot model. That is, for any agent, beliefs in time t are a weighted average of their own and their neighbours beliefs in time $t - 1$. The departure is then with regard to incorporating private signals in a sequential manner. We can think of the DeGroot model as a special case of the current model where all private signals arrive in round $t = 1$, and all that is left is for agents to repeatedly calculate weighted averages of their social network until a consensus emerges.

Similar to DeMarzo et al. (2003) who motivate the evolution of beliefs in the classic DeGroot mode as a bounded rationality model, we can think of a bounded rationality setup for the current setting. We can interpret the current setting as one in which a finite set of agent $N = \{1, 2, \ldots, n\}$ attempt to estimate a parameter of interest $\theta \in \mathbb{R}$. Signals participants receive can be thought of

[7] For the analysis in this paper this will be equivalent to information being processed at the beginning of the round, with the alternative release time $\tilde{t}_i = \hat{t}_i + 1$, and the equation representing the the evolution of the vector of beliefs would be

$$b_t = M \left((I - \Gamma_t) b_{t-1} + \Gamma_t \left((I - \Lambda) b_{t-1} + \Lambda s \right) \right)$$

as noisy measures of the parameter of interest, $s_i = \theta + \varepsilon_i$ with $\mathbb{E}[\varepsilon_i] = 0$. While each signal is unbiased with respect to the parameter of interest, each signal is noisy, thus, agents can improve their estimation of θ by paying attention to the actions of others. In line with this interpretation, we can think of the initial consensus c_0 as the mean of the prior distribution of θ. If the prior and the individual signals are normally distributed, and the signals are independent, the weight that a Bayesian agent would place on each signal would be proportional to the precision of the signal. We can then think of the weight that agent i places on the opinion of agent j, as being proportional to the signal precision of agent j. Similarly, the weight that the agent place on their own signal need not be ad hoc, rather, it may be proportional to the precision of the signal. Of course, the bounded rationality element comes into play as agents continue to use the same weight on thought all rounds of communication. For a more in depth discussion of the bounded rational interpretation refer to DeMarzo et al. (2003).

2.3 Belief Convergence

Let us begin by analyzing how beliefs evolve within rounds in which no new information is released. Let b_t represent the vector of beliefs that agent's hold in period t, and let M once more represent the network structure. No information release in equation 1 corresponds to Γ being a null matrix. As a result, beliefs in the next period can be expressed as $b_{t+1} = Mb_t$. That is, in line with the DeGroot model, the new formed beliefs of each agent are weighted averages of the beliefs of the agents they pay attention to, possibly including themselves. Continuing in this fashion, as long as no new information is released, beliefs in round $t+2$ can then be expressed as $b_{t+2} = Mb_{t+1} = M(Mb_t) = M^2 b_t$. We can express beliefs with k rounds of no information release as $b_{t+k} = M^k b_t$. A natural question then arises, with no further information release, if agents continue to communicate with one another, will their beliefs converge and thus give rise to a consensus in beliefs?

The matrix M can be thought of as a transition matrix of a Markov chain. It is well known that if the matrix M is strongly connected and aperiodic then M is convergent.[8]. This matrix

[8]See Kemeny and Snell (1960). A matrix is strongly connected if there is a path from any node, to any other node. The period $d(k)$ of a state k of a matrix M is given by $d(k) = gcd\left\{m \geq 1 : M^m_{k,k} > 0\right\}$. Where $gdc(\cdot)$ stands for greatest common divisor. If $d(k) = 1$, then state k is aperiodic. Matrix M is aperiodic if and only if all its states are aperiodic. An alternative definition of aperiodic is the following. The chain is aperiodic if and only if there exists a positive integer n such that all elements of the matrix M^n are strictly positive, $[M^n]_{ij} > 0 \; \forall \, i \, j$. A strongly connected matrix is aperiodic, trivially if each agent places at least some weight on their own past actions $m_{ii} > 0$.

has a unique eigenvalue equal to one, while all other eigenvalues have modulus smaller than one. Furthermore, there is a unique left row eigenvector π satisfying $\pi M = \pi$ corresponding to eigenvalue 1, with $\sum_i \pi_i = 1$, such that for any initial vector of beliefs b_t

$$\lim_{k\to\infty} b_{i,t+k} = \lim_{k\to\infty} \left(M^k b_t\right)_i = \pi b_t$$

Thus, with a strongly connected and aperiodic network matrix M if enough rounds of communication take place between information release rounds, beliefs will converge to a consensus. Furthermore, the consensus will be formed by a convex combination of the starting beliefs, where the weight of each belief is represented by the corresponding π_i value on the left eigenvector. We can think of π_i as the influence of agent i, as it captures the weight that the initial belief of agent i has on the consensus.

We now consider where beliefs converge after each round of information release, and where beliefs converge after the final round of information release. We will denote beliefs as $b_{i,t}^{(k)}$, where i represents the agent, k stands for the number of information release rounds, while t represents the round number since the last information release. Furthermore, initial beliefs after new information has just been released will be denoted as $\tilde{b}_i^{(k)} = b_{i,1}^{(k)}$ while beliefs after communication has taken place will be represented by $\hat{b}_i^{(k)} = b_{i,\infty}^{(k)}$. From the discussion above, we know that after enough rounds of communication beliefs will once more converge, and for any i and j, $\hat{b}_i^{(k)} = \hat{b}_j^{(k)} = c^{(k)}$, where $c^{(k)}$ represents the new consensus, after the k'th round of information release. From the above discussion we also know that the new consensus will be equal to $c^{(k)} = \sum_{i=1}^{N} \pi_i \tilde{b}_i^{(k)}$. Thus the beliefs immediately after new information is released are

$$\tilde{b}_i^{(k)} = \begin{cases} (1-\lambda_i)\, c^{(k-1)} + \lambda_i s_i & if \quad i \in \gamma(k) \\ c^{(k-1)} & if \quad i \notin \gamma(k) \end{cases} \tag{2}$$

While the beliefs after information has been released and communication takes place will be

$$\hat{b}_i^{(k)} = c^{(k)} = \left(1 - \sum_{j\in\gamma(k)} \pi_j \lambda_j\right) c^{(k-1)} + \sum_{j\in\gamma(k)} \pi_j \lambda_j s_j \tag{3}$$

After each round of information release, and after communication takes place, a new consensus arises. The new consensus is a convex combination of the old consensus and the signals released in that information round. The weight that the signals receive depends initially on the weights that the agents who received them place on their own signals, namely λ_i. Naturally, if agents ignore their own signals, the signals would fail to affect the beliefs of others. Furthermore, the impact that each signal has on the new consensus depends on the influence measure, π_i, of the agent receiving the signal. Not surprising, regardless of how much weight an agent places on their own signal, if the rest of society places little weight on them, their signal will have a modest influence on the new consensus. We have now laid the groundwork to state the following proposition.

Proposition 1 (Belief Convergence). *Let M be a strongly connected and aperiodic matrix representing the social network, Λ a vector of weights agents place on their own signal, once they arrive. Furthermore, assume that between each information round communication takes place for enough periods for a consensus to be reached. After the nal round of information release, the beliefs of all agents converge to the consensus belief*

$$c^{(K)} = \sum_{k=1}^{K} \left(\prod_{z=k+1}^{K} \left(1 - \sum_{j \in \gamma(z)} \pi_j \lambda_j \right) \right) \sum_{i \in \gamma(k)} \pi_i \lambda_i s_i + \prod_{k=1}^{K} \left(1 - \sum_{j \in \gamma(k)} \pi_j \lambda_j \right) c^{(0)}$$

$c^{(K)}$ represent the final consensus after all information has been released and communication has taken place. Recursive replacements of $c^{(k)}$ in equation 3 leads to the equation in Proposition 1. As can be seen, if a signal is released in information round k the weight it has on the final consensus is multiplied by $\prod_{z=k+1}^{K} \left(1 - \sum_{j\in\gamma(z)} \pi_j \lambda_j\right)$. Since each $\left(1 - \sum_{j\in\gamma(z)} \pi_j \lambda_j\right) < 1$, the earlier a signal is released, the more it will be discounted, and hence, the lower it's influence will be on the final consensus. By how much the weight of a signal is discounted, depends on the influence that agents for whom information arrives after have, as well as the weight they place on their own signals.

Hence, the final consensus will be a weighted average of the initial consensus and all the signals that agents receive. The influence of each signal depends on the weight that the agent who received it placed on the signal, the social influence of this agent, as well as the time in which the signal was received.

2.4 General Weights

From Proposition 1 it becomes clear that the influence the signals have on the final consensus depends on the timing of their arrival. Consequently, the final consensus itself depends on the sequencing of information arrival.

As emphasized above, one of the assumptions is that agent i places weight λ_i on her signal, once she receives the signal, regardless of when the signal arrives. However, Proposition 2 below reveals that this is not what drives the dependence of the final consensus on the sequencing of information. Let n represent the number of agents and K the number information release rounds.

Proposition 2 (General Weights)**.** *There exists no $\lambda_i^{(k)}$ with $i \in \{1, 2, \ldots, n\}$ and $k \in \{1, 2, \ldots, K\}$ such that the nal consensus $c^{(K)}$ is independent of the sequencing of information arrival.*

That is, even allowing for the weight agents place on their signal to differ for each agent, and each information arrival round, the final consensus continues to be affected by the sequencing of information arrival for all possible weights. Knowing that this is not the driving force of the dependence, for tractability purposes we maintain the assumption of a fixed λ_i.

3 Sequential Information Arrival

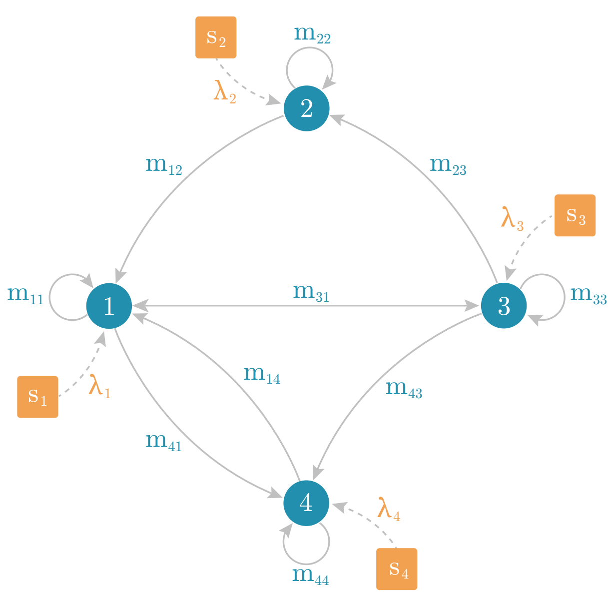
3.1 Example: A Committee Evaluating a Candidate

Consider a committee evaluating the quality of an applicant. Given the common information available for the candidate, such as the resume, presentation, and so on, the committee members form some belief c_0 with regard to the quality of the candidate. In addition to the common knowledge, each committee member receives an additional private signal with regard to the candidate's quality. In the case of a hiring committee evaluating an academic job market candidate, the private signals may be direct phone-calls with the candidates advisor, an additional signal drawn from reading the participants job market paper once more, and so on.[9]

[9] Consider the following alternative interpretation. c_0 may represent the expected value of the pool of candidates. In addition to the prior knowledge, committee members have access to the same information for the candidate, such as the candidate's resume, job market paper, etc. However, the committee members access this information at different times. The signals that each member receives can be thought of as their subjective interpretation of the information.

More formally, consider a social network composed of four members. Member i places weight m_{ij} on the opinion of member j, and weight λ_i on their own signal, once they receive the signal. Each communication period, which could be thought of as a chat with other members, committee members update their beliefs with regard to the quality of the candidate, where their new beliefs are simply a weighted average of the beliefs of the members they pay attention to, including themselves. One such network is depicted on Figure 1. A directed link from j to i implies that member i pays attention to member j, that is $m_{ij} > 0$. Missing links from some member j to some member i imply that member i pays no attention to, or places weight 0 on the opinion of member j. This could be a result of physical or organizational constraints, or it could simply imply that member i chooses to disregard the opinion of member j.

Figure 1: Comittee Example



The network structure can be represented by a matrix M, with entries m_{ij} representing the weight members place on each others opinions. Consider a concrete example with the following values

$$M = \begin{bmatrix} 0.40 & 0.15 & 0.10 & 0.35 \\ 0 & 0.65 & 0.35 & 0 \\ 0.55 & 0 & 0.45 & 0 \\ 0.25 & 0 & 0.20 & 0.55 \end{bmatrix} \qquad \lambda = \begin{bmatrix} 0.70 \\ 0.45 \\ 0.65 \\ 0.70 \end{bmatrix} \qquad c_0 = 7 \qquad S = \begin{bmatrix} 2 \\ 3 \\ 5 \\ 9 \end{bmatrix}$$

The vector λ represents the weights that members place on their own signals once they receive them, c_0 represents the initial beliefs, while the vector S represents the vector of signals that members eventually receive. The left row eigenvector π corresponding to eigenvalue 1 of the above matrix M is equal to $\pi = [\,0.34 \;\; 0.15 \;\; 0.25 \;\; 0.26\,]^T$. Hence, in this particular network the most influential agent turns out to be the first agent. This is according to the traditional influence measure, which would capture the amount that each agent influences the final consensus if all information is released jointly. We now consider two possible signal arrival sequences.

All Signals Jointly In the first case, all signals arrive jointly in the very first round. Each agent incorporates their private signal leading to the following beliefs $b_{i,1}^{(1)} = (1 - \lambda_i)c_0 + \lambda_i s_i$. From this point onward communication takes place. Committee members communicate their beliefs, and form new beliefs based on the other members they pay attention to. After enough rounds of communication beliefs converge.[10] From Section 2.3 we know that beliefs converge to the following value

$$c_T = \sum_{i=1}^{4} \pi_i b_{i,1}^{(1)} = \sum_{i=1}^{4} (1 - \pi_i \lambda_i) c_0 + \sum_{i=1}^{4} \pi_i \lambda_i s_i = 5.59$$

Thus, if all information arrives at the same time and communication takes place via the above described social network, eventually all members would end up believing that the quality of the candidate is 5.59.

Two Rounds of Information Arrival Now consider an alternative sequence of information arrival. The first two members receive their signals immediately, whereas the signals of the other

[10] It is straightforward to check that the matrix M is strongly connected and aperiodic, thus we know that eventually a consensus will arise.

two members are delayed. Beliefs after the first two signals are received will be

$$b_{i1}^{(1)} = \begin{cases} (1-\lambda_i)c_0 + \lambda_i s_i & if \quad i \in \{1,2\} \\ c_0 & if \quad i \in \{3,4\} \end{cases}$$

After the first two members receive and incorporate their signals communication takes place and the information disseminates across the network. Once more from Section 2.3 we know that the beliefs of the members eventually converge to

$$c_1 = \sum_{i=1}^{4} \pi_i b_{i,1}^{(1)} = 5.54$$

Where c_1 represents the consensus reached after the first round of information release, followed by communication. Afterwards, the signals the other two members also arrive, and the new beliefs are as follows

$$b_{i1}^{(2)} = \begin{cases} (1-\lambda_i)c_1 + \lambda_i s_i & if \quad i \in \{3,4\} \\ c_1 & if \quad i \in \{1,2\} \end{cases}$$

Communication takes place once more and the new information disseminates in the network. The final consensus, after all four members receive their private signals and communicate is then

$$c_2 = \sum_{i=1}^{4} \pi_i b_{i,1}^{(2)} = 6.10$$

With this particular sequence of information arrival, the committee members end up believing that the quality of the candidate is 6.10. As can be seen, although the network structure, the weight committee members place on their own signals and other group members opinions, and the signals associated with each member are unchanged, the sequence of information arrival affected the final consensus that the group reached. If the committee was comparing the current candidate with another candidate, the quality of whom they evaluate to be 6, the sequence of information arrival could have determined whether the candidate gets offered the job.

If we were to calculate the final consensus under other information release sequences, we would find that the group could have ended up with other final beliefs. Some of which would be lower, in between, and even higher than the beliefs reached by the two sequences considered in this example. A natural question then arises. Is it possible to bound all final consensuses that might arise as a result of different information arrival sequences, while keeping the network structure, the weights agents place on their own signal as well as the signal realizations unchanged? If such bounds exist, would it be possible to identify which information release sequence could produce the most extreme final beliefs? The next section gives a positive answer to both of these questions.

3.2 Optimal and Pessimal Information Release Sequences

In the example analyzed in Section 3.1 the timing of information arrival was found to affect the final consensus reached by the agents, even though the network structure and the information content remained unchanged. In this section we identify the information release sequence that leads to the highest and lowest attainable final consensus, and in doing so bound all possible final consensuses that may arise as a result of the timing of information arrival. That is, under the particular model specifications, whatever the actual sequencing of information turns out to be, the final consensus will fall withing our identified bounds.

Proposition 3 (Optimal Information Sequence)**.** *Without loss of generality let that* $s_i < s_j$ *if* $i < j$*. Then the nal consensus is maximized under the following information sequencing*

$$\overline{\gamma}(k) = \begin{cases} k & if \quad c^{(k-1)} \geq s_k \\ \{k, k+1, \ldots, n\} & if \quad c^{(k-1)} < s_k \end{cases}$$

That is, the information sequence that yields the highest attainable consensus releases information sequentially in a non-decreasing order, starting from the signal with the lowest value. Under this information release, the sequential release continues until the prevailing consensus, if ever, becomes lower than the lowest signal not yet released; in which case, all signals are released jointly. To build some intuition for this result, note from Proposition 1 that a signal released earlier will have a lower weight on the final consensus compared to the weight it would have if the same signal was released in later information rounds. Naturally then, if we aim to maximize the final consensus we would lead with the low valued signals first, as these signals will be more heavily discounted the earlier they are released.

Corollary 1 (Pessimal Information Sequence)**.** *Without loss of generality let that* $s_i < s_j$ *if* $i < j$. *Then the nal consensus is minimized under the following information sequencing*

$$\underline{\gamma}(k) = \begin{cases} n+1-k & if \quad c^{(k-1)} \leq s_{n+1-k} \\ \{1, \ldots, k\} & if \quad c^{(k-1)} > s_{n+1-k} \end{cases}$$

That is, the information sequence that yields the lowest attainable consensus releases information sequentially in a non-increasing order, starting from the signal with the highest value. Under this information release, the sequential release continues until the prevailing consensus, if ever, becomes higher than the highest signal not yet released; in which case, all signals are released jointly. The intuition for this corollary is similar to the intuition for Proposition 3.

While Proposition 3 identifies the information release sequence that yields the highest attainable final consensus, Corollary 1 identifies the information release sequence that yields the lowest attainable final consensus. In doing so, within the analyzed framework, we bound all the possible final consensuses that may arise as a result of the changes in the timing of information arrival.

Furthermore, Proposition 3 and Corollary 1 reveal that to identify the optimal and pessimal information release sequence we do not need to search through n^n possible sequences, which is the total number of possible sequences. Instead, for the optimal information release there are only n possible releases we need to consider, in which, only the "threshold" beyond which signals are released jointly differs. The same "threshold" logic holds for the pessimal sequence. Thus, the problem of finding the optimal and pessimal information release sequence greatly reduces from one searching through n^n possible sequences, to one searching through $2n$ possible sequences. Which sequence turns out to be optimal and pessimal is determined by the realization of the signals.

We can now revisit the example in 3.1 and see what the highest and the lowest attainable consensuses would be. Any other consensus arising from any other information release timing would fall in-between these boundaries. For that specific example the highest consensus would be attained if signals were released sequentially, from lowest to highest, yielding a consensus of 6.17. The lowest consensus would have been attained by releasing the highest signal first, followed by all three signals in the second information release round, yielding a consensus of 5.42. Recall that the consensuses found in 3.1 were 5.54 under the joint release and 6.10 in the two period release.

Figure 2: Upper and Lower bound Example

As can be seen, the consensus arising from the joint release, as well as the consensus arising from the two round release, both fall within the identified bounds. Neither of these sequences was optimal or pessimal, and consequently, the resulting consensuses is strictly within the bounds.

While in this section, for tractability purposes, we have assumed that after each round of information release communication takes place until a new consensus is reached, this is not necessary for the identified information release sequences to remain optimal/pessimal. This is made clear in the proposition below.

Proposition 4 (Communication Rounds)**.** *Let r represent the rounds of communication between information release rounds. Then $\exists \tilde{r}$: if $r > \tilde{r}$ the identi ed sequences leading to the extreme consensuses remain unchanged.*

Proposition 4 states that there always exists some $\tilde{r}$ such that if communication takes place for at least $\tilde{r}$ rounds between each round of information release, the optimal and pessimal information release sequences identified earlier, remain optimal and pessimal. That is, "enough" communication suffices, and consequently convergence of beliefs between information release rounds is not necessary for the identified sequences to generate the extreme consensuses.

3.3 A Large Society with Limited Information Release Rounds

In the analysis above we assumed that we may have as many information release rounds as needed. This is likely to be the case in small networks, since we only need a few rounds to accommodate the sequential release of the signals associated with each agent. However, in a large society with a high number of agents, it becomes likely that the number of signals far outweighs the number of information release rounds. In this section, we study properties of the sequences that lead to the highest and lowest attainable consensuses in a large society.

A large society is captured by a sequence of networks where the number of agents n grows. Specifically, $(M(n))_{n=1}^{\infty}$ represents a sequence of n-by-n interaction matrices indexed by n, the number of agents in each network. As we are interested in analyzing where the final beliefs of the group converge, we maintain the assumption that each matrix is strongly connected and aperiodic. Let $(\pi(n))_{n=1}^{\infty}$ represent a sequence of influence vectors associated with each network n. We impose the following assumption

$$\max_{i \in \{1,\ldots,n\}} \pi_i(n) \to 0 \quad as \quad n \to \infty$$

That is, as the network grows, the influence of the most influential agent goes to 0. This ensures that no agent is disproportionately influential. Furthermore, let $(\lambda(n))_{n=1}^{\infty}$ be a sequence of vectors, representing the weights that agents place on their private signals. To ensure a nontrivial problem we assume that $\frac{1}{n}\sum_{i=1}^{n}\lambda_i(n)$ is bounded away from 0 for all n. Finally let $s_i(n)$ represent a sequence of signals associated with agent i in network n.

Proposition 5 (Optimal Information Sequence with Limited Rounds)**.** *Let $\bar{\gamma}$ represent the optimal information release sequence that yields the highest attainable consensus. Furthermore, let $\overline{\gamma}(k)$ represent the set of agents for whom information arrives in information round k. Then*

$$\max_{i \in \overline{\gamma}(k)} s_i < \min_{i \in \overline{\gamma}(k+1)} s_i$$

That is, under the optimal information release sequence, the highest value signal released in information round k must be lower than the lowest value signal released in information round $k+1$. Hence, the feature of the optimal structure identified with no limits on information rounds, namely, that signals are released in a non-decreasing fashion, prevails in a large society even with limited information release rounds.

The proof of the above proposition relies on showing that as the number of agents n increases, the final consensus from a sequence that releases a group of signals whose values are larger than another group of signals, but whose release round is earlier than the later group, can be increased by swapping the release date of these groups. As n increases, the added granularity makes it possible to select sub-groups such that the swap has no impact on the discounting of all other signals, leading to an unambiguous increase in the final consensus. Then, if such a swap is beneficial for any sequence, it can not be that there is space for such a swap in the optimal information release sequence. Implying that under the optimal information release round, signals are released in a monotonic fashion.

3.4 Influence and Sequence Susceptibility

We now analyze features of groups that affect their susceptibility to the sequencing of information arrival. In particular, we study conditions that maximize the expected gap between the highest attainable consensus, $\overline{C}(\pi, \lambda, s)$, and the lowest attainable consensus $\underline{C}(\pi, \lambda, s)$. Where π, λ and s, represent the influence vector, the vectors of weights agents place on their signals, and the agents' signals respectively. Let $\tilde{\theta}$, the state of interest, be a random variable drawn from distribution G with mean $c^{(0)}$. Let signals s_i be $i.i.d$ random variables drawn from a distribution F with mean θ, the realized state. Let λ_i be a $i.i.d$ random variable with distribution H on $[0, 1]$.

Proposition 6 (Influence and Sequence Susceptibility)**.** *Let $\overline{C}(\pi, \lambda, s)$ and $\underline{C}(\pi, \lambda, s)$ represent the highest and lowest attainable consensus respectively. The expected gap $\mathsf{E}\left[\overline{C}(\pi, \lambda, s) - \underline{C}(\pi, \lambda, s)\right]$ is maximized for a in uence vector π with uniform in uences $\pi_i = \frac{1}{N}\ \forall\, i$, where N is the number of agents. As the in uence concentrates towards an opinion dictator, $\max_i \pi_i \to 1$, the gap shrinks to 0.*

Proposition 6 states that groups that are most susceptible to the different sequencing of information arrival are those in which each member has identical influence. To see why this is the case, recall that in the final consensus the weight of each signal, among other things, depends on the influence of the agent who receives it, and is discounted based on the influence of the agents receiving signals in later rounds. By having uniform weight across all agents, when constructing the optimal and pessimal information sequence, there are numerous configurations to choose from. Consequently, in expectation, reshuffling the signal release timing leads to substantial variation in the final consensus, leading to a wide gap between the highest and lowest attainable consensus.

On the other hand, when each entry of π is not identical, it is not clear which π_i value will be associated with which realized signal. Although the π values are fixed while signals are random, realizing that the optimal/pessimal information release sequence releases information in a monotonic fashion, leads to an alternative interpretation, in which, the order statistics of the signals are considered fixed, whereas there is variation with regard to which π_i value will be associated with which order statistic. It turns out that this variation does not benefit the signals released in the last round, while boosting the weight of all previously released signals. Consequently, the potency to generate a high/low final consensus shrinks, leading to a decrease in the expected gap.

In the extreme case, when $\max_i \pi_i = 1$, all sequences result in the same final consensus, as now, there is an "opinion dictator". If $\max_i \pi_i = 1$, the sequence of information arrival no longer plays a role. Whenever the agent with $\pi_i = 1$ receives her signal, be it immediately, slightly later, or much later, the beliefs of the whole group converge to the beliefs of this agent. Since no other agent has influence, even if other agents receive signals afterwards, the weigh of the "opinion dictator" is not discounted, and thus, the newly found consensus remains unchanged.

4 The Wisdom of Crowds under Sequential Information Arrival

This section analyzes whether ***wisdom***, a concept to be formalized below, persists when information arrives sequentially. The finding behind the ***wisdom*** of crowds, as introduced and analyzed in Golub and Jackson (2010), is that, in a society in which agents assign fixed weights on each other's opinions—in a society in which learning takes place as in the DeGroot fashion—under relatively weak conditions beliefs converge to the correct state. This is a rather important result, as it suggests that society as a whole does not need to act optimally for information to be aggregated desirably; rather, by relying on the numerous sources of information, even with an imperfect information aggregation process, the final beliefs converge to the truth. The focus of this section is to analyze whether this result persists when information arrives sequentially.

4.1 The Wisdom of Crowds

We begin by defining the ***wisdom*** of crowds, as introduced in Golub and Jackson (2010). As we specified in Section 3.3, let $(M(n))_{n=1}^{\infty}$ represents a sequence of n-by-n interaction matrices indexed by n, the number of agents in each network. We maintain the assumption that each network is convergent. There is a true state of nature, $\theta \in [0, 1]$. At time $t = 0$, each agent i has an initial belief $b_{i,0}(n) \in [0, 1]$. These beliefs are assumed to be drawn independently from a distribution with variance $\sigma^2 > 0$.[11] One of the main assumptions made in this paper is that the distribution of these beliefs has mean equal to θ. For any given n and realization of beliefs $b_0(n)$, the belief of agent i in network n approaches a limit denoted by $b_{i,\infty}(n)$. Since these limiting beliefs depend on the realization of the initial beliefs, each limiting belief itself is a random variable. The sequence $(M(n))_{n=1}^{\infty}$ is wise if

$$\operatorname*{plim}_{n\to\infty} \max_{i\leq n} |b_{i,\infty}(n) - \theta| = 0$$

That is, a society is considered wise if as n increases, the final beliefs converge in probability to the realized value θ. Note that since the assumption that each matrix is convergent is maintained, we have

$$b_{i,\infty}(n) = \sum_{j=1}^{n} \pi_j(n) b_{j,0}(n)$$

Where $\pi_j(n)$ represents the influence of agent j in network n, the j'th value in the left eigenvector with eigenvalue equal to 1 of matrix $M(n)$.

Proposition 2 in Golub and Jackson (2010)

If $(M(n))_{n=1}^{\infty}$ is a sequence of convergent stochastic matrices, then it is wise if and only if the associated in uence vectors are such that $\max_{i\leq n} \pi_i(n) \to 0$ as $n \to \infty$.

The intuition of the result above is as follows. If the influence of the most influential agent shrinks to 0 as the number of agents grows, then, the idiosyncratic error associated with each agent washes away, and thus, the consensus converges to the mean of the beliefs, which is θ. As highlighted above, on key assumption is that the initial beliefs are drawn from a distribution with mean θ. If we think that agent's initial beliefs are formed by some signals they receive with regard to the realized value θ, for the agents to fully follow these signals, it must be that their prior on θ is diffused; otherwise, their beliefs at time $t = 1$, their optimal guesses, would lie between the signal they received and the mean of the prior distribution of θ.

In the sequential model studied thus far we have explicitly introduced a prior, which plays a

[11] The results follow even when each signal is drawn from a different distribution, as long as the support of each signal is bounded.

key role in determining agents' beliefs. Modifying the model so as to get rid of the prior raises the issue of handling belief updating when facing agents whose beliefs are an empty set, these would be the agents who do not receive a signal in the first round. Banerjee et al. (2019) study a setup in which all information arrives in the first round, however, not all agents receive a signal. They propose a learning rule when the beliefs of some agents are an empty set. We study some of the main takeaways of the model below, as they will become relevant when studying the ***wisdom*** of crowds in a sequential information arrival setting.

4.2 Alternative Specification: A Prior-less Model

Generalized DeGroot Model Banerjee et al. (2019) builds on the DeGroot model by introducing uninformed agents. Within this setup, at any point in time t an agent is either informed or uninformed. An informed agent at time t holds belief $b_{i,t} \in \mathbb{R}$, while an uninformed agent holds the empty belief $b_{i,t} = \emptyset$. The initial opinions of informed agents are an unbiased signal of the true state θ, drawn from some distribution F with finite variance:

$$b_{i,0} = \theta + \varepsilon_i \qquad \varepsilon \sim F(0, \sigma^2)$$

At time $t = 0$, a subset of agents receive a signal, while the remaining agents never receive a signal. All agents with which a node is linked are called neighbors of that node. Let the set J_i^t denote the set of informed neighbors of agent i at time t. The authors specify, what they name the ***generalized DeGroot*** updating as follows [12]

$$b_{i,t+1} = \begin{cases} \emptyset & if \quad J_i^t = \emptyset \\ \frac{\sum_{j \in J_i^t} b_{j,t}}{|J_i^t|} & if \quad J_i^t \neq \emptyset \end{cases}$$

Under this specification, agents who do not receive a signal are initially uninformed, until one of their neighbours becomes informed; in which case, they adopt the belief of their neighbour. If more than one neighbour is informed, agents average out these beliefs. One of the crucial assumptions in this paper is that agents have an uninformative prior on θ. That is, if an agent does not receive a signal, their beliefs are an empty set, and if an agent receives a signal, their belief is set equal to

[12] As discussed in Banerjee et al. (2019), the results generalize to non-uniform weighting.

this signal. Again, this would correspond to a case in which the prior is diffused, or in other words, the private signals are infinitely more informative than the prior; leading agents to fully adopt their signal as their guess once they receive a signal.

The authors find that when there are uninformed agents, even if $\max_i \pi_i(n) \to 0$ as $n \to \infty$ *wisdom* may fail, and consequently, beliefs may converge away from the mean of the signals received by the informed agents. However, for lattice graphs with shortcuts, even if information is sparse, the authors find that *wisdom* prevails; and as such, the final beliefs of the society converge in probability to the mean of the signals the agents receive.[13] In what follows, we will assume that the network structure satisfies this condition, and thus, *wisdom* does not fail as a result of signal sparsity in the first information arrival round.

Prior-less Sequential Information Arrival We can adapt this specification in the sequential setup. That is, we assume that the initial prior agents have on θ is uninformative. Consequently, we assume that for agents who do not receive a signal in the first information round, their beliefs are an empty set $b_{i,0}^{(1)} = \emptyset$ if $i \notin \gamma(1)$ and $b_{i,0}^{(1)} = s_i$ if they receive a signal, $i \in \gamma(1)$. To accommodate this, in our initial setup we must modify the weights agents place on their own signal. Specifically, let $\tilde{\lambda}_i$ represent the original weight participant i places on their signal. We modify the new weight as follows

$$\lambda_i = \begin{cases} \tilde{\lambda}_i & if \quad \tilde{\gamma}_i \neq 1 \\ 1 & if \quad \tilde{\gamma}_i = 1 \end{cases}$$

That is, when participants receive a private signal in the first round, since they have no prior, they set their beliefs equal to the signal they received. After the first set of signals are released, beliefs update as in the *generalized DeGroot model*, that is, participants pay attention only to informed neighbours.

$$b_{i,t+1} = \begin{cases} \emptyset & if \quad J_i^t = \emptyset \\ \frac{\sum_{j \in J_i^t} m_{i,j} b_{j,t}}{\sum_{j \in J_i^t} m_{i,j}} & if \quad J_i^t \neq \emptyset \end{cases}$$

[13]For more details and a formal definition of lattice graphs with shortcuts, see Banerjee et al. (2019).

After the first round of information release, and after communication takes place, there will no longer be any agents whose beliefs are an empty set. Thus, after the first round of information release and communication takes place, the dynamics under this new specification becomes identical to the dynamics in the initial model. With the difference being that the consensus reached after the first round is a convex combination only of signals released in information round 1, excluding the prior.

4.3 Persistence and Failure of Wisdom

Having defined the concept of *wisdom*, we now turn to analyze conditions and specifications under which *wisdom* persist when information arrives sequentially. With regard to the evolution of beliefs, we will consider both the initial specification, in which participants have a prior about θ, as well as the prior-less specification, which models the evolution of beliefs in the absence of a prior. Furthermore, with regard to the information release rule, we will consider two cases. An information release rule that maps signals to information release rounds without conditioning on the signal realization, as well as an information rule that conditions on signal realization when assigning signals to information release rounds.

Let $\tilde{\theta}$, the state of interest, be a random variable drawn from a non-atomic distribution G on $[0, 1]$. Let $\mathsf{E}[\tilde{\theta}] = c^{(0)}$ be the mean of the distribution. Furthermore let θ represent the realized value of $\tilde{\theta}$. Associated with each agent i in matrix n is a signal $s_i(n) \in [0, 1]$ drawn from a distribution F with variance σ^2 and mean θ. Let the weight the agent places on her own signal $\lambda_i(n) \in [0, 1]$ be drawn from a distribution G with variance σ_λ^2 and mean $\bar{\lambda}$. There exists a rule γ that maps information release rounds $\{1, 2, \ldots, k\}$ to agents $\{1, 2, \ldots, n\}$ for whom information arrives in that rounds. In particular $\gamma(n)[k]$ represents the group of agents for whom information arrives in information round k in network n. We may assume that this rule assigns nodes to different information release rounds independently of the realized signal $s_i(n)$. In this case, we assume that the rule specifies probabilities $\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_k$ such that $\sum_{i=1}^k \tilde{w}_i = 1$; where $\tilde{w}_j$ represents the probability that an agent will receive their signal in information round j.

Alternatively, we also consider a case in which the the mapping depends on the realized signals. In this case we assume that $i(n) \in \gamma(n)[j]$ if $s_i(n) \in S_j$; where $\cup_{j=1}^{k} S_j = [0,1]$ and $S_j \cap S_i = \emptyset$. That is, agent i receives her signal in information round j, if her signal falls within the S_j partition. We let S_k be an element of a measurable partitions of the set of values that a signal can take, such that the intersection of each element is empty and the union of all elements is equal to the whole set. Define the following

$$w_k = \int \mathbb{1}\{x \in S_k\} f(x)dx \qquad \mu_k = \frac{\int x \mathbb{1}\{x \in S_k\} f(x)dx}{w_k} = \mathbb{E}[s|s \in S_k]$$

Where w_k represents the mass of signals released in round k, while μ_k represents their conditional expectation. Recall from the law of iterated expectations that the proper way to "stitch" together the conditional expectations so as to retrieve the unconditional expectation is $\theta = \sum_{i=1}^{k} w_k \mu_k$; that is, we take a convex combination of the conditional expectations with weights equal to the mass of the signal within the partitions.

Initial model with Conditioning Under the initial model and a rule that conditions on signal realization, the limiting beliefs are

$$\operatorname*{plim}_{n\to\infty} c^{(K)} = \sum_{j=1}^{K} \left(\prod_{k=j+1}^{K} \left(1 - w_k \bar{\lambda}\right) \right) w_j \bar{\lambda} \mu_j + \prod_{j=1}^{K} \left(1 - w_j \bar{\lambda}\right) c^{(0)} \neq \theta$$

Where w_j represents the mass of signals released in information round j, $\bar{\lambda}$ the average weight agents place on their signal and μ_j, the expected value of signals released in information round j. The above expression differs from θ for two reasons. First, the weights associated with the conditional expectations are a distorted version of the weights prescribed by the law of iterated expectations. Second, the mean of the prior $c^{(0)}$ appears on the final consensus, thus, beliefs are pulled towards this value.[14]

In cases with limited information, having a prior helps create a more efficient estimate of the realized state, especially if the prior is much more informative than a single signal. However, when

[14]Note that $c^{(0)}$ shows up in the final beliefs even if all agents had $\lambda_i = 1$. Under this condition, only under a joint release of all signals in the first round would wisdom prevail.

the total available information is abundant, and sufficient for the realized state to be fully revealed, the impact of the prior optimally washes away. Yet, in the current setting, since each agent places some weight on their prior, the prior affects the final consensus. Since for a non-atomic distribution of $\tilde{\theta}$, the probability that the realized value θ is equal to the mean of the prior is 0, the fact that the mean of the prior does not wash away pulls final beliefs away from the optimal guess.

Initial model without Conditioning Under the initial model and a rule that does not conditions on signal realization, the limiting beliefs are

$$\operatorname*{plim}_{n\to\infty} c^{(K)} = \left(1 - \prod_{j=1}^{K}\left(1 - \tilde{w}_j\bar{\lambda}\right)\right)\theta + \prod_{j=1}^{K}\left(1 - \tilde{w}_j\bar{\lambda}\right)c^{(0)} \neq \theta$$

In this case the weights on the signals continue to be distorted, yet, this distortion does not pull away from the optimal guess, since the expected signal value within each information round is θ, the unconditional mean of the signals. This comes as a result of the lack of conditioning on the signal realization when assigning an information release round. However, as was the case in the previous setting, the final beliefs do not converge to the realized value θ, as once more the mean of the prior $c^{(0)}$ influences the final beliefs.[15] Notice that this will be the case even if $\lambda_i(n) = 1\,\forall i$ and n.

Prior-less Model with Conditioning We now consider the setup in which agents who do not receive a signal in the first round have their beliefs equal to an empty set. After the first round of information arrival beliefs are updated according to the *GDG* model. In all remaining information rounds, since each agent will be informed, the belief dynamics become identical with that of the initial model. We start by considering the case in which the information release rule conditions on signal realizations. In this case, the limiting beliefs are

$$\operatorname*{plim}_{n\to\infty} c^{(K)} = \sum_{j=2}^{K}\left(\prod_{k=j+1}^{K}\left(1 - w_k\bar{\lambda}\right)\right)w_j\bar{\lambda}\mu_j + \prod_{j=2}^{K}\left(1 - w_j\bar{\lambda}\right)\mu_1 \neq \theta$$

[15] Only if the realized value θ was exactly equal to the mean of the distribution $c^{(0)}$ would beliefs have converged to θ. However, this is a probability 0 event.

What makes this case distinct from the previous ones is that the mean of the prior no longer appears on the final beliefs. The limiting belief is now a convex combination of the conditional means. However, the weights assigned with the conditional means are not w_k, which following the law of iterated expectations, would be the adequate weights. Since these weights are distorted, generically the limiting beliefs will differ from θ.

Thus, even in cases in which agents may have non-informative priors—learning about a new product, a new political candidate, etc.—in environments in which otherwise their beliefs would converge to the truth, if information is released sequentially their beliefs may nonetheless still converge away from the truth. What causes *wisdom* to fail in this case is the distorted weights that signals receive depending on the timing of their arrival. Since the weight signals receive depends on the timing of their release, if there is any correlation between the value of the signals, and the round at which they are released, the final beliefs converge away from the truth.

Prior-less model without Conditioning Finally, once again we consider a setup in which agents who do not receive a signal in the first round have their beliefs set to an empty set. Once again, after the first round of information arrival beliefs are updated according to the *GDG* model. In all remaining information rounds, since each agent will be informed, the belief dynamics become identical with that of the initial model. We now consider the case in which the information release rule does not conditions on signal realizations. In this case, the limiting beliefs are

$$\operatorname*{plim}_{n\to\infty} c^{(K)} = \sum_{j=2}^{K} \left(\prod_{k=j+1}^{K} \left(1 - \tilde{w}_k \bar{\lambda}\right) \right) \tilde{w}_j \bar{\lambda} \theta + \prod_{j=2}^{K} \left(1 - \tilde{w}_j \bar{\lambda}\right) \theta = \theta$$

Once more, the prior no longer appears on the final beliefs. Furthermore, since the information release rule does not condition on signal realization, each conditional mean is equal to the unconditional mean of the signals. Although the weights associated with the signals released in different rounds are distorted, the convex combinations are simply mixtures of the unconditional expectation θ. Consequently, final beliefs are equal to the realized value θ. Hence, in the setup in which agents have an uninformative prior, and there is no correlation between the information release round and the signal realizations, *wisdom* persists.

We summarize these results in the proposition below.

Proposition 7 (The Persistence and Failure of *Wisdom*)**.**

	Initial Model	Prior-less Model
Conditioning on s_i	*Wisdom* Fails	*Wisdom* Fails
Not conditioning on s_i	*Wisdom* Fails	*Wisdom* Persists

Table 1: The persistence and failure of *wisdom*

Thus, within this setup, when agents have an informative prior, by affecting the initial guesses of these agents, the prior distort the final beliefs. Consequently final beliefs do not converge to the realized value θ. In addition, under sequential information arrival, an adequate weighting of the conditional means, as prescribed by the law of iterated expectations, can not be maintained. The only way around this is for each one of the conditional means to be equal to the unconditional mean, which happens when the rule that determines the information release rounds does not depend on signal realization.

In sum, this set of findings reveal that once we consider an environment with sequential information arrival, the ability to adequately aggregate information is compromised, in all but very specific cases *wisdom* fails.

5 Conclusions

We extend the DeGroot model to allow for sequential information arrival. We show that the final consensus formed within the group is affected by the information arrival sequence, even when the information content is held constant. By identifying the optimal and pessimal information release sequence that yields the highest and lowest attainable final consensus, we bound the variation in the final consensus that can be attributed to the sequencing of information. Taking an ex-ante approach, we find that the expected range between the highest and lowest attainable final consensus is maximized when each group member has identical influence. We further analyze the robustness of the wisdom of crowds in a large society, where the number of agents grows. We find that, to a large extent, wisdom fails when information arrives sequentially: in all but particular cases, the beliefs of the group converge away from the truth, even as the number of agents grows arbitrarily large.

All things considered, this work emphasizes that not only the information content but the timing and order by which this information arrives play a role in determining the final beliefs formed via social learning. This would not be the case under optimal Bayesian learning. However, in our setting, Bayesian updating would require an immense amount of computational power. Consequently, we expect individuals to rely on heuristics for aggregating information. Indeed, experimental evidence from a companion paper, Reshidi (2020), reveals that the timing and order of information arrival affect the final beliefs formed by groups. Furthermore, individuals update their beliefs following heuristics akin to those studied in this paper.

6 Appendix

6.1 Proofs

Proof of Proposition 2.

For the final consensus $c^{(K)}$ to be independent of the information sequencing, generically, it must be that the weight each signal has on the final consensus is unchanged regardless of the round in which any signal is released. From Proposition 1 we know that after all information has been released, the final consensus that emerges is.

$$c^{(K)} = \sum_{k=1}^{K} \left(\prod_{z=k+1}^{K} \left(1 - \sum_{j \in \gamma(z)} \pi_j \lambda_j^{(z)} \right) \right) \sum_{i \in \gamma(k)} \pi_i \lambda_i^{(k)} s_i + \prod_{k=1}^{K} \left(1 - \sum_{j \in \gamma(k)} \pi_j \lambda_j^{(k)} \right) c^{(0)}$$

Where in the above expression we have replaced λ_i with $\lambda_i^{(k)}$, thus allowing for agents to place different weights on their signals based on when they receive the signals. Consider changing the release timing of a subset of signals G that would otherwise arrive in round g, with $\sum_{i \in G} \pi_i \lambda_i^{(g)} > 0$. As long as $\lambda_i^{(k)}$ is not equal to 0 for all i and k, finding such a group is always possible. Denote $\tilde{\lambda}_i = \lambda_i^{(g)}$. To ensure that a change of the release timing does not change the weight that these signals have on the final consensus, it must be that for all $i \in G$:

$$\lambda_i^{(k)} = \begin{cases} \frac{\tilde{\lambda}_i}{\left(1 - \sum_{i \in \gamma(g) \setminus G} \pi_i \lambda_i(g)\right) \prod_{l=1}^{q-1} \left(1 - \sum_{i \in \gamma(g-l)} \pi_i \lambda_i^{(g-l)}\right)} & if \quad k = g - q \\ \vdots & \vdots \\ \frac{\tilde{\lambda}_i}{\left(1 - \sum_{i \in \gamma(g) \setminus G} \pi_i \lambda_i(g)\right)} & if \quad k = g - 1 \\ \tilde{\lambda}_i & if \quad k = g \\ \left(1 - \sum_{i \in \gamma(g+1)} \pi_i \lambda_i^{(g+1)}\right) \tilde{\lambda}_i & if \quad k = g + 1 \\ \vdots & \vdots \\ \prod_{l=1}^{q} \left(1 - \sum_{i \in \gamma(g+l)} \pi_i \lambda_i^{(g+l)}\right) \tilde{\lambda}_i & if \quad k = g + q \end{cases}$$

This is the only specification of $\lambda_i^{(k)}$ for agents $i \in G$ which ensures that the weight these signals command on the final consensus does not change as we change their information release round.

Now consider changing the information release round of signals in G, from g to $g - q$. A similar argument would follow if we changed the release round to $g + q$. The above specification of $\lambda_i^{(k)}$ ensures that signals in G command the same weight as before. However, by changing the release round of signals in G, the discounting of all signals in rounds $\{1, 2, \ldots, g-1\}$ changes, while the weights $\lambda_i^{(k)}$ remain unchanged for all $i \notin G$ since their release round did not change. In particular, the weight of signals released in rounds $\{g-q, g-q+1, \ldots, g-1\}$ strictly increases. Concretely, for a signal r released in round z, where $g - q \le z < g$ the weight on the final consensus changes from

$$\prod_{\{j>z\}\backslash g}\left(1-\sum_{i\in\gamma(j)}\pi_i\lambda_i^{(j)}\right)\left(1-\sum_{i\in\gamma(g)}\pi_i\lambda_i^{(g)}\right)\pi_r\lambda_r^{(z)}$$

to

$$\prod_{\{j>z\}\backslash g}\left(1-\sum_{i\in\gamma(j)}\pi_i\lambda_i^{(j)}\right)\left(1-\sum_{i\in\gamma(g)\backslash G}\pi_i\lambda_i^{(g)}\right)\pi_r\lambda_r^{(z)}$$

Which is strictly larger since $\sum_{i\in G}\pi_i\lambda_i^{(g)}\geq 0$. Since the weight on at least a subset of signals changes, generically, the final consensus is altered. □

Proof of Proposition 3.

Lemma 1. *Given n signals, the optimal sequence can not have s_j released individually in information release round k and s_i released individually in information release round $k+1$ if $s_i < s_j$.*

Proof of Lemma 1.

Two Signal Case

Let π_i represent the i'th value of the left eigenvector of the network matrix M. Let $c^{(0)}$ represent the initial consensus, and let λ_i represent the weight that agent i places on her signal. Releasing only signal s_i in the first information release round leads to the new consensus $c^{(1)}=\pi_i\lambda_i s_i+(1-\pi_i\lambda_i)c^{(0)}$. Afterwards, releasing s_j in the second information release round leads to the following final consensus

$$c^{(2)}=\pi_j\lambda_j s_j+(1-\pi_j\lambda_j)\pi_i\lambda_i s_i+(1-\pi_j\lambda_j)(1-\pi_i\lambda_i)c^{(0)}$$

Denote by $\tilde{c}^{(2)}$ the alternative final consensus reached by swapping the release time of signal s_i and s_j. The difference between the final consensuses will be $c^{(2)}-\tilde{c}^{(2)}=\lambda_i\pi_i\lambda_j\pi_j(s_j-s_i)$, which is positive as long as $s_j>s_i$. Thus, given two signals and an initial consensus $c^{(0)}$ it is never optimal to release s_i after s_j if $s_i<s_j$.

General Case

Assume by contradiction that the optimal sequence has signal s_j released in information release round k and s_i released in $k+1$ while $s_i<s_j$. Notice that regardless of what the information release sequence from $k+2$ and onward is, the final consensus will be a weighted average of the consensus $c^{(k+1)}$ and the signals released at $k+2$ and onward. Holding fixed the sequence of information arrival before k and after $k+1$, the problem of maximizing $c^{(k+1)}$ by choosing when to release s_i and s_j reduces to the optimization problem with two signals. From the results in the **Two Signal Case**, we know that the value of $c^{(k+1)}$ can be increased by releasing s_i before s_j, thus violating the claim that the initial sequence was optimal. This concludes the proof of Lemma 1. □

Splitting Information Sets

Consider a set $J \subseteq \gamma(k)$ of agents for whom information arrives in information round k. The effect of releasing the signals of these agents one information period earlier without joining any other set of agents (or equivalently, the effect of releasing the signals for all agents in $\gamma(k) \setminus J$ and all agents in $\gamma(k')$ for all $k' > k$ one period later) would shift the final consensus from

$$c^{(K)} = \sum_{v=k+1}^{K} \prod_{z=v+1}^{K} \left(1 - \sum_{j \in \gamma(z)} \pi_j \lambda_j\right) \sum_{i \in \gamma(v)} \pi_i \lambda_i s_i + \left(\prod_{z=k+1}^{K} \left(1 - \sum_{j \in \gamma(z)} \pi_i \lambda_i\right)\right) \sum_{i \in \gamma(k)} \pi_i \lambda_i s_i$$
$$+ \left(\prod_{z=k+1}^{T} \left(1 - \sum_{j \in \gamma(z)} \pi_j \lambda_j\right)\right) \left(1 - \sum_{i \in \gamma(k)} \pi_i \lambda_i\right) c^{(k-1)}$$

To the new final consensus c'_T

$$\tilde{c}^{(K)} = \sum_{v=k+1}^{K} \prod_{z=v+1}^{K} \left(1 - \sum_{j \in \gamma(z)} \pi_j \lambda_j\right) \sum_{i \in \gamma(v)} \pi_i \lambda_i s_i + \left(\prod_{z=k+1}^{K} \left(1 - \sum_{j \in \gamma(z)} \pi_i \lambda_i\right)\right) \sum_{i \in \gamma(k) \setminus J} \pi_i \lambda_i s_i$$
$$+ \left(\prod_{z=k+1}^{K} \left(1 - \sum_{j \in \gamma(z)} \pi_i \lambda_i\right)\right) \left(1 - \sum_{i \in \gamma(k) \setminus J} \pi_i \lambda_i\right) \sum_{i \in J} \pi_i \lambda_i s_i$$
$$+ \left(\prod_{z=k+1}^{T} \left(1 - \sum_{j \in \gamma(z)} \pi_j \lambda_j\right)\right) \left(1 - \sum_{i \in \gamma(k) \setminus J} \pi_i \lambda_i\right) \left(1 - \sum_{i \in J} \pi_i \lambda_i\right) c^{(k-1)}$$

First, note that the weights on all other signals released at or after k remain unaffected. Second, the weight that signals of agents in set J have on the finl consensus is lower under $\tilde{c}^{(K)}$. Furthermore, since $\sum_{i=1}^{N} \pi_i = 1$ and $\lambda_i \leq 1 \; \forall \; i$, it is clear that $0 < 1 - \sum_{i \in \gamma(k) \setminus J} \pi_i \lambda_i < 1$, as long as all signals are not released in information round k and the set J does not consist of all the agents for whom information initially arrived in round k. Hence, this shift decreases the weight that signals of agents in set J have on the final consensus. To see that this swap increases the weight on $c^{(k-1)}$ note that:

$$\left(1 - \sum_{i \in \gamma(k) \setminus J} \pi_i \lambda_i\right) \left(1 - \sum_{i \in J} \pi_i \lambda_i\right) = 1 - \sum_{i \in \gamma(k)} \pi_i \lambda_i + \left(\sum_{i \in \gamma(k) \setminus J} \pi_i \lambda_i\right) \left(\sum_{i \in J} \pi_i \lambda_i\right)$$

As long as J is not an empty set or does not contain all elements in $\gamma(k)$:

$$\left(1 - \sum_{i \in \gamma(k)} \pi_i \lambda_i\right) < \left(1 - \sum_{i \in \gamma(k) \setminus J} \pi_i \lambda_i\right) \left(1 - \sum_{i \in J} \pi_i \lambda_i\right)$$

Furthermore, note that the increased weight of consensus $c^{(k-1)}$ is exactly equal to the decreased weight on signals in J. Then, the total impact of such a shift is:

- The weights on the final consensus of signals released in information round k that are not shifted, as well as all signals released after information round k, are unaffected.
- The weights on the final consensus of signals in set J is decreases.
- The weights on the final consensus of $c^{(k-1)}$, or equivalently, the weights on the final consensus of all signals released before information round k including the initial consensus $c^{(0)}$ increases.

Merging Information Sets

Consider the set $\gamma(k)$ of agents for whom information arrives in information round k. The effect of releasing the signals of these agents one information release round earlier by joining a group of agents for whom information arrives one information round earlier, would shift the final consensus from

$$\begin{aligned} c^{(K)} = &\sum_{v=k+1}^{K} \prod_{z=v+1}^{K} \left(1 - \sum_{j\in\gamma(z)} \pi_j\lambda_j\right) \sum_{i\in\gamma(v)} \pi_i\lambda_i s_i + \left(\prod_{z=k+1}^{K} \left(1 - \sum_{j\in\gamma(z)} \pi_i\lambda_i\right)\right) \sum_{i\in\gamma(k)} \pi_i\lambda_i s_i \\ &+ \left(\prod_{z=k+1}^{T} \left(1 - \sum_{j\in\gamma(z)} \pi_j\lambda_j\right)\right) \left(1 - \sum_{i\in\gamma(k)} \pi_i\lambda_i\right) \sum_{j\in\gamma(k-1)} \pi_j\lambda_j s_j \\ &+ \left(\prod_{z=k+1}^{T} \left(1 - \sum_{j\in\gamma(z)} \pi_j\lambda_j\right)\right) \left(1 - \sum_{i\in\gamma(k)} \pi_i\lambda_i\right) \left(1 - \sum_{j\in\gamma(k-1)} \pi_j\lambda_j\right) c^{(k-2)} \end{aligned}$$

To the new final consensus $\tilde{c}^{(K)}$:

$$\begin{aligned} \tilde{c}^{(K)} = &\sum_{v=k+1}^{K} \prod_{z=v+1}^{K} \left(1 - \sum_{j\in\gamma(z)} \pi_j\lambda_j\right) \sum_{i\in\gamma(v)} \pi_i\lambda_i s_i \\ &+ \left(\prod_{z=k+1}^{T} \left(1 - \sum_{j\in\gamma(z)} \pi_j\lambda_j\right)\right) \sum_{j\in\gamma(k-1)\cup\gamma(k)} \pi_j\lambda_j s_j \\ &+ \left(\prod_{z=k+1}^{T} \left(1 - \sum_{j\in\gamma(z)} \pi_j\lambda_j\right)\right) \left(1 - \sum_{j\in\gamma(k-1)\cup\gamma(k)} \pi_j\lambda_j\right) c^{(k-2)} \end{aligned}$$

This shift does not affect the weight on the signals in set $\gamma(k)$, it increase the weight on the signals in set $\gamma(k-1)$ and decreases the weight on the consensus c_{t-2}. To see the later, note that

$$\left(1 - \sum_{i\in\gamma(k)} \pi_i\lambda_i\right) \left(1 - \sum_{i\in\gamma(k-1)} \pi_i\lambda_i\right) = 1 - \sum_{i\in\gamma(k)\cup\gamma(k-1)} \pi_i\lambda_i + \left(\sum_{i\in\gamma(k)} \pi_i\lambda_i\right) \left(\sum_{i\in\gamma(k-1)} \pi_i\lambda_i\right)$$

Hence, as long as $\gamma(k)$ and $\gamma(t-1)$ are not empty:

$$\left(1 - \sum_{i \in \gamma(k) \cup \gamma(k-1)} \pi_i \lambda_i\right) < \left(1 - \sum_{i \in \gamma(k)} \pi_i \lambda_i\right)\left(1 - \sum_{i \in \gamma(k-1)} \pi_i \lambda_i\right)$$

The weights on all other signals released after t remain unaffected. Furthermore, note that the increased weight of signals in $\gamma(k-1)$is exactly equal to the decreased weight the consensus $c^{(k-2)}$. Then, the total impact of such a shift is:

- The weights on the final consensus of signals released after information round k are unaffected.
- The weights on the final consensus of signals originally released in information round k are unaffected.
- The weights on the final consensus of signals originally released in information round $k-1$ increases.
- The weight on the consensus $c^{(k-2)}$, or equivalently, the weight on all signals released before information round $k-1$, including here the initial consensus $c^{(0)}$ decreases.

Knowing the effect that **Splitting Information Sets** and **Merging Information Sets** has, allows us to know the effect that any relevant reshuffling of signals has. For example, if a set J leaves a group of agents for whom signals are released in information release round k, and are merged with a group of agents for whom signals are released at information round $k-1$, the effect will be the joint effect of **Splitting Information Sets** and **Merging Information Sets**. If the timing of information arrival changes from k to $k+1$ for a set of J agents, with no other agents receiving signals in round k, the total effect will simply be the opposite of **Merging Information Sets**, and so on.

Possible Joint Releases

Assume that under the optimal information release sequence multiple signals are released in information round $\hat{k}$.

Then, it must be that all of the signals released in information round $\hat{k}$ are greater than the consensus $c^{(\hat{k}-1)}$. Otherwise, if there was a signal $s_i < c^{(\hat{k}-1)}$ with $i \in \gamma(\hat{k})$, releasing it earlier would increase the value of the final consensus $c^{(K)}$. To see why this would be the case, note that if one of these signals were to be released earlier, via the **Splitting Information Sets** effect, the weight it would lose would be exactly equal to the weight that the consensus at $c^{(\hat{k}-1)}$ would gain.

Assume that a single signal, or a group of signals, are released after the joint release in $\hat{k}$. Since all signals released in $\hat{k}$ must be greater than $c^{(\hat{k}-1)}$, the signals released at $\hat{k}+1$ can be merged with the signals released at $\hat{t}$, thus having a **Merging Information Sets** effect, boosting the weight of the signals at $\hat{k}$ and decreasing the weight of the consensus at $c^{(\hat{k}-1)}$ by exactly the same amount. This would lead to an increase of the final consensus, thus contradicting the optimality of the information release sequence.

Hence, in the optimal information release sequence a group release of signals can only occur in the final information release round. Furthermore, all signals released in this round must be greater than the previous consensus.

Constructing the Optimal Information Release Sequence

From **Possible Joint Releases** we know that the optimal sequence either has no joint release of signals, or if it does, it can have at most one joint release which must occur at the last information release round K. Thus, $\gamma(k)$ for all $k \in \{1, \ldots, K-1\}$ must be singletons. From Lemma 1 we know that whenever there are singleton releases, the signal s_i released in information round k must be larger than the signal s_j released in $k-1$. This implies that for any $i \in \{1, \ldots, K-1\}$ and any $j \in \{1, \ldots, K-1\}$ if $s_i < s_j$ then s_i is released earlier than s_j.

Without loss of generality assume that $s_i < s_j$ if $i < j$. To construct the optimal information release sequence we can initially associate each signal s_i with in information round i. Afterwards we can see if there is a benefit of having s_K jointly released with s_{K-1}. This merging will be optimal only if $s_{K-1} > c^{(K-2)}$. If this is not the case, it must be that the optimal sequence is to release all signals one after the other, from lowest to highest. If however, having s_K be released the same time with s_{K-1} is optimal, we can continue this process by evaluating if having s_K and s_{K-1} released jointly with s_{K-2} improves the final consensus. This will be the case if $s_{K-2} > c^{(K-3)}$. Continuing in this fashion, possibly joining the group release in the last information arrival round with the last signal released individually until the last singleton $s_k < c_{k-1}$ constructs the optimal information release sequence. □

Proof of Corollary 1.

Corollary 1 follows from the proof of Proposition 3 and symmetry. □

Proof of proposition Proposition 4.

Once more, let the vector representing beliefs immediately after information arrival be denoted by $\tilde{b}^{(k)}$, while the vector representing beliefs after communication takes place be denoted by $\hat{b}^{(k)}$. Let $\gamma(k)$ be a predetermined set of nodes who's signals will be released in round k. That is $\gamma : \mathbb{N} \to \mathbb{N}$ represents a mapping from natural numbers, representing rounds, to the set of agents. When there were enough rounds of communication for convergence to take place, the latter beliefs were equal to $\hat{b}^{(k)} = M^{\infty}\tilde{b}^{(k)}$. We now replace M^{∞} with M^r representing r rounds of communication between information release. Define the beginning of the period beliefs $\tilde{c}_t$ as follows:

$$\tilde{b}_i^{(k)} = \begin{cases} (1-\lambda_i)\,\hat{b}_i^{(k-1)} + \lambda_i s_i & if \quad i \in \gamma(t) \\ \hat{b}_i^{(k-1)} & if \quad i \notin \gamma(t) \end{cases}$$

Where the difference from (2) is that we no longer assume a consensus has been reached. Note that $\tilde{b}_i^{(k)}$ is continuous in all it of it's components. Specifically, a small change in $\hat{b}_i^{(k-1)}$ leads to a small change in $\tilde{b}_i^{(k)}$. We can express the end of communication beliefs of agent i as

$$\hat{b}_i^{(k)} = \sum_j [M^r]_{ij}\, \tilde{b}_j^{(k)}$$

Where $[M^r]_{ij}$ represents the entry in row i and column j of matrix M^r. Note that $\hat{b}_i^{(k)}$ as a function of these entries, is continuous in each $[M^r]_{ij}$ coefficient, as well as in $\tilde{b}_j^{(k)}$. Consider the diagonal decomposition of matrix M

$$M^r = \Pi^{-1}\tilde{\Lambda}^r\Pi$$

Where $\tilde{\Lambda}$ is a diagonal matrix, with entry (i, j) equal to the i'th eigenvalue. While Π represents the matrix of left hand eigenvectors of M [16]. It then follows that

$$[M^r]_{ij} = \pi_j + \sum_{k\geq 2} \lambda_k^r \pi_{ik}^{-1}\pi_{kj}$$

Since we can think of M as a transition matrix of an ergodic Markov chain we have $\lambda_1 = 1 \geq \lambda_2 \cdots \geq \lambda_n$. Hence, the difference between $[M^r]_{ij}$ and π_j goes to 0 exponentially in r. Thus, for any given $\tilde{\varepsilon} > 0$, we can find a $r_{\tilde{\varepsilon}}$ large enough such that $\left| [M^{r_{\tilde{\varepsilon}}}]_{ij} - \pi_j \right| < \varepsilon$ for all (i, j). Consequently, given some release sequence γ, a vector of weights agents place on their private signals λ, a vector of initial beliefs $\tilde{b}^{(0)}$, and a network matrix M, for any given $\varepsilon > 0$, we can find a $\tilde{r}$ large enough such that $\left| c_T(\gamma, r = \infty) - c_T(\gamma, r \geq \tilde{r}) \right| < \varepsilon$. This follows from the continuity of $\tilde{b}^{(k)}$ and $\hat{b}^{(k)}$, as well as from the fact that M^r converges to a constant stochastic matrix as r increases. Hence, given an optimal release sequence γ^* that maximized $c^{(K)}$ under $r = \infty$ we had

$$c^{(K)}(\gamma^*, r = \infty) > c^{(K)}(\gamma', r = \infty) \quad \forall\, \gamma' \neq \gamma^*$$

Then, for $\tilde{r}$ large enough

$$c^{(K)}(\gamma^*, r \geq \tilde{r}) > c^{(K)}(\gamma', r \geq \tilde{r}) \quad \forall\, \gamma' \neq \gamma^*$$

Thus, although we identify the optimal release sequence for $r = \infty$, this release sequence will be optimal for all cases in which $r \geq \tilde{r}$, where the exact value of $\tilde{r}$ depends on the network structure. □

Proof of proposition Proposition 5.

Associated with each agent i in matrix n is a signal $s_i(n)$. For the purposes of this proof we can assume that $s(n)$ is a sequence of signals bounded within $[0, 1]$. Let γ be some rule mapping $\{1, \ldots, K\}$ information release rounds to the $\{1, \ldots, n\}$ agents. It could be that agents are distributed randomly in the K information rounds, it could be that which round they are associated with depends on their signal realization, or any other mapping. In particular $\gamma(n)[k]$ represents the

[16] See Jackson (2010) for further discussion and intuition behind the link between the speed of convergence and eigenvalues.

group of agents for whom information arrives in information round k in network n. Let ψ represent the mapping from agents to information release rounds according to the aforementioned sequence. Thus, $\psi(n)[i]$ maps agents $i \in \{1, \dots, n\}$ in network n, to the $k \in \{1, 2, \dots, K\}$ release rounds. Define

$$\overline{s}(n)[k] := \min_{i \in \gamma(n)[k]} s_i(n) \qquad \overline{s}(n)[k] := \max_{i \in \gamma(n)[k]} s_i(n) \qquad r(n)[k] := \{\underline{s}(n)[k], \overline{s}(n)[k]\}$$

$\underline{s}(n)[k]$ and $\overline{s}(n)[k]$ represent the lowest and highest signal released in information round k for network n respectively. While $r(n)[k]$ represents the range of signals released in information round k in network n. Further define

$$\check{W}(n)[k, z] := \sum_{i \in \gamma(n)[k] : s_i(n) \leq z} \pi_i(n)\lambda_i(n) \qquad \hat{W}(n)[j, z] := \sum_{i \in \gamma(n)[j] : s_i(n) > z} \pi_i(n)\lambda_i(n)$$

$\check{W}(n)[k, z]$ and $\hat{W}(n)[j, z]$ represents the impact that signals released on information round k or j in network n, with values lower than z and higher than z respectively, have. Where the impact of a signal refers to the influence of the agent receiving it $\pi_i(n)$ multiplied by the value this agent assigns to the signal $\lambda_i(n)$.

$$\bar{W}(n)[k, j] := \max_{z(n)[k,j] \in [\underline{s}(n)[k], \bar{s}(n)[j]]} \left\{ \min \left\{ \hat{W}(n)[j, z] \, , \, \check{W}(n)[k, z] \right\} \right\} \qquad \forall j < k, \;\; k \in \{1, \dots, K\}$$

In network n, $\bar{W}(n)[k, j]$ is equal to the impact that signals released in round k have, with signal values lower than $z^*(n)[k, j]$, or the impact that signals released in round j have, with signal values larger than $z^*(n)[k, j]$, where $z^*(n)[k, j]$ is the argument that maximizes this value. Note that $z(n)[k, j] \in [\underline{s}(n)[k], \bar{s}(n)[j]]$, thus, we are interested in the value of signals that are lower than the highest signal released in round j, but larger than the lowest signal released in round k.

$$\bar{W}(n) := \max_{j,k} \bar{W}(n)[k, j] \qquad \forall j < k, \;\; k \in \{1, \dots, K\}$$

That is, $\bar{W}(n)$ chooses among all pairings k and j, and chooses the ones for whom the impact of the overlapping signals is the highest. Associated with $\bar{W}(n)$ are $j^*(n)$ and $k^*(n)$, which are the arguments the maximize the function above, as well as $z^*(n)$ which is the argument that had maximized $\bar{W}(n)[k^*(n), j^*(n)]$. Note that if $\bar{W}(n) = 0$ then released signals are all sorted, such that the lowest signal released in round t must me larger than the highest signal of nodes released in time $t - 1$. Define

$$\check{K}(n) := \{i \in \gamma(n)[k^*(n)] : s_i(n) \leq z^*(n)\} \qquad \hat{K}(n) := \{i \in \gamma(n)[j^*(n)] : s_i(n) > z^*(n)\}$$

For network n, $\hat{K}(n)$ and $\check{K}(n)$ represent the set of agents whose signals are released in round $k^*(n)$ or $j^*(n)$ respectively, with values smaller than or lower than $z^*(n)$ respectively. Note that since $\max_{i \in \{1,\dots,n\}} \pi_i(n) \to 0$ as $n \to \infty$, if $\#\hat{K}(n)$ and $\#\check{K}(n)$ do not increase $\bar{W}(n) \to 0$. In what follows without loss of generality assume that $\hat{W}(n)[j^*(n), z^*(n)] \leq \check{W}(n)[k^*(n), z^*(n)]$.

$$\tilde{K}(n) \coloneqq$$
$$\left\{ i : i \in \arg\min_{\{i\} \in \check{K}(n)} \left\{ \begin{array}{cc} \hat{W}(n)[j^*(n), z^*(n)] - \sum_i \pi_i(n)\lambda_i(n) & if \quad \sum_i \pi_i(n)\lambda_i(n) \leq \hat{W}(n)[j^*(n), z^*(n)] \\ 2\max_i \pi_i(n) & otherwise \end{array} \right. \right\}$$

Hence, $\tilde{K}(n)$ is the subset of elements in $\check{K}(n)$ whose influence most closely approximates the value $\check{W}(n)[j^*(n), z^*(n)]$ from below. To see this, note that

$$\triangle(n) \coloneqq \hat{W}(n)[j^*(n), z^*(n)] - \sum_{i \in \tilde{K}(n)} \pi_i(n)\lambda_i(n) \leq \max_i \pi_i(n)$$

If this was not the case, then $\tilde{K}(n)$ would include at least one more element from $\check{K}(n)$. Now, consider the consequence of swapping the release time of elements in $\tilde{K}(n)$ with those of elements in $\hat{K}(n)$. Denote the initial final consensus as $c^{(K)}(n)$, and let this swap lead to the new final consensus $\tilde{c}^{(K)}(n)$. Then we have

$$\begin{aligned}
&\tilde{c}^{(K)} - c^{(K)} = \\
&\quad - \triangle(n) \left(\sum_{k \in \gamma(n)[k^*(n)] : k \notin \tilde{K}(n)} \pi_k(n)\lambda_k(n) - \sum_{j \in \gamma(n)[j^*(n)] : j \notin \hat{K}(n)} \pi_j(n)\lambda_j(n) \right) \\
&\quad \times \left(\prod_{m \notin \{j^*(n), k^*(n)\}} \left(1 - \sum_m \pi_m(n)\lambda_m(n) \right) \right) c^{(0)} \\
&\quad - \triangle(n) \left(\sum_{k \in \gamma(n)[k^*(n)] : k \notin \tilde{K}(n)} \pi_k(n)\lambda_k(n) - \sum_{j \in \gamma(n)[j^*(n)] : j \notin \hat{K}(n)} \pi_j(n)\lambda_j(n) \right) \\
&\quad \times \left(\sum_{r < j^*(n)} \left(\prod_{m > r : m \neq \{j^*(n), k^*(n)\}} \left(1 - \sum_m \pi_m(n)\lambda_m(n) \right) \right) \sum_r \pi_r(n)\lambda_r(n)s_r(n) \right) \\
&\quad - \triangle(n) \left(\sum_{j^*(n) < r < k^*(n)} \left(\prod_{m > r : m \neq k^*(n)} \left(1 - \sum_m \pi_m(n)\lambda_m(n) \right) \right) \sum_r \pi_r(n)\lambda_r(n)s_r(n) \right) \\
&\quad - \triangle(n) \left(\prod_{m > j^*(n) : m \neq k^*(n)} \left(1 - \sum_m \pi_m(n)\lambda_m(n) \right) \right) \sum_{j \in \gamma(n)[j^*(n)]} \pi_j(n)\lambda_j(n)s_j(n) \\
&\quad + \left(\sum_{j \in \hat{K}(n)} \pi_j(n)\lambda_j(n)s_j(n) - \sum_{k \in \tilde{K}(n)} \pi_k(n)\lambda_k(n)s_k(n) \right) \left(\prod_{m > k^*(n)} \left(1 - \sum_m \pi_m(n)\lambda_m(n) \right) \right) \\
&\quad \left(1 - \left(\prod_{j^*(n) < m < k^*(n)} \left(1 - \sum_m \pi_m(n)\lambda_m(n) \right) \right) \left(1 - \sum_{v \in \left\{ \hat{K}(n) \cup \gamma(n)[k^*(n)] \setminus \tilde{K}(n) \right\}} \pi_v(n)\lambda_v(n) \right) \right)
\end{aligned}$$

All terms, besides the last, are multiplied by $\triangle(n)$. All parts multiplying $\triangle(n)$ are weighted convex combinations of the prior and the signals, which lie in a bounded set, thus, they themselves are bounded. Since $\triangle(n) \leq \max_i \pi_i(n)$, and since $\max_i \pi_i(n) \to 0$ as $n \to \infty$, all terms but the final term converge to 0 as $n \to \infty$. Focusing on the last term, since $\sum_{i=1}^{n} \pi_i(n) = 1$ and since $\forall\ i$ and $\forall\ n$ $\lambda_i(n) \leq 1$, it is straightforward to see that the last three multiplicative parts are always non-negative. While the following inequality holds for the first part of the last term

$$\sum_{j \in \hat{K}(n)} \pi_j(n)\lambda_j(n)s_j(n) > \sum_{j \in \hat{K}(n)} \pi_j(n)\lambda_j(n)z^*(n) \geq \sum_{j \in \tilde{K}(n)} \pi_j(n)\lambda_j(n)z^*(n) \geq \sum_{k \in \tilde{K}(n)} \pi_k(n)\lambda_k(n)s_k(n)$$

The first inequality follows from the definition of $\hat{K}(n)$, since each element in $\hat{K}(n)$ is larger than $z^*(n)$. The second inequality follows from the definition of $\tilde{K}(n)$, since it approximates the influence of signals in $\hat{K}(n)$ from below. The third inequality follows from the fact that each element in $\tilde{K}(n)$ is drawn from $\check{K}(n)$. By definition, each signal associated with each element in the later set has value lower than or equal to $z^*(n)$. Hence, under the assumption that in the initial sequence the intersection of the range of signals released in different periods is not empty, the last term is bounded away from 0, which leads to

$$\lim_{n \to \infty} \tilde{c}^{(K)}(n) - c^{(K)}(n) > 0$$

Thus, as $n \to \infty$, if there is overlap between signal values released in different rounds, the final consensus can be increased by shifting signals with lower values to earlier information release rounds, and shifting signals with higher values to later information release rounds. Since this improves the final consensus for any sequence, it can not be that the optimal information release sequence has overlapping signals. □

Proof of proposition Proposition 6.

From Proposition 3 it follows that for any realization of signals, and weights participants place on their signals, the optimal sequence is one of n possible sequences. These n sequences release information in a monotonic manner, starting with the lowest signal. The difference between the sequences is a round $K \in \{1, 2, \ldots, n\}$ after which the sequence releases all remaining signals jointly. Since λ_i are independently drawn, and since each one enters linearly in the final consensus, in expectation we can replace all λ_i with the mean of their distribution, denoted by $\bar{\lambda}$. Since the sequence being analyzed releases information monotonically, the expected value of signals released in the first round is simply the first order statistic of the distribution of signals $S_{(1)}$. This follows all the way to K, in which the last $n - K$ signals are released, which in expectation are equal to the $S_{(K)}, S_{(K+1)}, \ldots, S_{(n)}$ order statistics of the distribution of signals. Then, the expected final consensus for one of these sequences is

$$
\begin{aligned}
\mathsf{E}\left[c^{(K)}(\pi,\lambda,s)\right] =&\mathsf{E}\left[\left(1-\sum_{i\in\gamma(K)}\pi_i\lambda_i\right)\sum_{k=1}^{K-1}\left(\prod_{z=k+1}^{K-1}\left(1-\pi_z\lambda_z\right)\right)\pi_k\lambda_k s_k\right.\\
&\left.+\sum_{k\in\gamma(K)}\pi_k\lambda_k s_k+\left(1-\sum_{i\in\gamma(K)}\pi_i\lambda_i\right)\prod_{z=1}^{K-1}\left(1-\pi_z\lambda_z\right)c^{(0)}\right]\\
=&\mathsf{E}\left[\left(1-\sum_{i\in\gamma(K)}\tilde{\pi}_i\bar{\lambda}\right)\sum_{k=1}^{K-1}\left(\prod_{z=k+1}^{K-1}\left(1-\tilde{\pi}_z\bar{\lambda}\right)\right)\tilde{\pi}_k\right]\bar{\lambda}S_{(k)}\\
&+\sum_{k=K}^{n}\mathsf{E}\left[\tilde{\pi}_k\right]\bar{\lambda}S_{(k)}+\mathsf{E}\left[\left(1-\sum_{i\in\gamma(K)}\tilde{\pi}_i\lambda_i\right)\prod_{z=1}^{K-1}\left(1-\tilde{\pi}_z\lambda_z\right)\right]c^{(0)}
\end{aligned}
$$

Where $c^{(0)}$, the initial consensus, represents the ex-ante mean of the distribution of the signals. Although the influence values π_i are deterministic, ex-ante we do not know which signal they will be associated with. Thus, given a particular sequence, from an ex-ante point of view, after taking expectations, the only remaining uncertainty is with regard to the π_i values associated to each order statistic. Hence, we can continue the analysis as if $\lambda_i = \bar{\lambda}$ and the order statistics $S_{(i)}$ are fixed, while the π_i values vary. Let $\pi = \{\pi_1, \dots, \pi_n\}$ represent the vector of influence values that can be associated with the order statistics, and let $\tilde{\pi}_i$ represent the realized value associated with order statistic i. That is, $\tilde{\pi}_i$ associated with order statistic $i \in \{1, 2 \dots, n\}$, are drawn from π uniformly without replacement.

Consider the case in which $K = n$, and thus, one signal is released in each information round. In all other $n-1$ information release sequences, the analysis follows the same steps as the analysis below; the only difference is that multiple higher order statistics are released in the last round, making the effect further pronounced. Let $p(\pi_i)$ represent the probability that the realized value of $\tilde{\pi}_n = \pi_i$. First, note that

$$
p(\pi_i) = \frac{1}{n} \qquad \mathsf{E}[\tilde{\pi}_n] = \sum_{i=1}^{n} p(\pi_i)\pi_i = \sum_{i=1}^{n}\frac{1}{n}\pi_i = \frac{1}{n}
$$

We have $p(\pi_i) = \frac{1}{n}$ since each value is equally likely to be associated with the n'th order statistic. As can be seen, the expected influence value associated with the signal released in the last round is $\mathsf{E}\left[\tilde{\pi}_n\right] = \frac{1}{n}$. Since the weight the nth order statistic has on the final consensus is $\mathsf{E}[\tilde{\pi}_n]\bar{\lambda}$, whether $\pi = \frac{1}{n}\ \forall\ i$, or whether there is variation in the values π_i, in expectation, the weight signals released in the last round receive is not affected. On the other hand, the weight associated with the $n-1$ order statistic is

$$
\mathsf{E}\left[(1-\tilde{\pi}_n\bar{\lambda})\tilde{\pi}_{n-1}\right]\bar{\lambda} = \left(\mathsf{E}\left[\tilde{\pi}_{n-1}\right] - \mathsf{E}\left[\tilde{\pi}_n\tilde{\pi}_{n-1}\right]\bar{\lambda}\right)\bar{\lambda} = \left(\frac{1}{n} - \mathsf{E}\left[\tilde{\pi}_n\tilde{\pi}_{n-1}\right]\bar{\lambda}\right)\bar{\lambda}
$$

To calculate $\mathsf{E}\left[\tilde{\pi}_n\tilde{\pi}_{n-1}\right]$ we define[17]

$$p(\pi_i,\pi_j)=\begin{cases}\frac{1}{n(n-1)} & if \quad \pi_i\neq\pi_j\\ 0 & if \quad \pi_i=\pi_j\end{cases}$$

$$\begin{aligned}\mathsf{E}[\tilde{\pi}_n\tilde{\pi}_{n-1}] &= \sum_{i=1}^{n}\sum_{j\neq i}p(\pi_i,\pi_j)\pi_i\pi_j=\sum_{i=1}^{n}\sum_{j\neq i}\frac{1}{n(n-1)}\pi_i\pi_j=\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{n(n-1)}\pi_i\pi_j-\sum_{i=1}^{n}\frac{1}{n(n-1)}\pi_i^2\\ &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\pi_i\sum_{j=1}^{n}\pi_j-\frac{1}{n(n-1)}\sum_{i=1}^{n}\pi_i^2=\frac{1}{n(n-1)}-\frac{1}{n(n-1)}\sum_{i=1}^{n}\pi_i^2\end{aligned}$$

Since π_i^2 is a convex function, and the condition $\sum_{i=1}^{n}\pi_i=1$ has to be satisfied, $\sum_{i=1}^{n}\pi_i^2$ is minimized when $\pi_i=\frac{1}{n}\ \forall\ i$, and maximized when $\pi_i=1$ for some i and $\pi_j=0$ for all $j\neq i$. Hence, $\mathsf{E}[\tilde{\pi}_n\tilde{\pi}_{n-1}]\in\left[0,\frac{1}{n^2}\right]$, and so, the weight placed on the $n-1$ order statistic is within $\left[\frac{n-\bar{\lambda}}{n^2}\bar{\lambda},\frac{1}{n}\bar{\lambda}\right]$. The weight placed on the $n-2$ order statistic is

$$\mathsf{E}\left[(1-\tilde{\pi}_n\bar{\lambda})(1-\tilde{\pi}_{n-1}\bar{\lambda})\tilde{\pi}_{n-2}\right]\bar{\lambda}=\left(\mathsf{E}\left[\tilde{\pi}_{n-2}\right]-\mathsf{E}\left[\tilde{\pi}_n\tilde{\pi}_{n-2}\right]\bar{\lambda}-\mathsf{E}\left[\tilde{\pi}_{n-1}\tilde{\pi}_{n-2}\right]\bar{\lambda}+\mathsf{E}\left[\tilde{\pi}_n\tilde{\pi}_{n-1}\tilde{\pi}_{n-2}\right]\bar{\lambda}^2\right)\bar{\lambda}$$

To calculate $\mathsf{E}\left[\tilde{\pi}_n\tilde{\pi}_{n-1}\tilde{\pi}_{n-2}\right]$ we define

$$p(\pi_i,\pi_j,\pi_k)=\begin{cases}\frac{1}{n(n-1)(n-2)} & if \quad \pi_i\neq\pi_j\neq\pi_k\\ 0 & otherwise\end{cases}$$

$$\begin{aligned}\mathsf{E}[\tilde{\pi}_n\tilde{\pi}_{n-1}\tilde{\pi}_{n-2}] &= \sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq\{i,j\}}p(\pi_i,\pi_j,\pi_k)\pi_i=\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq\{i,j\}}\frac{1}{n(n-1)(n-2)}\pi_i\pi_j\pi_k\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{1}{n(n-1)(n-2)}\pi_i\pi_j\pi_k-\frac{1}{n(n-1)(n-2)}\left(3\sum_{i=1}^{n}\pi_i^2-2\sum_{i=1}^{n}\pi_i^3\right)\\ &= \frac{1}{n(n-1)(n-2)}-\frac{1}{n(n-1)(n-2)}\left(3\sum_{i=1}^{n}\pi_i^2-2\sum_{i=1}^{n}\pi_i^3\right)\end{aligned}$$

Once more, the highest value $\left(3\sum_{i=1}^{n}\pi_i^2-2\sum_{i=1}^{n}\pi_i^3\right)=1$ is achieved when $\pi_i=1$ for some i and $\pi_j=0\ \forall\ j\neq i$, while the lowest value $\left(3\sum_{i=1}^{n}\pi_i^2-2\sum_{i=1}^{n}\pi_i^3\right)=\frac{3n-2}{n^2}$ is achieved when $\pi_i=\frac{1}{n}$ $\forall i$. Thus, $\mathsf{E}[\tilde{\pi}_n\tilde{\pi}_{n-1}\tilde{\pi}_{n-2}]\in\left[0,\frac{1}{n^3}\right]$, and so, the weight placed on the $n-1$ order statistic is within $\left[\frac{(n-\bar{\lambda})^2}{n^3}\bar{\lambda},\frac{1}{n}\bar{\lambda}\right]$. It is straightforward to see that the additional weight gained by the $n-2$ order statistic is larger than that gained by the $n-1$ order statistic. Continuing in this fashion we see that the weight on the $n-k$ order statistic will be within $\left[\frac{(n-\bar{\lambda})^k}{n^{k+1}},\frac{1}{n}\bar{\lambda}\right]$. With the weight of all order

[17]With a slight abuse of notation, $\pi_i\neq\pi_j$ does not exclude the possibility that these values are equal. Rather, it excludes the possibility of choosing the same π_i for two different signals, which follows from the fact that we draw without replacement.

statistics lower than n minimized when $\pi_i = \frac{1}{n}\,\forall\, i$, and being boosted towards $\frac{1}{n}\bar{\lambda}$ as $\max_i \pi_i \to 1$. Furthermore, it is also straightforward to see that the the additional weight pulls towards a convex combination lower than the mean of the distribution. Naturally, the weight that each signal has on the final consensus, including the weight of the prior, sums up to 1. The weight that the prior has on the final consensus is $\mathsf{E}\left[\prod_{z=1}^{K}(1-\tilde{\pi}_z\lambda_z)\right]$, which is maximized when $\pi_i = \frac{1}{n}\,\forall\, n$ and is minimized when $\pi_i = 1$ with $\pi_j = 0\,\forall\, j \neq i$. The additional weight gained by all order statistics lower than n comes to the cost of the weight on the prior, which in expectation is equal to the mean of the distribution. Thus, this transfer shifts weight from the mean, towards a convex combination lower than the mean, and consequently decreases the final consensus. Since this weight transfer decreases the final consensus for each one of the n information release sequences, it must decrease the final consensus for the maximum of these sequences as well. Hence moving away from $\pi_i = \frac{1}{n}\,\forall\, i$ to any other vector π decreases the value of $\mathsf{E}\left[\overline{C}(\pi,\lambda,s)\right]$, with it's lowest value reached when $\max_i \pi_i \to 1$. When $\max_i \pi_i = 1$ the $\mathsf{E}\left[\overline{C}(\pi,\lambda,s)\right]$ is equal to the unconditional expectation of s. By symmetry $\mathsf{E}\left[\underline{C}(\pi,\lambda,s)\right]$ is minimized with a uniform π vector, and consequently, the gap is maximized. $\mathsf{E}\left[\underline{C}(\pi,\lambda,s)\right]$ also collapses towards the unconditional expectation value as $\max_i \pi_i \to 1$, and consequently the gap shrinks to 0. □

Proof of proposition Proposition 7.

Let ψ represent the mapping from agents to information release rounds according to the aforementioned rule. Thus, $\psi(n)[i]$ maps agents $i \in \{1,\ldots,n\}$ in network n, to the $k \in \{1,2,\ldots,K\}$ release rounds.

Initial Model with Signal Conditioning

When γ can not condition on signal realization, the rule amounts to choosing which share of signals is released in which information round. If all signals were released in the first period

$$c^{(K)} = \sum_{i=1}^{n}\pi_i(n)\lambda_i(n)s_i(n) + \left(1-\sum_{i=1}^{n}\pi_i(n)\lambda_i(n)\right)c^{(0)}$$

Note that $\operatorname{plim}_{n\to\infty}\sum_{i=1}^{n}\pi(n)\lambda_i(n) = \bar{\lambda}$. This follows from **Lemma 1** in Golub and Jackson (2010). We can define a new random variable, $\kappa_i(n) = \lambda_i(n)s_i(n)$. Since $\lambda_i(n)$ and $s_i(n)$ are independent $\mathbb{E}\left[\kappa_i(n)\right] = \bar{\lambda}\theta$. Furthermore

$$Var\left(\kappa_i(n)\right) = \mathbb{E}[\lambda_i(n)^2]\mathbb{E}[s_i(n)^2] - \bar{\lambda}\theta \leq 1$$

The inequality holds since both $\lambda_i(n)$ and $s_i(n)$ are bounded within 0 and 1. Once more, from the aformentioned lemma, it follows that $\operatorname{plim}_{n\to\infty}\sum_{i=1}^{n}\pi_i(n)\lambda_i(n)s_i(n) = \bar{\lambda}\theta$. Which leads to

$$\operatorname*{plim}_{n\to\infty} c^{(K)} = \bar{\lambda}\theta + (1-\bar{\lambda})c^{(0)}$$

Thus, if all signals are released jointly at the very first round, if $\lambda_i(n) = 1 \; \forall i, n$, then ***wisdom*** would prevail. That is, if agents fully listened to their signals, the prior would wash away, and since their signals are centered around θ, the consensus would eventually converge there. Let S_k be an element of a measurable partitions of the set of values that a signal can take, such that the intersection of each element is empty and the union of all elements is equal to the whole set. Define the following

$$w_k = \int \mathbb{1}\{x \in S_k\} f(x)dx \qquad \mu_k = \frac{\int x\mathbb{1}\{x \in S_k\} f(x)dx}{w_k} = \mathbb{E}[s|s \in S_k]$$

Further define

$$w_j(n) = \sum_{i\in\gamma(n)[j]} \pi_i(n) \qquad \tilde{\pi}_i(n) = \frac{\pi_i(n)}{w_{\psi(n)[i]}(n)}$$

If signals are released sequentially then the final consensus will be

$$\begin{aligned} c^{(K)} =& \sum_{j=1}^{K} \left(\prod_{k=j+1}^{K} \left(1 - \sum_{i\in\gamma(n)[k]} \pi_k(n)\lambda_k(n) \right) \right) \sum_{i\in\gamma(n)[j]} \pi_i(n)\lambda_i(n)s_i(n) \\ &+ \prod_{j=1}^{K} \left(1 - \sum_{i\in\gamma(n)[j]} \pi_i(n)\lambda_i(n) \right) c^{(0)} \end{aligned}$$

Replacing $\pi_i(n)$ with $\tilde{\pi}_i(n)$ and realizing that $w_{\psi(n)[i]}(n)$ is the same for all $i \in \gamma(n)[i]$, we can rewrite the above as

$$\begin{aligned} c_T^{(K)} =& \sum_{j=1}^{K} \left(\prod_{k=j+1}^{K} \left(1 - w_k(n) \sum_{i\in\gamma(n)[k]} \tilde{\pi}_k(n)\lambda_k(n) \right) \right) w_j(n) \sum_{i\in\gamma(n)[j]} \tilde{\pi}_i(n)\lambda_i(n)s_i(n) \\ &+ \prod_{j=1}^{K} \left(1 - w_j(n) \sum_{i\in\gamma(n)[j]} \tilde{\pi}_i(n)\lambda_i(n) \right) c^{(0)} \end{aligned}$$

First, note that $\text{plim}_{n\to\infty} w_k(n) = w_k$. To see this, notice that $\sum_i \pi_i \mathbb{1}_{i\in\gamma(i)}$ is unbiased for w_i. Then

$$\begin{aligned} \mathsf{P}\left[\Big| \sum_{i=1}^{n} \pi_i \mathbb{1}_{i\in\gamma(i)} - w_k \Big| > \varepsilon \right] &\leq \frac{Var\left(\sum_{i=1}^{n} \pi_i \mathbb{1}_{i\in\gamma(i)}\right)}{\varepsilon^2} = \frac{Var\left(\mathbb{1}_{i\in\gamma(n)[k]}\right) \sum_{i=1}^{n} \pi_i^2(n)}{\varepsilon^2} \\ &\leq \frac{Var\left(\mathbb{1}_{i\in\gamma(n)[k]}\right) \max_{i\leq n} \pi_i(n) \sum_{i=1}^{n} \pi_i(n)}{\varepsilon^2} \to 0 \end{aligned}$$

Where the last part follows from the fact that $Var\left(\mathbb{1}_{i\in\gamma(n)[k]}\right)$ is bounded, $\sum_{i=1}^{n} \pi_i = 1$ and the assumption that $\max_{i\leq n} \pi_i \to 0$ as $n \to \infty$. Then, having re-normalized $\tilde{\pi}_i(n)$, note that $\sum_{i\in\gamma(n)[k]} \tilde{\pi}_i(n) = 1$ while still $\max_{i\in\gamma(n)[k]} \tilde{\pi}_i(n) \to 0$ as $n \to \infty$. Once more, the aforementioned lemma implies that

$$\operatorname*{plim}_{n\to\infty} \sum_{i\in\gamma(n)[k]} \tilde{\pi}_i(n)\lambda_i(n) = \bar{\lambda} \qquad\qquad \operatorname*{plim}_{n\to\infty} \sum_{i\in\gamma(n)[k]} \tilde{\pi}_i(n)\lambda_i(n)s_i(n) = \bar{\lambda}\mu_k$$

Which leads to

$$\operatorname*{plim}_{n\to\infty} c^{(K)} = \sum_{j=1}^{K} \left(\prod_{k=j+1}^{K} \left(1 - w_k\bar{\lambda}\right) \right) w_j\bar{\lambda}\mu_j + \prod_{j=1}^{K} \left(1 - w_j\bar{\lambda}\right) c^{(0)}$$

Even under the assumption that $\lambda_i(n) = 1\ \forall\, i$, the final consensus is not equal to θ. However $\operatorname{plim}_{n\to\infty} c^{(K)}$ differs from θ for two reasons. First, even if $\lambda_i = 1\ \forall\, i$, $c^{(0)}$ crawls in and pulls away from the actual realized value θ. Second, notice that a convex combination of the conditional expectations μ_k with weights exactly equal to w_k is equal to the unconditional expected value of the signals $\mathsf{E}[s_i] = \sum_{i=1}^{K} w_k\mu_k = \theta$, which follows from the law of iterated expectations. However, in the expression above, these weights are distorted. In particular, the set of signals released earlier receives weight $\sum_{j=1}^{K} \left(\prod_{k=j+1}^{K} \left(1 - w_k\bar{\lambda}\right)\right) w_j\bar{\lambda} \le w_j$.

Initial Model without Signal Conditioning

Assume now that the mapping between information rounds and signals does not depend on signal realization. Then, the assignment rule reduces to choosing a probability w_k with which signal $s_i(n)$ is released in information round k. Naturally $\sum_{k=1}^{K} w_k = 1$. We maintain the previous definitions of $w_j(n)$ and $\tilde{\pi}_i(n)$. Following identical steps as above, it is straightforward to see that the probability limit of $\sum_{i\in\gamma(n)[k]} \tilde{\pi}_i(n)\lambda_i(n)$ and $s_i(n) = \bar{\lambda}\mu_k$, and $w_k(n)$ remain unchanged; with the only difference being that w_k now represents the probability of being released in information round k, which in the limit, corresponds to the share of signals released in round k. The main difference from the previous specification is the following probability limit

$$\operatorname*{plim}_{n\to\infty} \sum_{i\in\gamma(n)[k]} \tilde{\pi}_i(n)\lambda_i(n)s_i(n) = \bar{\lambda}\theta$$

Which can be shown following the same steps as before. Since the release rule can not condition on signal realization, $\mu_k = \theta\ \forall\, k$. Which leads to

$$\operatorname*{plim}_{n\to\infty} c^{(K)} = \left(1 - \prod_{j=1}^{K} \left(1 - w_j\bar{\lambda}\right) \right) \theta + \prod_{j=1}^{K} \left(1 - w_j\bar{\lambda}\right) c^{(0)}$$

In this case, the final beliefs do not converge to the realized value θ only because the mean of the prior $c^{(0)}$ has crept in.[18] Notice that this will be the case even if $\lambda_i(n) = 1\, \forall i$ and n.

[18] Only if the realized value θ was exactly equal to the mean of the distribution $c^{(0)}$ would beliefs have converged to θ. However, this is a probability 0 event.

Prior-less Model with Signal Conditioning

Adopting the GDG model implies that once the first round of signals is released, the prior is completely washed away. From **Theorem 2** in Banerjee et al. (2019) we know that for lattice graphs with shortcuts, *wisdom* prevails.[19] This implies that, the consensus reached after the first set of signals is released converges in probability to the mean of these signals, $\text{plim}_{n\to\infty} c^{(1)} = \mu_1$, as $n \to \infty$. After this initial release of signals there are no more uninformed agents, thus, the analysis follows that of the sequential DeGroot model, in which we simply replace $c^{(0)}$ with μ_1. Following identical steps as in the analysis before leads to

$$\underset{n\to\infty}{\text{plim}}\, c^{(K)} = \sum_{j=2}^{K} \left(\prod_{k=j+1}^{K} \left(1 - w_k \bar{\lambda}\right) \right) w_j \bar{\lambda} \mu_j + \prod_{j=2}^{K} \left(1 - w_j \bar{\lambda}\right) \mu_1$$

Which will generically not be equal to μ.[20] Hence, in this case, even tho the mean of the prior has been washed out, beliefs do not converge to θ as the conditional means of the signals are weighted with distorted weights.

Prior-less Model without Signal Conditioning

Following identical steps as in the **GDG Model with Signal Conditioning**, but realizing that when the information release rule can not condition on signal realization $\mu_k = \theta \ \forall k$ leads to

$$\underset{n\to\infty}{\text{plim}}\, c^{(K)} = \sum_{j=2}^{K} \left(\prod_{k=j+1}^{K} \left(1 - w_k \bar{\lambda}\right) \right) w_j \bar{\lambda} \theta + \prod_{j=2}^{K} \left(1 - w_j \bar{\lambda}\right) \theta = \theta$$

Hence, once more, the weights of the signals are distorted, however now wisdom holds within each information release round, and consequently the consensus will be equal to θ after each round. Thus, under this specification, *wisdom* prevails. □

References

Acemoglu, D., Dahleh, M. A., Lobel, I., and Ozdaglar, A. (2011). Bayesian learning in social networks. *The Review of Economic Studies*, 78(4):1201–1236.

Bala, V. and Goyal, S. (1998). Learning from neighbours. *The review of economic studies*, 65(3):595–621.

[19] Note that the cardinality of $\gamma(1)$ in expectation is equal to $w_1 n$. Hence, the cardinality of $\gamma(1)$ increases as $n \to \infty$. If $w_1 = 0$, we can simply re-define the second round as the first round, and the analysis follows through.

[20] To see this, notice that the attainable space of parameters for w_k and μ_k with the constraint that $\sum w_k = 1$ and $\sum w_k \mu_k = \mu$, lies in a subset of $[0,1]^{2K-2}$. Having $\text{plim}_{n\to\infty} c^{(K)}$ equal to μ adds a new constraint not implied by the previous constraints, reducing the dimensionality to $2K - 3$.

Banerjee, A., Breza, E., Chandrasekhar, A. G., and Mobius, M. (2019). Naive learning with uninformed agents. Technical report, National Bureau of Economic Research.

Banerjee, A. V. (1992). A simple model of herd behavior. *The quarterly journal of economics*, 107(3):797–817.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, 100(5):992–1026.

Bohren, J. A. (2016). Informational herding with model misspecification. *Journal of Economic Theory*, 163:222–247.

Brandts, J., Giritligil, A. E., and Weber, R. A. (2015). An experimental study of persuasion bias and social influence in networks. *European Economic Review*, 80:214–229.

Chandrasekhar, A. G., Larreguy, H., and Xandri, J. P. (2020). Testing models of social learning on networks: Evidence from two experiments. *Econometrica*, 88(1):1–32.

Choi, S., Gale, D., and Kariv, S. (2008). Sequential equilibrium in monotone games: A theory-based analysis of experimental data. *Journal of Economic Theory*, 143(1):302–330.

Choi, S., Gale, D., and Kariv, S. (2012). Social learning in networks: a quantal response equilibrium analysis of experimental data. *Review of Economic Design*, 16(2-3):135–157.

Corazzini, L., Pavesi, F., Petrovich, B., and Stanca, L. (2012). Influential listeners: An experiment on persuasion bias in social networks. *European Economic Review*, 56(6):1276–1288.

Dasaratha, K. and He, K. (2020). Network structure and naive sequential learning. *Theoretical Economics*, 15(2):415–444.

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

DeMarzo, P. M., Vayanos, D., and Zwiebel, J. (2003). Persuasion bias, social influence, and unidimensional opinions. *The Quarterly journal of economics*, 118(3):909–968.

Enke, B. and Zimmermann, F. (2017). Correlation neglect in belief formation. *The Review of Economic Studies*, 86(1):313–332.

Eyster, E. and Rabin, M. (2010). Naive herding in rich-information settings. *American economic journal: microeconomics*, 2(4):221–43.

Eyster, E., Rabin, M., and Weizsacker, G. (2015). An experiment on social mislearning.

Gale, D. and Kariv, S. (2003). Bayesian learning in social networks. *Games and economic behavior*, 45(2):329–346.

Golub, B. and Jackson, M. O. (2010). Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–49.

Grimm, V. and Mengel, F. (2020). Experiments on belief formation in networks. *Journal of the European Economic Association*, 18(1):49–82.

Hazla, J., Jadbabaie, A., Mossel, E., and Rahimian, M. A. (2019). Reasoning in bayesian opinion exchange networks is pspace-hard. In *Conference on Learning Theory*, pages 1614–1648.

Jackson, M. O. (2010). *Social and economic networks*. Princeton university press.

Kemeny, J. G. and Snell, J. L. (1960). Finite markov chains. d van nostad co. *Inc., Princeton, NJ*.

Kübler, D. and Weizsäcker, G. (2004). Limited depth of reasoning and failure of cascade formation in the laboratory. *The Review of Economic Studies*, 71(2):425–441.

Levy, G. and Razin, R. (2018). Information diffusion in networks with the bayesian peer influence heuristic. *Games and Economic Behavior*, 109:262–270.

Lobel, I. and Sadler, E. (2015). Information diffusion in networks through social learning. *Theoretical Economics*, 10(3):807–851.

Molavi, P., Tahbaz-Salehi, A., and Jadbabaie, A. (2018). A theory of non-bayesian social learning. *Econometrica*, 86(2):445–490.

Mossel, E., Mueller-Frank, M., Sly, A., and Tamuz, O. (2020). Social learning equilibria. *Econometrica*, 88(3):1235–1267.

Mossel, E., Sly, A., and Tamuz, O. (2015). Strategic learning and the topology of social networks. *Econometrica*, 83(5):1755–1794.

Mueller-Frank, M. (2013). A general framework for rational learning in social networks. *Theoretical Economics*, 8(1):1–40.

Mueller-Frank, M. and Neri, C. (2020). A general analysis of boundedly rational learning in social networks. *Available at SSRN 2933411*.

Reshidi, P. (2020). An experiment on social learning with information sequencing.

Rosenberg, D., Solan, E., and Vieille, N. (2009). Informational externalities and emergence of consensus. *Games and Economic Behavior*, 66(2):979–994.

Smith, L. and Sørensen, P. (2000). Pathological outcomes of observational learning. *Econometrica*, 68(2):371–398.