Information Sequencing with Naive Social Learning*

Pëllumb Reshidi †

December 2023

Abstract

In complex environments—where carrying out Bayesian updating is computationally unfeasible—the DeGroot model has emerged as a reliable and heavily utilized alternative. An assumption present in practically all versions of this model is that agents receive information simultaneously. We depart from this setup by allowing for the sequential arrival of information. We find final beliefs can be altered by varying only the sequencing of information, keeping the information content unchanged. Sequential information arrival often undermines the *wisdom of crowds*, leading to belief divergence from the truth even as group size increases without bounds. We identify the sequences that produce the highest and lowest attainable consensus levels, thereby bounding the range of belief variation due to information sequencing. Groups in which all members are equally influential turn out to be most susceptible to information sequencing.

*I thank Leeat Yariv, Roland Bénabou, Alessandro Lizzeri, Ben Golub, Pietro Ortoleva, Can Urgun, Xiaosheng Mu, Joao Thereze, Mu Zhang, Franz Ostrizek, Denis Shishkin, and Evgenii Safonov, and participants at many seminars and conferences for helpful comments and suggestions.

†Department of Economics, Duke University pellumb.reshidi@duke.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, which schools to send their children to, or which political candidates to support. In such learning environments, information often arrives over time. Consider a committee evaluating a candidate, where, in addition to common information—the candidate's CV, letters of recommendation, and so on—each member receives additional private information—e-mails from previous employers, conference interactions, and so forth. Some of the information might work in the candidate's favor, and 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 ordering of information? These questions are at the heart of this project.

Under Bayesian learning, the order in which agents receive information has no impact in and of itself. However, when agents in a network 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 with high computational and information demands.[1] As a consequence, to a large extent, the literature has focused on naive social learning, where agents use simpler, less demanding heuristics to update their beliefs. In particular, the DeGroot (1974) model has emerged as the canonical boundedly rational model and remains one of the most heavily utilized social-learning models.[2] In this paper, we study the impact of information ordering when agents learn via these simpler heuristics. However, taking the DeGroot model off the shelf and studying how information ordering affects beliefs is not possible due to an underlying assumption that the model has, namely, that in period one, agents start with a set of beliefs that evolve according to the DeGroot rule from that point onward. Interpreting the difference between these initial beliefs

[1] 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; in particular, they show that the problem is PSPACE-hard.

[2] 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. Several experimental papers find observed data often, in contrast to the Bayesian model, follows patterns predicted by the DeGroot model. The model has attractive long-run information-aggregation properties and has been recently axiomatized. See Section 1.2 for relevant papers.

as a result of different information, the above boils down to assuming that all information arrives before communication begins. In this paper, we depart from this setup and allow for information to arrive sequentially. In doing so, we aim to achieve two goals: to study how information sequencing affects beliefs when belief updating is carried out through feasible means, and to relax a prevailing assumption in the canonical DeGroot model and see which of its features carry through.

To study the effect information ordering has on beliefs formed within 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 DeGroot 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. In the model we analyze, each agent is associated with a single private signal, a time at which the signal is received, and a weight the agent places on the signal once it is received, which can be history-dependent. Information arrives at specific rounds, in which a subset of agents receive their private signals. 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 standard DeGroot 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 in 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. We find no weighting profiles exist that agents can use on their own signals that would offset this effect. Thus, regardless of how agents incorporate their own signals, be it in very mechanical or highly sophisticated ways, as long as learning takes place through a DeGroot fashion, information sequencing will affect the final beliefs.

Our next set of results relates to the group's ability to adequately aggregate information as the number of group members grows. The *wisdom of crowds*, 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. By contrast, once we allow for sequential information arrival, we find *wisdom* typically fails, and beliefs converge away from the truth even as the number of agents grows without bounds. We emphasize that *wisdom* typically fails if agents have an informative 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 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—learning about a new product, a new political candidate, and so on—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 signals have based on the time in which they were released. If any correlation exists between the realized signal values and the round at which they are released, the final beliefs converge away from the truth. However, in specific networks, *wisdom* persists if agents have an uninformative prior and no relation exists between realized signal values and their release time. This set of findings reveals 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.

In the rest of the paper, we assume agents place a fixed weight on their own signal, regardless of when they receive the signal. We consider this assumption the most faithful extension of the DeGroot model, the motivation behind which is the reliance on simple heuristics. Furthermore, Reshidi (2023) finds experimental evidence in support of such fixed weights. Under this assumption, 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 ordering of information release. Thus, within this framework, whatever the order of information turns out to be, the final beliefs must fall within the identified bounds. We also derive features of the maximal consensus sequence under limited information rounds for a large society.

Finally, 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 beliefs differs as the *influence* of each group member changes. Formally, *influence* corresponds to the eigenvector centrality of an agent, which is a commonly used measure in the social network literature. 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 the expected gap is maximized when influence is uniformly distributed across all group members. That is, groups in which each member has an equal voice are groups whose beliefs are most susceptible to manipulation through information ordering.

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, agents' strategic concerns are rather limited, making Bayesian learning a reasonable assumption. However, non-Bayesian models are still 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 in social-learning settings with repeated actions in papers such as Gale and Kariv (2003), Rosenberg et al. (2009), Mueller-Frank (2013), Mossel et al. (2015), Mossel et al. (2020). These papers may limit strategic interactions; for example, Gale and Kariv (2003) assume agents observe only the distribution of actions and that each agent is a small enough part of society for their actions not to affect aggregate outcomes and consequently greatly reduce strategic concerns. Alternatively, these papers study aspects such as whether agreement and proper information aggregation will occur at the asymptotic steady state. Although understanding the 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 it is rather extreme. In particular, they show the problem is PSPACE-hard.

The computational complexity of carrying out proper Bayesian updating in such settings has been 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), and Chandrasekhar et al. (2020), show people tend to make far-from-optimal choices in rather simple environments requiring only basic inference. To avoid computational complexity, which seems far beyond what humans can reasonably handle, and to gain additional tractability, a large body of literature studies 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), Golub and Jackson (2010), and Banerjee et al. (2021) fall within this category. Our primary difference from these papers is our focus on the ordering of information arrival and how it shapes final beliefs.

Among the non-Bayesian models, the DeGroot model, as first analyzed in DeGroot (1974), has emerged as the predominant alternative. A more behavioral interpretation of the model is covered in DeMarzo et al. (2003), whereas Golub and Jackson (2010) highlight desirable features of the model. Concretely, they find that despite its reliance on simple heuristics, under mild assumptions, the final consensus of beliefs updated in a DeGroot manner converges to the realized value of the parameter of interest. Furthermore, Molavi et al. (2018) develop axiomatic foundations of the DeGroot model. They find DeGroot learning is the unique learning rule to satisfy *imperfect recall*, *label neutrality*, *monotonicity*, and *seperability*.

Also of relevance is the work of Banerjee et al. (2021), who extend the DeGroot model to allow for uninformed agents. Although this work is not directly related to information sequencing, we use this framework in a particular case when analyzing the robustness of the *wisdom of crowds*.

In addition to the simplicity and tractability that the DeGroot model offers, experimental studies such as Chandrasekhar et al. (2020), Grimm and Mengel (2020), and Brandts et al. (2015) suggest the model predicts observed behavior often better than the Bayesian model.[3] In Reshidi (2023), we reach similar conclusions in a setting where information arrives sequentially. In the setting analyzed in that paper, with Bayesian agents, information sequencing is expected to generate no difference compared to a simultaneous release of information. However, we demonstrate information sequencing alone causes substantial and statistically significant differences. We reveal that not only does the sequencing of information affect final beliefs, but it does so in a manner well predicted by a sequential version of the DeGroot model.

2 Model

A finite set $N = \{1, 2, \ldots, n\}$ of agents interact through a social network. Each agent is represented by a node, whereas their interactions are captured by an $n \times n$ nonnegative matrix M. The matrix M is a stochastic matrix: the sum of each row is normalized to add up to 1. The matrix 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.

[3]In a field experiment, and in an additional lab experiment, Chandrasekhar et al. (2020) find the share of DeGroot agents to be 90% and 50%, respectively. Grimm and Mengel (2020) find that in explaining individual-level decisions, the DeGroot model outperforms the Bayesian model. Brandts et al. (2015) find that, in line with the DeGroot model, agents with more outgoing degrees have a greater influence on the final consensus.

Agent i in time t has belief $b_{i,t} \in \mathbb{R}$. Importantly, although in line with the existing literature, we refer to these values as beliefs, they do not describe agents' posterior distributions of a state of interest; rather, these beliefs can be thought of as agents' opinions, guesses, or best estimates of a state of interest. 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 H^t represent the history of information releases up to time t. In practice, H^t can be thought of as a vector of length n with the jth entry being equal to $\hat{t}_j$ if $\hat{t}_j < t$ and 0 otherwise. Let the weight an agent places on their signal $\lambda_i(H^t)$ be a mapping from the set of all possible histories up to time t to weights in $[0, 1]$. That is, how much agent i is influenced by her private signal may depend on the agent's identity, the time of her information arrival, and the history of information release. Let $\gamma(k)$ represent the set of agents for whom information arrives in time k, that is, $i \in \gamma(k)$ if $\hat{t}_i = k$. Signals arrive after communication has taken place in that round; hence, the vector of beliefs evolves as follows:[4]

$$b_t = (I - \Gamma_t) \circ M b_{t-1} + \Gamma_t \left((I - \Lambda_t) \circ M b_{t-1} + \Lambda_t s \right), \tag{1}$$

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_t is also a diagonal matrix with $\lambda_i(H^t)$ being the diagonal entry in row and column i. Thus, the beliefs of agent i are as follows

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

If agent i does not receive their private signal in round t, 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 direct attention to. When agent i receives a signal, beliefs are updated in the same manner, and afterward, the signal is incorporated with weight $\lambda_i(H^t)$. Note that in all rounds in which agents 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

[4] Final beliefs are identical to the case in which information is received at the beginning of the round with the alternative release time $\tilde{t}_i = \hat{t}_i + 1$, and alternative signal weights $\tilde{\lambda}_i(H^t) = \lambda_i(H^t)/m_{ii}$, leading to the following equation representing the evolution of beliefs

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

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 until a consensus emerges.

Similar to DeMarzo et al. (2003), who motivate the evolution of beliefs in the classic DeGroot model as a bounded rationality model, we can think of a bounded rationality interpretation for the current setting. We can interpret the current setting as one in which a set of agents attempt to estimate a parameter of interest $\theta \in \mathbb{R}$. Signals agents receive can be thought of as noisy measures of the parameter of interest, $s_i = \theta + \varepsilon_i$ with $\mathbb{E}[\varepsilon_i] = 0$. Although each signal is unbiased, each signal is noisy; thus, agents can improve their estimation by paying attention to others' estimations. 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 independently drawn signals are normally distributed, a Bayesian agent's estimation of θ will be a weighted average of the prior and the signals.[5] The weight a Bayesian agent places on a signal is proportional to the precision of their signal.[6] We can then think of the weight that agent i places on the opinion of agent j as proportional to the signal precision of agent j. Similarly, the weight the agent place on their 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 from the baseline DeGroot learning setup, in which agents continue to use the same weight on each other through all rounds of communication.[7]

3 Information Sequencing and Beliefs

3.1 Belief Convergence

We begin by analyzing how beliefs evolve within rounds in which no new information is released. 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 $b_{t+1} = Mb_t$. In line with the DeGroot model, the newly formed beliefs of each agent are weighted averages of the beliefs of the agents they pay direct attention to. Continuing in this fashion, as long as no new information is released, beliefs in round $t + 2$ can be expressed as $b_{t+2} = Mb_{t+1} = M(Mb_t) = M^2 b_t$. We can express beliefs after j rounds of no information release as $b_{t+j} = M^j b_t$. The matrix M can be thought of as a transition matrix of a Markov chain. It is well known that if

[5] This holds for any symmetric loss function.

[6] By precision, we refer to the inverse variance of the signals.

[7] For a more in-depth discussion of the bounded rational interpretation of the DeGroot model, refer to DeMarzo et al. (2003).

matrix M is strongly connected and aperiodic, then it is convergent.[8] This matrix has a unique eigenvalue equal to 1, whereas all other eigenvalues have modulus smaller than 1. Furthermore, a unique left-row eigenvector π exists satisfying $\pi M = \pi$, corresponding to eigenvalue 1, with $\sum_i \pi_i = 1$, such that for any initial vector of beliefs b_t,

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

Thus, with a strongly connected and aperiodic network matrix M, if enough rounds of communication take place between information-release rounds, beliefs 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, because it captures the weight the initial belief of agent i has on the consensus.

Note that if there are no rounds of communication between information release-rounds, the problem reduces to one in which all information is released simultaneously. Hence, for this setup to differ from the benchmark DeGroot model, communication must take place between information-release rounds. If these communication rounds are limited, the distinction between sequential and simultaneous information release is limited by the particular network structure.[9] To isolate the effect of information sequencing, throughout the analysis, we assume that between each round of information release, communication takes place until a consensus is formed. All results up to Section 5 go through even if we assume consensus does not emerge between information rounds; however, the assumption greatly helps with our ability to write final beliefs explicitly and keep proofs concise. Even in Section 5 where the assumption that a consensus emerges between information-release rounds greatly helps with tractability, we show that results go through with finite, albeit sufficiently large, rounds of communication.

We now consider where beliefs converge after each—and specifically after the final—round of information release. We relabel beliefs as $b_{i,t}^{(k)}$, where i represents an agent, k stands for the number of information-release round, and t represents the number of rounds since the last

[8]See Kemeny and Snell (1960). A matrix is strongly connected if a path exists from any node, to any other node. The period $d(k)$ of a state k of a matrix M is given by $d(k) = gcd\left\{m \geq 1 : M_{k,k}^m > 0\right\}$. Where $gcd\{\cdot\}$ stands for greatest common divisor. If $d(k) = 1$, 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 a positive integer n exists 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 $m_{ii} > 0$.

[9]If agents incorporate others' beliefs slowly or, equivalently, place a great amount of weight on their own previous period beliefs (high m_{ii}), then, even with several rounds of communication, agents' beliefs may barely differ from the starting beliefs.

information release. Furthermore, we denote initial beliefs after new information has just been released as $\tilde{b}_i^{(k)} = b_{i,1}^{(k)}$, and beliefs after communication has taken place as $\hat{b}_i^{(k)} = b_{i,\infty}^{(k)}$. 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 communication takes place in the k'th round of information release. From the above discussion, we also know the new consensus will be equal to $c^{(k)} = \sum_{i=1}^{N} \pi_i \tilde{b}_i^{(k)}$. Thus, beliefs immediately after new information is released will be

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

whereas beliefs after information is released and communication takes place will be

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

After each round of information release and after communication takes place, a new consensus arises. The new consensus is a convex combination of the previous consensus and the signals released in that information round. How much these signals affect the new consensus depends initially on the weights the agents who received them place on their own signals, namely, $\lambda_i(H^k)$. Naturally, if an agent ignores their own signal, the signal will fail to affect the beliefs of others. Furthermore, each signal's impact on the new consensus depends on the influence measure π_i of the agent receiving the signal. Regardless of how much weight an agent places on their own signal, if the rest of society places little to no weight on this agent, their signal will have little to no effect on the new consensus.

Proposition 1 (Belief Convergence)**.** *Let M be a strongly connected and aperiodic matrix, Let Λ_t be a vector of weights agents place on their private signal, and let communication between information-release rounds occur until convergence. After all signals are released, the beliefs of each agent converge to the consensus belief*

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

Recursive replacement of $c^{(k)}$ in equation (3) leads to the equation in Proposition 1. As can be seen, the final consensus will be a weighted average of the initial consensus $c^{(0)}$ and all the signals agents receive.[10] The influence of each signal on the final consensus, the value

[10]That is, the above is written in the following form $c^{(K)} = \sum_{i=1}^{n} \alpha_i s_i + \alpha_0 c^{(0)}$, with $\sum_{i=0}^{n} \alpha_i = 1$.

multiplying the signal, depends on the weight the agent who received it placed on the signal, the social influence of said agent—and the time when the signal was released. If a signal is released in information round k, the weight it has on the final consensus is multiplied by $\prod_{z=k+1}^{K}\left(1-\sum_{j\in\gamma(z)}\pi_j\lambda_j(H^z)\right)$. Since each $\left(1-\sum_{j\in\gamma(z)}\pi_j\lambda_j(H^z)\right)<1$, this multiplier serves as a discounter. Once a signal is released, its weight diminishes with each additional wave of new signals.

3.2 Sequencing Dependence

Proposition 1 reveals that the ordering of information affects a signal's influence on the final consensus. Thus, other things equal, we expect changing the ordering of information release to affect the final consensus that is eventually formed. However, there may be one way to eliminate the influence of information sequencing on the final consensus. Because the weight an agent places on their signal may be history-dependent, we next see whether there exist weights that can offset the impact of information sequencing, making the final consensus once more independent of the order of information arrival.

Definition 1 (Sequencing Independence)**.** *Beliefs in a network are sequencing independent if, for any vector of signals, the final consensus remains unchanged after changing any signal's order of arrival.*

Once more, let n represent the number of agents and K represent the number of information-release rounds.[11]

Proposition 2 (Sequencing Dependence)**.** *No $\lambda_i(H^k)$ weights exist that lead to sequencing-independent beliefs.*

Thus, even if we handpick the weight each agent placed on their signal after each observed history, we would not be able to find weights that make beliefs sequencing independent. Proposition 2 highlights an important feature that accompanies DeGroot learning when information no longer arrives simultaneously. If information arrives sequentially, regardless of how this information is incorporated—be it quite mechanically, using the same λ_i weight irrespective of when information arrives; or through optimal Bayesian weights, adequately assessing the existing informativeness in the network at the time when their signal arrives and weighting their signals accordingly—the particular sequencing of information will affect

[11] We assume there exists at least one information release history such that for at least two agents i and j, $\max_k \pi_i\lambda_i(H^k)>0$, and $\max_k \pi_j\lambda_j(H^k)>0$. By doing so, we simply exclude trivial cases in which beliefs are always equal to the initial consensus or cases in which only one agent places a non-zero weight on their signal—making sequential information arrival meaningless.

the final consensus. Hence, the ability to affect the final consensus by changing the order of information is not driven by how information is incorporated; rather, it is fundamentally related to DeGroot learning.

4 The Wisdom of Crowds under Sequential Information Arrival

4.1 Wisdom of Crowds Prerequisites

Despite its reliance on simple heuristics, the DeGroot model has appealing features beyond the gains in tractability. One such feature is the tendency of social learning in this environment to converge to the truth. Golub and Jackson (2010) find that in a society where learning takes place in a DeGroot fashion, beliefs converge to the correct state under relatively weak conditions. This feature is labeled the *wisdom* of crowds, an important feature as it suggests society as a whole does not need to act optimally for information to be correctly aggregated; rather, by relying on numerous sources of information, even with an imperfect information-aggregation process, the final beliefs converge to the truth. The objective of this section is to examine whether this result persists when information arrives sequentially. In Section 7.1 in the Appendix, we provide a brief overview of the formal concept of the wisdom of crowds.

In the sequential-information-arrival model studied thus far, the initial consensus serves a role comparable to that of a prior. As shown, the initial consensus plays a key role in determining agents' beliefs since it shows up even in the final consensus. To create an environment more aligned with Golub and Jackson (2010), we have to get rid of the initial consensus. However, modifying the model to get rid of the prior raises the issue of handling belief updating when facing agents whose beliefs are an empty set; these agents would be the ones who do not receive a signal in the first round. In Section 7.1 in the Appendix we introduce a modified version of the model presented in Section 2. This modified model can be interpreted as a setup without prior information. This modification brings the model into closer alignment with the aforementioned paper as well as other recent works in this topic.

4.2 Persistence and Failure of Wisdom

We now analyze different specifications and see whether *wisdom* persists when information arrives sequentially. With regard to the evolution of beliefs, we consider both the initial specification described in Section 2, which can be interpreted as one in which agents have an informative prior, and the prior-free specification described in Section 7.1 in the Appendix.

We consider two information-release rules: one that conditions and one that does not condition on signal realizations when assigning information-release rounds. Both scenarios have relevance in varying contexts. In certain situations, the timing of information, whether it's positive, negative, or somewhere in between, may not be tied to any specific temporal pattern. However, in other cases, there could be an association between the nature of the information and when it is disclosed: A significant breakthrough in a particular drug's development (a positive signal) may require a longer timeframe and may naturally occur after several unsuccessful attempts (negative signals); When news outlets have control over the release of information, they often resort to sensationalist headlines to attract more attention and viewership, prioritizing sharing positive (or negative) news before unveiling more neutral or mundane updates; In a corporate setting, certain departments, such as sales, might find it easier to report information quickly, typically with neutral updates as performance can be tracked year-round. In contrast, research and development teams may report less frequently and might disclose either very positive or highly concerning news regarding anticipated product improvements.

Let $(M(n))_{n=1}^{\infty}$ represent a sequence of n-by-n interaction matrices indexed by n, the number of agents in each network. Since 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. Let $\tilde{\theta}$, the state of interest, be a random variable drawn from a non-atomic distribution H normalized without loss of generality to be in $[0,1]$, with mean $\mathbb{E}[\tilde{\theta}] = c^{(0)}$. 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.[12] Let the weight an agent places on her signal received in information round k, $\lambda_i^{(k)}(n)$, be drawn from a distribution $G^{(k)}$ on $[0,1]$ with variance $\sigma_\lambda^{2(k)}$ and mean $\bar{\lambda}^{(k)}$.[13]

Consider a mapping γ that links information-release rounds $\{1,2,\ldots,K\}$ to agents $\{1,2,\ldots,n\}$ for whom information arrives in that round. In particular, $\gamma(n)[k]$ represents the group of agents for whom information arrives in information round k in network n. We

[12]Results go through if the state of interest, as well as the signals, lie in a multidimensional Euclidean space.

[13]In effect, $\lambda_i^{(k)(n)}$ weights do not need to be random. However, for ease of exposition, we assume these weights are drawn from some distribution. To accommodate the fact that these weights may be history-dependent, the distribution can be defined after the sequencing of information is specified and can be different for each information round and history of information release $G^{(k)} = G(H^k)$. Results carry through with deterministic $\lambda_i(H^k)(n)$ under the assumption that the mean (across agents) of weights utilized in information round k converges; otherwise, beliefs are not guaranteed to converge. With deterministic weights, $\bar{\lambda}^{(k)}$ can simply be thought of as the average weight utilized in information round k.

can assume this mapping does not depend on signal realizations, in which case, we assume the rule specifies probabilities $\hat{w}_1, \hat{w}_2, \ldots, \hat{w}_K$ such that $\sum_{k=1}^{K} \hat{w}_k = 1$, where $\hat{w}_k$ represents the probability that an agent will receive their signal in information round k. Alternatively, we consider a case in which the mapping depends on the realized signals. In this case, we assume $i(n) \in \gamma(n)[j]$ 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 partition 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. Let $w_k = \int \mathbb{1}\{x \in S_k\} dF$ represent the mass of signals released in round k, whereas $\mu_k = \mathbb{E}[s | s \in S_k]$ represents their conditional expectation.

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

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

The above expression differs from θ for two reasons. First, the weights associated with the conditional expectations are distorted away from the mass of signals released in that round.[14] Second, the mean of the prior $c^{(0)}$ appears on the final consensus; thus, beliefs are drawn toward this value.[15] 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 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, because each agent places some weight on their prior and afterward updates beliefs by averaging previous beliefs, the prior affects the final consensus. Because for a non-atomic distribution of $\tilde{\theta}$, the event in which the realized value θ is exactly equal to the mean of the prior has mass zero, the fact that the mean of the prior does not wash away pulls final beliefs away from the realized value θ.

[14]Recall from the law of iterated expectations that the proper way to *stitch* together conditional expectations so as to retrieve the unconditional expectation is $\theta = \sum_{i=1}^{k} w_k \mu_k$.

[15]Note that, when information is released sequentially, $c^{(0)}$ shows up on the final beliefs even if all agents had $\lambda_i^{(k)}(n) = 1$ weights $\forall k$. Only under a joint release of all signals in the first round and $\lambda_i^{(1)}(n) = 1$ weights would wisdom prevail. This would reduce the current setup to that analyzed in Jackson (2010).

Initial model without Conditioning Under the initial model and a rule that does not condition on realized signals, the limit beliefs are

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

In this case, the weights on the signals continue to be distorted, yet this distortion does not pull away from the optimal guess because the expected signal value within each information round is θ. This comes as a result of the lack of conditioning on the signal realizations when assigning information-release rounds. However, as was the case in the previous setting, the final beliefs do not converge to the realized value θ, because, once more, the mean of the prior $c^{(0)}$ influences the final beliefs.[16]

Prior-free Model with Conditioning We now consider the setup in which agents who do not receive a signal in the first round have their beliefs set to an empty set. After the first round of information arrival, beliefs are updated according to the *generalized DeGroot model*. In all remaining information rounds, because each agent will be informed, the belief dynamics become identical to 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

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

What makes this case distinct from the previous ones is that the mean of the prior no longer appears on the final beliefs. The final consensus is now a convex combination of conditional means. However, the weights associated with the conditional means are not w_k. Because the weight signals receive depends on the order of their release, if a correlation exists between the value of the signals and the round in which they are released, the final beliefs generically converge away from the truth.

Prior-free Model without Conditioning Finally, we consider the prior-free model and a rule that does not condition on realized signals. Once again, after the first round of information arrival, beliefs are updated according to the *generalized DeGroot model*. In all remaining information rounds, because each agent will be informed, the belief dynamics

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

become identical to that of the initial model. In this case, the limit beliefs are

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

Again, the prior does not appear on the final beliefs. Furthermore, because 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 combination is simply a mixture of the unconditional expectation θ. Consequently, final beliefs are equal to this value. Thus, in a setup in which agents have an uninformative prior, and no correlation exists between information-release rounds and signal realizations, *wisdom* persists.

We summarize the above results in the proposition below.

Proposition 3 (The Persistence and Failure of *Wisdom*)**.**

	Sequential DeGroot Model	Prior-free Sequential DeGroot Model
Signal Conditioning	*Wisdom* Fails	*Wisdom* Fails
No Signal Conditioning	*Wisdom* Fails	*Wisdom* Persists

When agents have an informative prior, by affecting their initial guesses, the prior distorts the final beliefs. Consequently, final beliefs do not converge to the realized value θ. The only way around this is to have an uninformative prior. In addition, under sequential information arrival, an adequate weighting of the conditional means, as prescribed by the law of iterated expectations, cannot be maintained. The only way around this is for each of the conditional means to be equal to the unconditional mean, which happens if the rule that determines the information-release rounds does not condition on realized signals.

In sum, this set of findings highlights that once we allow for sequential information arrival in the DeGroot setup, the ability to adequately aggregate information becomes challenging; for *wisdom* to go through, specific conditions must be met.[17]

[17] In the prior-free sequential DeGroot setup, for *wisdom* to persist, the network structure must be such that wisdom persists in the *generalized DeGroot model* as analyzed in Banerjee et al. (2021), which happens for a select set of networks, see Section 7.1 for more details.

5 Information Sequencing under Fixed Signal Weights

In the DeGroot model, despite the evolution of information, agents place fixed weights on their neighbors' beliefs. In the previous sections, we maintained this assumption while allowing for weights on agents' signals to be responsive to the history of information arrival. There is a bit of a contradiction here. After all, the prevalence of the DeGroot model relies on its simple heuristics. Proper usage of weights on agents' own signals that can be adjusted depending on the history of information arrival requires knowledge of the network structure, knowledge of others' signal precisions, knowledge of the timing of information arrival, and the ability to incorporate this information. Although we allowed for such weights to highlight the generality of the results, the most faithful extension of the DeGroot model would be one in which the weights on participants' own signals are also fixed.[18] Furthermore, Reshidi (2023) finds that participants in the lab, placed in groups, facing sequential information arrival while attempting to estimate a common parameter of interest, largely place fixed weights on their own signal regardless of the timing of the signal's arrival.[19] Because fixed weights on agents' own signals seem most in line with DeGroot updating, because empirical evidence supports such weights, and because this specification gives additional tractability, in this section, we continue the analysis under the assumption that agents place a fixed weight on the signal they receive regardless of when they receive this signal, namely, $\lambda_i(H^k) = \lambda_i$.

5.1 Consensus Maximizing and Minimizing Sequences

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 ordering of information. Thus, whatever the actual sequencing of information turns out to be, the final consensus will fall within our identified bounds.

Proposition 4 (Maximal Consensus Sequence)**.** *Without loss of generality, let $s_i < s_j$ if $i < j$. The final consensus is maximized under the following information sequencing:*

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

The information sequence that yields the highest attainable consensus releases information sequentially in a non-decreasing order, starting from the signal with the lowest realized

[18] By replacing $\lambda_i(H^k) = \lambda_i$, it is straightforward to see all previous results follow through with fixed weights.

[19] See section 5, in particular, Table 4, Table 5, and Figure 5 in Reshidi (2023).

value. Under this information release, the sequential release continues until the prevailing consensus, if ever, becomes lower than the lowest realized signal not yet released, in which case all signals are released jointly. To build some intuition for this result, note from Proposition 1 that the earlier a signal is released, the more it will be discounted from signals released after it. The only possible countervailing force is the weight agents place on their own signals. But when these weights are fixed, the connection between the order of information and the signal's influence on the final consensus becomes deterministic. Namely, if a signal is released early on, it will have a lower weight on the final consensus than the weight it would have if it were released in later information rounds. Then, if we aim to maximize the final consensus, we would lead with the low-valued signals since these signals will be more heavily discounted the earlier they are released.

Corollary 1 (Minimal Consensus Sequence)**.** *Without loss of generality, let $s_i < s_j$ if $i < j$. The final 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 realized value. Under this information release, the sequential release continues until the prevailing consensus, if ever, becomes higher than the highest realized signal not yet released, in which case, all signals are released jointly. Whereas Proposition 4 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 possible final consensuses that may arise as a result of changes in the order of information.[20]

These identified bounds are important regardless of whether the sequencing of information is handled by a decision-maker with a particular agenda or whether all agents receive their information at random times. That is, there can be two interpretations of the work in this section. First, by identifying the information release sequences that lead to the highest and lowest attainable final consensus, we analyze *how* a decision maker may release information in an attempt to manipulate the beliefs of a group. We can alternatively interpret this work in a framework with a complete absence of a decision-maker. The above results can simply help

[20] Proposition 4 and Corollary 1 also reveal that to bound the final consensus, we do not need to search through all n^n possible information-release sequences, but only $2n$ sequences. For the maximal consensus sequence, we need to consider only n possible releases, in which only the *threshold* beyond which signals are released jointly differs. The same *threshold* logic holds for the minimal sequence.

us identify the susceptibility of a society's beliefs due to the ordering of information. In other words, even if the ordering of information is completely random, it is informative to know the bounds within which a society's final consensus may fall. As long as information arrives sequentially, a gap will exist between the highest and lowest attainable final consensus—a gap identified by the above proposition and corollary.

Although in this section we maintained the assumption that after each round of information release, communication takes place until a new consensus is reached, this assumption is not necessary for the identified information-release sequences to remain maximal/minimal.

Proposition 5 (Communication Rounds)**.** *Let r represent the rounds of communication between information-release rounds. $\exists\, \tilde{r}$: if $r \geq \tilde{r}$, the identified sequences leading to the extreme consensuses remain unchanged.*

Thus, $\tilde{r}$ rounds of communication are sufficient for the results to hold. Although the exact $\tilde{r}$ value depends on the network structure, note the network matrix M^r converges exponentially to the matrix composed of the left eigenvectors. Therefore, for a large class of networks, $\tilde{r}$ will be fairly low.

5.2 A Large Society with Limited Information-Release rounds

The above analysis revealed that to achieve the highest and lowest attainable final consensus, within an information round often a single signal is released. This can potentially be accommodated in small networks because only a few rounds of information release are needed for the few agents who are present. However, in a large society with many agents, the number of agents likely 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 with limited rounds of information release.

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. Because 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:

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

As the network grows, the influence of the most influential agent goes to 0. This ensures 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 $\frac{1}{n}\sum_{i=1}^{n}\lambda_i(n)$ 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 6 (Maximal Consensus Sequence with Limited Rounds)**.** *Let $\bar{\gamma}$ represent the maximal consensus sequence yielding 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 maximal consensus 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 maximal structure identified with no limits on information rounds, namely, that signals are released in a non-decreasing fashion, prevails in a large society with limited information-release rounds. Naturally, due to the symmetry of the problem, a corollary can be derived for the minimal information-release sequence with limited rounds.

The proof of the above proposition relies on showing the final consensus may be increased by swapping the release time of a subset of signals with the release time of some other subset of signals, where signals in the former subset have an earlier release date and higher realized signal values compared to the later subset. As n increases, the added granularity makes it possible to select subsets such that this swap has no impact on the discounting of any other signals, leading to an unambiguous increase in the final consensus.

5.3 Influence and Sequence Susceptibility

We next analyze features of groups that affect their susceptibility to the sequencing of information arrival. In particular, we study how the distribution of the influence vector—which is a direct result of the network structure—affects 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 signals s_i be *i.i.d.* random variables drawn from a distribution F, and let λ_i be an *i.i.d.* random variable with distribution H on $[0,1]$.[21]

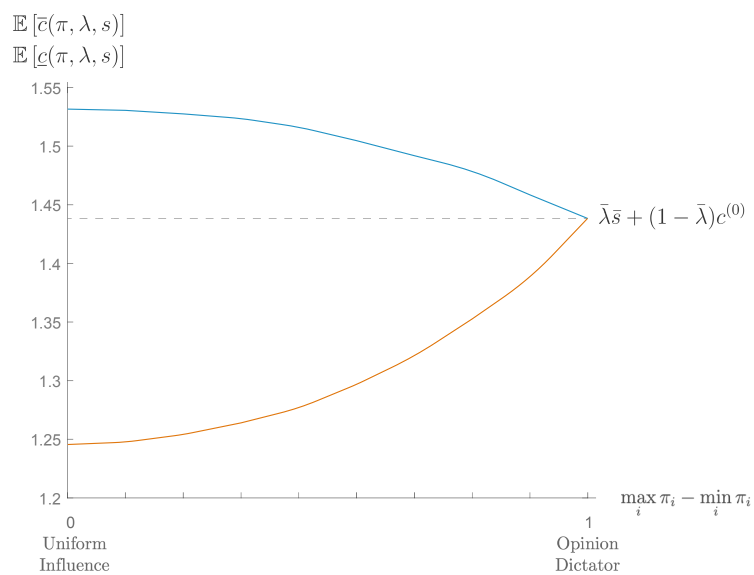
Proposition 7 (Influence and Sequence Susceptibility)**.** $\mathbb{E}\left[\overline{c}(\pi,\lambda,s)-\underline{c}(\pi,\lambda,s)\right]$ *is maximized for a influence vector π with uniform influences $\pi_i=\frac{1}{n}\ \forall\, i$. As influence concentrates,*

[21] These weights are considered fixed as they are drawn once and remain unchanged regardless of the history of information release.

$\max_i \pi_i \to 1$, *toward a single agent (an opinion dictator), the gap shrinks to* 0.

Proposition 7 reveals that groups that are the most susceptible to the ordering of information are those in which each member has identical influence. To see why, recall that in the final consensus, the weight of each signal depends on, among other things, 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 maximal and minimal information sequence, numerous configurations are possible. Consequently, in expectation, changing the signal-release order causes a 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, which π_i value will be associated with which realized signal is not known ex-ante. Although the π values are fixed while signals are random, realizing that the maximal/minimal information-release sequence releases information in a monotonic fashion leads to an alternative interpretation. The order statistics of the signals may be considered fixed, whereas there is variation regarding which π_i value will be associated with which order statistic. This uncertainty does not affect the weight of 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.

Figure 1: Expected Maximal(Minimal) Final Consensus



The above graph is constructed from a network comprising five nodes. Signals are drawn from a normal distribution with a mean of 1.5 and a standard deviation of 1. The initial consensus is set to $c^{(0)} = 1.2$, while the mean weight placed on signals is $\bar{\lambda} = 0.8$. On the far left, the influence vector is set to $\pi = [0.2, 0.2, 0.2, 0.2, 0.2]$. In each iteration, the influence of nodes 2 to 4 is reduced by 0.02, while the influence of the first node increases by 0.08. On the far right, the influence vector becomes $\pi = [1, 0, 0, 0, 0]$.

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

Figure 1 illustrates how the gap between the expected maximal and minimal final consensus shrinks as the influence vector shifts from one in which all agents have equal influence to an influence vector that places weight 1 on a single agent. As can be seen, as the influence focuses on a single individual, the gap between the expected highest and lowest attainable final consensus shrinks, and consequently, beliefs become less and less affected by the sequencing of information.[22] If a decision-maker had control over the sequencing of information and wanted to manipulate the final consensus, they would prefer influence to be equally distributed among group members, which would maximize the attainable final consensuses, giving the decision-maker a richer set of choices.

6 Conclusions

We depart from the canonical DeGroot setup by allowing for information to arrive sequentially. In doing so, we study an important dimension that arises once Bayesian learning is no longer assumed: the order of information arrival affects final beliefs even when the information content is unchanged. We find the influence a signal has on the final consensus is no longer only a function of the social influence of the agent receiving it; the order of information release also plays a crucial role, as does the weight the agent places on the signal. Sequential arrival of information undermines proper information aggregation, with beliefs typically converging away from the truth regardless of the abundance of information; thus, celebrated features of the DeGroot model may be more fragile than previously thought. For fixed weights on agents' own signals, we identify the information sequence attaining the maximal and minimal posteriors and show that in expectation, groups most susceptible to the order of information are those in which all agents have equal influence. Experimental evidence from Reshidi (2023) reveals the order of information arrival indeed affects final beliefs formed by groups and that individuals update their beliefs following heuristics akin to those studied in this paper.

[22] The fact that the expected maximal(minimal) final consensus is maximized for a uniform influence vector is not a coincidence. The proof Proposition 7 reveals that not only the gap but also $\mathbb{E}[\bar{c}(\pi, \lambda, s)]$ ($\mathbb{E}[\underline{c}(\pi, \lambda, s)]$), is maximized(minimized) under uniform influences.

7 Appendix

7.1 Wisdom of Crowds Prerequisites: In More Detail

The Wisdom of Crowds We begin by defining the *wisdom* of crowds, as introduced in Golub and Jackson (2010). Let $(M(n))_{n=1}^{\infty}$ represent 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 θ, normalized without loss of generality to be 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 $\sigma^2 > 0$.[23] One of the main assumptions in the aforementioned paper is that the distribution of these beliefs has mean equal to θ. For any given n and realization of beliefs, as communication takes place, the belief of agent i in network n approaches a limit denoted by $b_{i,\infty}(n)$. Because these limit 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 the number of agents increases, the final beliefs converge in probability to the realized value θ. The assumption that each matrix is convergent implies

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

where $\pi_j(n)$ represents the influence of agent j in-network n, the jth entry in the left eigenvector corresponding to the eigenvalue equal to 1.

Proposition 2 in Golub and Jackson (2010)
If $(M(n))_{n=1}^{\infty}$ is a sequence of convergent stochastic matrices, it is wise if and only if the associated influence vectors are such that $\max_{i\leq n} \pi_i(n) \to 0$ as $n \to \infty$.

The intuition of the above result is as follows. If, as the number of agents grows, the influence of the most influential agent shrinks to 0, the idiosyncratic errors associated with each agent wash away, and consequently, the consensus converges to the mean of the initial beliefs, which is θ.

A key assumption here is that initial beliefs are centered around θ. If we interpret

[23] The results follow even when each belief is drawn from a different distribution with bounded support.

agents' initial beliefs as driven by unbiased signals they receive (signals centered around the realized value θ), for the agents to follow these signals fully, it must be that their prior is uninformative; otherwise, their beliefs at time $t = 0$, their optimal guesses, would not simply be equal to the signal they receive but would instead incorporate information from their prior.[24] Hence, if we want to interpret the setup analyzed in Golub and Jackson (2010) as one driven by information, we have to acknowledge this underlying assumption of an uninformative prior.

In our sequential-information-arrival model studied thus far, the initial consensus serves a comparable role to that of a prior. As shown, the initial consensus plays a key role in determining agents' beliefs since it shows up even in the final consensus. To create an environment more aligned with Golub and Jackson (2010), we have to get rid of the initial consensus. However, modifying the model to get rid of the prior raises the issue of handling belief updating when facing agents whose beliefs are an empty set; these agents would be the ones who do not receive a signal in the first round. Banerjee et al. (2021) study an extension of the DeGroot model in which all information arrives in the first round; however, not all agents receive a signal. They modify the DeGroot model by proposing a learning rule for agents whose beliefs are an empty set. Below, we highlight some of the main takeaways of this model since they will shortly become relevant.

The Generalized DeGroot Model Banerjee et al. (2021) 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 $b_{i,t} \in \mathbb{R}$, whereas 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 $b_{i,0} = \theta + \varepsilon_i$, $\varepsilon \sim F(0, \sigma^2)$. At time $t = 0$, a subset of agents receive a signal, while the remaining agents never receive a signal. All agents with whom 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 *generalized DeGroot* updating as follows [25]

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

Under this specification, agents who do not receive a signal are initially uninformed until one

[24] For example, in the case of a normally distributed prior and signals, their optimal guess would be a convex combination of the mean of the prior and the realized signal, which would generically not be centered around the realized value θ.

[25] As discussed in Banerjee et al. (2021), the results generalize to non-uniform weighting.

of their neighbors becomes informed, in which case, they adopt the belief of their neighbor. If more than one neighbor is informed, agents average out these beliefs. A crucial assumption in this paper is that 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 exactly equal to this signal. Again, an information-driven interpretation of this setup corresponds to the uninformative prior case, leading agents who receive signals to fully adopt their signal as their optimal guess.

The authors find that when uninformed agents are present, even if $\max_i \pi_i(n) \to 0$ as $n \to \infty$, *wisdom* may fail. However, they identify network structures for which *wisdom* prevails; as such, the final beliefs converge in probability to the mean of the signals the agents receive. Because this section aims to identify conditions in which wisdom may break down due to the sequential arrival of information and not due to the sparsity of information in the first information round, in what follows, we assume the network structure permits *wisdom* to go through in a *generalized DeGroot model.*[26]

Prior-free Sequential Belief Updating We now incorporate the above specification in our sequential setup. That is, we assume agents who do not receive a signal in the first information round, $i \notin \gamma(1)$, have beliefs set to an empty set $b_{i,0}^{(1)} = \emptyset$. On the other hand, agents who receive a signal in the first information round, $i \in \gamma(1)$, set their beliefs equal to their signals $b_{i,0}^{(1)} = s_i$. To accommodate this, in our initial setup, we must modify the weights agents place on their own signal. Specifically, let $\tilde{\lambda}_i(H^k)$ represent the original weight participant i places on their signal. We make the following modification: $\lambda_i(H^1) = 1$ if $\gamma_i = 1$, and $\lambda_i(H^k) = \tilde{\lambda}_i(H^k)$ if $\gamma_i \neq 1$. When the first set of signals are released, beliefs update as in the *generalized DeGroot model*, with agents paying attention only to informed neighbors

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

After the first round of information release, and after information disseminates across the network, there will no longer be any agents whose beliefs are an empty set. Thus, afterwards, belief updating under this new specification becomes identical to belief updating in the initial setup. The difference is that the first consensus $c^{(1)}$ is a convex combination of only the signals released in the first information round, excluding the initial consensus $c^{(0)}$.

We have now laid the groundwork for analyzing whether *wisdom* persists in a sequential-information-arrival setting. We can do so both for the initial and the prior-free setup.

[26]For more details on the structure of these networks, see Banerjee et al. (2021).

7.2 Proofs

Proof of Proposition 2.

Assume there exists at least one information release history such that for at least two agents i and j, $\max_k \pi_i \lambda_i(H^k) > 0$, and $\max_k \pi_j \lambda_j(H^k) > 0$. By doing so, we simply exclude trivial cases in which beliefs are always equal to the initial consensus or cases in which only one agent places a non-zero weight on their signal—making sequential information arrival meaningless. Let ψ represent the mapping from agents to information-release rounds. Thus, $\psi(i)$ maps agents $i \in \{1, \ldots, n\}$ to their information arrival round $k \in \{1, 2, \ldots, K\}$. Without loss of generality, let $\psi(i) < \psi(j)$ in the information release history in which the above assumption holds.[27] Consider the case in which $s_i \neq c^{(0)}$. From Proposition 1 we know that after all information has been released, in the final consensus that emerges, the weight on signal s_i is

$$\prod_{z=\psi(i)+1}^{K} \left(1 - \sum_{l \in \gamma(z)} \pi_l \lambda_l(H^z)\right) \pi_i \lambda_i \left(H^{\psi(i)}\right).$$

Consider the alternative information release sequence in which for all $l : \psi(l) > \psi(i)$ the new information arrival time is set to $\psi(i)$. Then, the new weight on signal s_i will be $\pi_i \lambda_i \left(H^{\psi(i)}\right)$, an unambiguous increase, since we know that at least for j $\pi_j \lambda_j(H^{(\psi(j)})$ was greater than 0. There are now three cases to consider. First, if weights associated with other signals remain unchanged, the final consensus changes since the weight on s_i increased, thus beliefs are not *sequencing independent*. Second, if weights associated with other signals change in a way that does not perfectly offset the change of the weight in s_i, leading to a change in the final consensus, beliefs are once more not *sequencing independent*. Finally, consider the case in which the weights on other signals changed in a way that perfectly offset the increase in the weight of signal s_i, leading to the same $c^{(K)}$ consensus once more. For brevity, let the original weights associated with each signal be denoted as α_i and the new weights as α'_i. Relabel $c^{(0)}$ as s_{n+1} and define $\alpha_{n+1} := (1 - \sum_i \alpha_i)$. Then, we know that $\sum_i \alpha_i s_i = \sum_i \alpha'_i s_i = c^{(K)}$, with $\alpha_i \neq \alpha'_i$, and for at least one other k, $\alpha_k \neq \alpha'_k$. Define the vectors of weights for which the convex combination of the signals leads to $c^{(K)}$ as follows $\tilde{\alpha} := \left\{\alpha : \sum_i \alpha_i s_i = c^{(K)}\right\}$, which is a hyperplane within the n+1 dimensional polytope. The additional constraint reduces the dimensionality of the attainable weights from n to $n-1$. Thus, we can always find a vector of signals s' such that $\sum_i \alpha_i s'_i = c^{(K)}$, and yet $\sum_i \alpha'_i s'_i \neq c^{(K)}$, implying once more that beliefs are not *sequencing independent*.

□

[27] If this is not the case simply relabel i and j or follow the process in reverse.

Proof of proposition Proposition 3.

Let ψ represent the mapping from agents to information-release rounds. Thus, $\psi(n)[i]$ maps agents $i \in \{1, \dots, n\}$ in network n, to the their information arrival round $k \in \{1, 2, \dots, K\}$.

Sequential DeGroot Model with Signal Conditioning

Let S_k be an element of a measurable partitions of the set of values that signals can take, such that the intersection of each element is empty and the union of all elements is equal to the whole set. Let

$$w_k = \int \mathbb{1}\{x \in S_k\} dF \qquad\qquad \mu_k = \mathbb{E}[s | s \in S_k]$$

Further define

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

Replacing $\pi_i(n)$ with $\tilde{\pi}_i(n)$, realizing that $\tilde{w}_{\psi(n)[i]}(n)$ is the same for all $j \in \gamma(n)[i]$, we can write the final consensus after a sequential release of information as

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

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

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

Where the last part follows from the fact that $Var\left(\mathbb{1}_{i \in \gamma(n)[k]}\right)$ is bounded, $\sum_{i=1}^{n} \pi_i(n) = 1$ and the assumption that $\max_{i \leq n} \pi_i(n) \to 0$ as $n \to \infty$. 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 $n \to \infty$. Since $\lambda_i^{(k)}(n)$ and $s_i(n)$ are independent, $\mathbb{E}\left[\lambda_i^{(k)} s_i(n)\right] = \bar{\lambda}^{(k)} \theta$. Since both are bounded within $[0, 1]$, $Var\left(\lambda_i^{(k)} s_i(n)\right) \leq$

1. From **Lemma 1** in Golub and Jackson (2010), it follows that[28]

$$\underset{n\to\infty}{\text{plim}} \sum_{i\in\gamma(n)[k]} \tilde{\pi}_i(n)\lambda_i^{(k)}(n)s_i(n) = \bar{\lambda}^{(k)}\mu_k$$

We have thus identified the probability limit of each element in $c_T^{(K)}(n)$. From the properties of probability limits it then follows that

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

The above limit differs from θ for two reasons. First the weight on $c^{(0)}$ is nonzero, since H is non-atomic, the probability that θ is equal to $c^{(0)}$ is zero. Second, note 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 $\mathbb{E}[s_i] = \sum_{i=1}^{K} w_k\mu_k = \theta$, 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 $\sum_{j=1}^{K}\left(\prod_{k=j+1}^{K}\left(1 - w_k\bar{\lambda}^{(k)}\right)\right) w_j\bar{\lambda}^{(j)} \leq w_j\bar{\lambda}^{(j)}$.

Sequential DeGroot 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 $\tilde{w}_i(n)$ and $\tilde{\pi}_i(n)$. Following identical steps as above, it is straightforward to see that the probability limit of $\tilde{w}_j(n)$ is $\hat{w}_j$, the probability of a signal being released in information round j. The main difference from the previous specification is the following probability limit

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

[28] The expected weight in round k is $\bar{\lambda}^{(k)}$ since the distribution of weights is history but not network dependent. It is straightforward to see that all results go through if we allow for network dependence, with the minor change of $\bar{\lambda}^{(k)}$ being replaced with whatever the expectation might be.

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

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

In this case, the final beliefs do not converge to the realized value θ only because the mean of the prior $c^{(0)}$ has non-zero weight. Note that this will be the case even if $\lambda_i^{(k)}(n) = 1 \, \forall i, k$ and n.

Prior-free Sequential DeGroot Model with Signal Conditioning

Adopting the *generalized DeGroot model* implies that once the first round of signals is released, the prior is completely washed away.

As mentioned in the main text, so as to not conflate the channels of why *wisdom* may fail, we assume that the network structure permits *wisdom* to go through in a *generalized DeGroot model*.[29] This implies that, the consensus reached after the first set of signals is released converges in probability to the mean of these signals, $\operatorname{plim}_{n\to\infty} c^{(1)} = \mu_1$, as $n \to \infty$.[30] 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, and effectively start the process from the second information round. Following identical steps as in the analysis before leads to

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

Which will generically not be equal to μ.[31] Hence, in this case, although the mean of the prior has washed out, beliefs do not converge to θ as the conditional means of the signals are aggregated with distorted weights.

Prior-free Sequential DeGroot Model without Signal Conditioning

Once more, following almost identical steps as above, but realizing that when the information

[29]For more details on the structure of these networks see Banerjee et al. (2021).

[30]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.

[31]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 $\operatorname{plim}_{n\to\infty} c^{(K)}$ equal to μ adds a new constraint not implied by the previous constraints, reducing the dimensionality to $2K-3$.

release rule can not condition on signal realization $\mu_k = \theta \ \forall k$ leads to

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

Thus, under this specification, *wisdom* prevails despite of the distorted weights associated with signals on different information-release rounds. This follows since within each round, the mean of the signals is θ.

In all above cases, the converse, that is, if $\max_{i\leq n} \pi_i$ converges to some value strictly larger than 0, follows from the last part of the proof of **Lemma 1** in Golub and Jackson (2010). Converge in probability to 0 will fail for at least one of the $c^{(k)}$ consensuses and consequently for the final consensus $c^{(K)}$. □

Proof of Proposition 4.

Lemma 1. *If $s_i < s_j$, the maximal consensus release sequence can not have s_j released individually in information-release round k and s_i released individually in information-release round $k + 1$.*

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 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 maximal 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 maximal. 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 $\gamma(z)$ for all $z > k$ one period later) would shift the final consensus from

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

To the new final consensus c'_T

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

First, note that the weights on all other signals released at, or after k, remain unaffected. 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 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 on the weights on the final consensus:

- Is unaffected, for signals released in information round k that are not shifted, as well as all signals released after information round k.
- Decreases, for signals in set J.
- Increases, for all signals initially released before information round k, including the initial consensus $c^{(0)}$.

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 in information round $k-1$, can be calculated following similar steps as above. The total impact of such a shift on the weights on the final consensus:

- Is unaffected, for signals initially released after or on information round k.
- Increases, for all signals initially released in information round $k-1$.
- Decreases, for all signals initially released before information round $k-1$, including the initial consensus $c^{(0)}$.

Knowing the effect that **Splitting Information Sets** and **Merging Information Sets** have, allows us to know the effect that any relevant reshuffling of signals has.

Possible Joint Releases
Assume that under the maximal consensus release sequence, multiple signals are released in information round $\hat{k}$.

First, it must be that all 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 $c^{(K)}$. This follows from the **Splitting Information Sets** effect. The weight this signal would lose is exactly the weight at $c^{(\hat{k}-1)}$ would gain.

Second, 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 $\hat{k}$, thus having a **Merging Information Sets** effect. This boosts the weight of the signals at $\hat{k}$ and decreases the weight of the consensus at $c^{(\hat{k}-1)}$ by exactly the same amount. This would lead to an increase in the final consensus, thus contradicting the optimality of the information release sequence.

Consequently, in the maximal consensus sequence, a joint 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 Maximal Consensus Sequence

From **Possible Joint Releases** we know that the maximal 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. If a joint release occurs, all signals must be greater than the previous consensus.

Thus, $\gamma(k)$ for all $k \in \{1, \dots, K-1\}$ 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 $i \in \{1, \dots, K-1\}$ and any $j \in \{1, \dots, K-1\}$ if $s_i < s_j$ then s_i is released earlier than s_j. □

Proof of Corollary 1.

Corollary 1 follows from the proof of Proposition 4 and symmetry. □

Proof of proposition Proposition 5.

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 whose signals will be released in round k. 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 as follows:

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

The difference from equation (2) is that we no longer assume a consensus has been reached. Note that $\tilde{b}_i^{(k)}$ is continuous in all of its components. 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^r

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

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 [32]. 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 $\left|[M^{r_{\tilde{\varepsilon}}}]_{ij} - \pi_j\right| < \tilde{\varepsilon}$ 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 maximal release sequence γ^* that maximized $c^{(K)}$ under $r = \infty$ we had

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

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

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 maximal release sequence for $r = \infty$, this release sequence will be maximal for all cases in which $r \geq \tilde{r}$, where the exact value of $\tilde{r}$ depends on the network structure. $\square$

Proof of proposition Proposition 6.
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. In particular $\gamma(n)[k]$ represents the 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 $i \in \{1, \ldots, n\}$ in network n, to the $k \in \{1, 2, \ldots, K\}$ release rounds. Define

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

$\underline{s}(n)[k]$ and $\overline{s}(n)[k]$ represent the lowest and highest signal released in information round k in network n. Define

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

$\check{W}(n)[k, z]$ and $\tilde{w}(n)[j, z]$ represents the impact of signals released in information round k in network n, with values lower than z and higher than z respectively. Where by impact of a signal we refer to the influence of the agent receiving it $\pi_i(n)$ multiplied by the value this agent assigns to the signal $\lambda_i(n)$. Let $j < k$ and further define

$$\bar{W}(n)[k, j] := \max_{z(n)[k,j] \in [\underline{s}(n)[k], \bar{s}(n)[j]]} \left\{ \min \left\{ \tilde{w}(n)[j, z] \, , \, \check{W}(n)[k, z] \right\} \right\}$$

In network n, $\bar{W}(n)[k, j]$ is either equal to the impact that signals released in round k have, with signal values lower than $z^*(n)[k, j]$, or it's equal to the impact that signals released in round j have, with signal values larger than $z^*(n)[k, j]$, whichever value is smaller. In particular, $z^*(n)[k, j]$ 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.

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

That is, among all pairings k and j, $\bar{W}(n)$ 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 that maximize the function above, as well as $z^*(n)$ which is the argument that maximizes $\bar{W}(n)[k^*(n), j^*(n)]$. Note that if $\bar{W}(n) = 0$ then released signals are all sorted, such that the lowest signal released in any round k must be larger than the highest signal released in round $k-1$. Define

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

For network n, $\hat{K}(n)$ and $\check{K}(n)$ represent the set of agents whose signals are released in round $k^*(n)$ or $j^*(n)$, with values smaller than or lower than $z^*(n)$ respectively. Note that since $\max_{i\in\{1,\ldots,n\}} \pi_i(n) \to 0$ 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 $\tilde{w}(n)[j^*(n), z^*(n)] \leq \check{W}(n)[k^*(n), z^*(n)]$. Define

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

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

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

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

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

$$
\begin{aligned}
&+\left(\sum_{j\in\hat{K}(n)}\pi_j(n)\lambda_j(n)s_j(n) - \sum_{k\in\tilde{K}(n)}\pi_k(n)\lambda_k(n)s_k(n)\right)\left(\prod_{m>k^*(n)}\left(1-\sum_m \pi_m(n)\lambda_m(n)\right)\right)\\
&\left(1-\left(\prod_{j^*(n)<m<k^*(n)}\left(1-\sum_m \pi_m(n)\lambda_m(n)\right)\right)\left(1-\sum_{v\in\left\{\hat{K}(n)\cup\gamma(n)[k^*(n)]\setminus\tilde{K}(n)\right\}}\pi_v(n)\lambda_v(n)\right)\right)
\end{aligned}
$$

All terms, besides the last, are multiplied by $\triangle(n)$. All parts multiplying $\triangle(n)$ are weighted convex combinations of the prior and 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 greater than 0. 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 maximal consensus release sequence has overlapping signals. □

Proof of proposition Proposition 7.
We show that $\mathbb{E}\left[\bar{c}(\pi,\lambda,s)\right]$ is maximized with uniform influences. That $\mathbb{E}\left[\underline{c}(\pi,\lambda,s)\right]$ is minimized for uniform influences will follow from symmetry, implying that the gap will be maximized under uniform influences.

From Proposition 4, it follows that for any realization of signals and weights agents place on their signals, the maximal 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 $k \in \{0, 1, \dots, n-1\}$, after which all remaining signals are released jointly. Denote such a sequence that releases signals monotonically up to k, and jointly releases all signals afterwards by $\gamma^{(k)}$. 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, after 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 under the $\gamma^{(k)}$ information release sequence is

$$\begin{aligned}
\mathbb{E}\left[c^{(k+1)}(\gamma^{(k)}, \pi, \lambda, s)\right] = \mathbb{E}&\left[\left(1 - \sum_{i \in \gamma^{(k)}(k+1)} \pi_i \lambda_i\right) \sum_{j=1}^{k} \left(\prod_{z=j+1}^{k} (1 - \pi_z \lambda_z)\right) \pi_j \lambda_j s_j\right. \\
&+ \left. \sum_{j \in \gamma^{(k)}(k+1)} \pi_j \lambda_j s_j + \left(1 - \sum_{i \in \gamma^{(k)}(k+1)} \pi_i \lambda_i\right) \prod_{z=1}^{k} (1 - \pi_z \lambda_z)\, c^{(0)}\right] \\
=\mathbb{E}&\left[\left(1 - \sum_{i \in \gamma^{(k)}(k+1)} \tilde{\pi}_i \bar{\lambda}\right) \sum_{j=1}^{k} \left(\prod_{z=j+1}^{k} \left(1 - \tilde{\pi}_z \bar{\lambda}\right)\right) \tilde{\pi}_j\right] \bar{\lambda} S_{(j)} \\
&+ \sum_{j=k+1}^{n} \mathbb{E}\left[\tilde{\pi}_j\right] \bar{\lambda} S_{(j)} + \mathbb{E}\left[\left(1 - \sum_{i \in \gamma^{(k)}(k+1)} \tilde{\pi}_i \bar{\lambda}\right) \prod_{z=1}^{k} \left(1 - \tilde{\pi}_z \bar{\lambda}\right)\right] c^{(0)}
\end{aligned}$$

Where $c^{(0)}$ is the initial consensus. 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 $\pi = \{\pi_1, \ldots, \pi_n\}$ 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 $i \in \{1, 2 \ldots, n\}$, are drawn from π uniformly without replacement. Let a special case of the influence vector be denoted as $\pi_u = \{1/n, 1/n, \cdots, 1/n\}$, where u stands for uniform. Let another special case of the influence vector be denoted as $\pi_f = \{1, 0, \ldots, 0\}$ where f stands for focused. The particular coordinate of the weight 1 influence value is irrelevant. Let $p(\pi_i)$ represent the probability that the realized value of $\tilde{\pi}_n = \pi_i$. First, note that

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

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 values associated with signal released in the last round $(k+1)$ is $\mathbb{E}\left[\tilde{\pi}_n\right] = \frac{1}{n}$. Since the expected weight these signals have on the final consensus is $\mathbb{E}[\tilde{\pi}_n]\bar{\lambda}$, this weight remains unaffected regardless of the distribution of influences in the π vector. On the other hand, the weight associated with the k order statistic is

$$\mathbb{E}\left[\left(1-\sum_{i=k+1}^{n}\tilde{\pi}_i\bar{\lambda}\right)\tilde{\pi}_k\right]\bar{\lambda} = \left(\mathbb{E}\left[\tilde{\pi}_k\right]-\sum_{i=k+1}^{n}\mathbb{E}\left[\tilde{\pi}_k\tilde{\pi}_i\right]\bar{\lambda}\right)\bar{\lambda}$$
$$= \left(\frac{1}{n}-\sum_{i=k+1}^{n}\mathbb{E}\left[\tilde{\pi}_k\tilde{\pi}_i\right]\bar{\lambda}\right)\bar{\lambda}$$

To calculate $\mathbb{E}\left[\tilde{\pi}_k\tilde{\pi}_i\right]$ we define[33]

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

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

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 $\pi_j = 1$ for some j and $\pi_i = 0$ for all $i \neq j$. Consequently $\mathbb{E}[\tilde{\pi}_k\tilde{\pi}_i] \in \left[0, \frac{1}{n^2}\right]$, hence, the weight placed on the k'th order statistic is within $\left[\frac{n-(n-k)\bar{\lambda}}{n^2}\bar{\lambda}, \frac{1}{n}\bar{\lambda}\right]$. The weight placed on the $k-1$ order statistic is

$$\begin{aligned}&\mathbb{E}\left[\left(1-\sum_{i=k+1}^{n}\tilde{\pi}_i\bar{\lambda}\right)\left(1-\tilde{\pi}_k\bar{\lambda}\right)\tilde{\pi}_{k-1}\right]\bar{\lambda} \\ &= \left(\mathbb{E}\left[\tilde{\pi}_{k-1}\right]-\sum_{i=k}^{n}\mathbb{E}\left[\tilde{\pi}_k\tilde{\pi}_i\right]\bar{\lambda}+\sum_{j=k+1}^{n}\mathbb{E}\left[\tilde{\pi}_{k-1}\tilde{\pi}_k\tilde{\pi}_j\right]\bar{\lambda}^2\right)\bar{\lambda}\end{aligned}$$

To calculate $\mathbb{E}\left[\tilde{\pi}_{k-1}\tilde{\pi}_k\tilde{\pi}_j\right]$ we define

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

[33] 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.

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

Once more, the highest value $(3\sum_{i=1}^{n}\pi_i^2 - 2\sum_{i=1}^{n}\pi_i^3) = 1$ is achieved when $\pi_i = 1$ for some i and $\pi_j = 0\ \forall\, j \neq i$, while the lowest value $(3\sum_{i=1}^{n}\pi_i^2 - 2\sum_{i=1}^{n}\pi_i^3) = \frac{3n-2}{n^2}$ is achieved when $\pi_i = \frac{1}{n}\ \forall\, i$. Thus, $\mathbb{E}[\tilde{\pi}_{k-1}\tilde{\pi}_k\tilde{\pi}_j] \in \left[0, \frac{1}{n^3}\right]$, and so, the weight placed on the $k-1$ order statistic is within $\left[\frac{(n-\bar{\lambda})(n(1-\bar{\lambda})+k\bar{\lambda})}{n^3}\bar{\lambda}, \frac{1}{n}\bar{\lambda}\right]$. Continuing in this fashion, it is straightforward to show that the weight on the j'th order statistic, for any $j \leq k$, will fall within an interval bounded from below when $\pi_i = \frac{1}{n}\ \forall\, i$, and bounded from above when $\pi_i = 1$ for some i and 0 for all $j \neq i$.[34] The weight on each order statistic can be at most $\bar{\lambda}/n$, but will be lower if information is released sequentially and $\pi \neq \pi_f$. Note that $\mathbb{E}\left[c^{(k+1)}\left(\gamma^{(0)}, \pi\right)\right] = \mathbb{E}\left[c^{(k+1)}\left(\gamma^{(k)}, \pi_f\right)\right]\ \forall k$, that is, for a given π, releasing all information jointly generates the same final consensus as any order of information release under π_f. Furthermore, for all $j \leq k$, the weight on the j'th order statistic can be written as $\frac{1}{n} - \psi_j(\pi)$. Where by how much the weight on the j'th order statistic is reduced $\psi_j(\pi)$ depends on the vector π. Again, $\psi_j(\pi) = 0$ under π_f, while it is maximized under π_u, where π_f and π_u are as defined above. Since the weight on all order statistics and the initial prior sums up to one, this weight reduction is directly transferred to the initial prior. Then, for a information release rule $\gamma^{(k)}$ and vector of influences π, we can write the expected final consensus as

$$
\mathbb{E}\left[c^{(k+1)}\left(\gamma^{(k)}, \pi\right)\right] = \mathbb{E}\left[c^{(k+1)}\left(\gamma^{(0)}, \pi\right)\right] + \sum_{i=1}^{k}\psi_i(\pi)\left(c^{(0)} - S_{(i)}\right)
$$

For the special case of $\pi = \pi_u$ we write the final consensus as

$$
\mathbb{E}\left[c^{(k+1)}\left(\gamma^{(k)}, \pi_u\right)\right] = \mathbb{E}\left[c^{(k+1)}\left(\gamma^{(0)}, \pi_u\right)\right] + \sum_{i=1}^{k}\phi_i\left(c^{(0)} - S_{(i)}\right)
$$

Where $\phi_i = \psi_i(\pi_u)$. From the arguments above we know that $\phi_i \geq \psi_i(\pi)\ \forall\, i$ and for all π. From the arguments above we also know that $\mathbb{E}\left[c^{(k+1)}\left(\gamma^{(0)}, \pi\right)\right] = \mathbb{E}\left[c^{(k+1)}\left(\gamma^{(0)}, \pi_u\right)\right]$, since, when all information is released jointly, the weight on each order statistic will be $\bar{\lambda}/n$ regard-

[34] In the special case of $k = n-1$, where each signal is released in a separate round, it is straightforward to see that the weight placed on the j'th order statistic will be within $\left[\frac{(n-\bar{\lambda})^{n-j}}{n^{n-j+1}}\bar{\lambda}, \frac{1}{n}\bar{\lambda}\right]$.

less of the vector π. It then follows that $\mathbb{E}\left[c^{(k+1)}\left(\gamma^{(k)},\pi\right)\right]$ differs from $\mathbb{E}\left[c^{(k+1)}\left(\gamma^{(k)},\pi_u\right)\right]$ only through the difference in the $\psi_i(\pi)$ and ϕ_i weights, with the later being larger. From Proposition 4 we know that in the maximal release sequence $\gamma^{(k^*)}$, the signal of agent i is released sequentially only if $s_i < c^{(0)}$. Thus, under $\gamma^{(k^*)}$, $c^{(0)} - S_{(i)} > 0$ for $i \leq k^*$. Consequently, as the $\psi_i(\pi)$ values increase, so does the value of $\mathbb{E}\left[c^{(k+1)}\left(\gamma^{(k^*)},\pi\right)\right]$. Hence, $\mathbb{E}\left[c^{(k+1)}\left(\gamma^{(k^*)},\pi\right)\right]$ is maximized under $\pi = \pi_u$. By symmetry and Corollary 1, we know that the expected minimal final consensus is also minimized under $\pi = \pi_u$. Therefore, the expected gap between the maximal and minimal final consensus is maximized under $\pi = \pi_u$.

From the fact that $\mathbb{E}\left[c^{(k+1)}\left(\gamma^{(k)},\pi_f\right)\right] = \mathbb{E}\left[c^{(k+1)}\left(\gamma^{(0)},\pi_f\right)\right]$, and that this holds for the miniimal sequence as well, it follows that the expected gap shrinks to 0 as $\max_i \pi_i \to 1$.

$\square$

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.

Banerjee, A., Breza, E., Chandrasekhar, A. G., and Mobius, M. (2021). Naive learning with uninformed agents. *American Economic Review*, 111(11):3540–74.

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. *Available at SSRN 2704746*.

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. (2023). An experiment on social learning with information sequencing. *Working Paper*.

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.