Information Aggregation under (not so) Naive Learning

Abhijit Banerjee and Olivier Compte

March 2021

Abstract

We explore a model of non-Bayesian information aggregation in networks. Agents non-cooperatively choose an aggregation rule from the Friedkin-Johnsen (FJ) class to maximize private payoffs in the presence of noise in information transmission. We characterize rules that get chosen. The well-known DeGroot (DG) rule, nested in FJ, is never chosen – all near optimal rules have individuals putting enough weight on their own initial opinion in every period unlike in DG. This precludes full consensus even in the long-run but ensures better information aggregation. This trade-off extends to a wider range of environments and a broader class of rules.

1 Introduction

Living in a world dominated by Facebook, Twitter and their ilk, it is hard to avoid wondering about the quality of information aggregation on networks. We constantly get information passed onto us by others and in turn pass it on to our network neighbors. What are properties of such a process? How well is information aggregated?

The literature that theoretically explores these questions typically takes one of two routes: a *Bayesian route*, in which agents make correct inferences based on an understanding of all the possible ways information can transmit through the network; and a *non-Bayesian route*, which avoids these very demanding assumptions about information processing by postulating a simple rule that individuals use to aggregate own and neighbors opinions.[1]

[1] A Bayesian needs to think through all possible sequences of signals that could be received as a function of the underlying state and all the possible pathways through which each observed sequence of signals could have reached them. There is obviously a very large

One such simple rule is the DeGroot (DG) rule, where agents update their current opinion by averaging their neighbors' most recent opinions with their own. This is usually justified by arguments that this is like Bayesian in the sense that it coincides with the Bayesian decision rule in certain simple cases, though recently Molavi et al. (2018) provide an axiomatic justification for DG-like (e.g. Log-linear learning) rules.[2]

This paper proposes a *third route*. It considers a *class* of simple aggregation rules and postulates that, within this class, each individual selects his or her favorite aggregation rule, based on its instrumental value.[3] We focus on the class of Friedkin-Johnsen (FJ) rules (Friedkin and Johnsen (1990)), which nest DG, but allow each individual to keep putting some weight on their own initial opinion. Within this limited class of "natural" rules, we allow agents substantial discretion in the choice of rules and assume that each individual selects the one that best aggregates information (for her) in the long-run.[4] In other words, instead of requiring our simple rule to be like Bayesian, we ask whether the agents would want a rule that has this characteristic.

We explore this third route in a variant of the standard setup where each individual initially gets a noisy signal correlated with some underlying state of the world, shares his current aggregate of information with his network neighbors, and importantly, information transmission is assumed to be noisy. The class of Friedkin-Johnsen (FJ) models can formally be written as

$$y_i^t = (1-\gamma_i)y_i^{t-1} + \gamma_i(m_i x_i + (1-m_i)z_i^{t-1}) \tag{FJ}$$

number of such pathways. Alatas et al. (2016) remark "To give a sense of scale to this computation, note that enumerating all such paths is # P-complete and a random graph with n nodes and edges with probability p_n has an expected number of paths between nodes 1 and n given by $(n-2)!p_n^{n-1}e\,(1+o\,(1))$, which is potentially an enormous number (Roberts and Kroese (2007)."

[2]For example, DeMarzo et al. (2003), who brought it into economics literature and following them, Golub and Jackson (2010), justify DG by arguing that it coincides with the Bayesian rule in the static case. In an environment where the underlying state changes over time, Alatas et al. (2016) justify a simple updating rule by its similarity to a Bayesian rule in special cases, and Dasaratha et al. 2020 argue that in their set-up the Bayesian rule is DG-like. Finally, Levy and Razin (2015) develop the Bayesian Peer Influence Paradigm to capture the idea of an almost Bayesian aggregation rule.

[3]This is in the spirit of the approach advocated in Compte and Postlewaite (2018) to model mildly sophisticated agents.

[4]The limitation to a class of rules is key. With no limitation, the Bayesian rule is the individually optimal way to process signals among all possible signal processing rules.

where y_i^t is i's belief in period t, x_i is the initial signal that i received and

$$z_i^t = \frac{1}{\mid N_i \mid} \sum_{j \in N_i} y_j^t + \varepsilon_i^t. \tag{1}$$

is the average report received by i from his neighbors (denoted by N_i) plus any noise in the transmission (or reception) of that signal. When the weight m_i is 0, individual i is using a DG rule.[5] One key assumption is that there is noise in communication of signals (or alternately, there are biases in the reports individuals get from others). Another assumption is that only m_i and γ_i) will be subject to choice, not the relative weights put on neighbors:[6] this is to reduce the dimensionality of the rule-choice problem.

We then assume that the initial signals x_i are correlated with some underlying given state of the world θ and that the individuals have a utility function which is a decreasing function of distance between the state θ and their long-run belief about θ, formed after the signal exchange process has had a long enough run. Given these preferences, one can examine the performance of rules by computing expected utilities *on average* over the many opinion-formation problems that agents face, that is, on average over realizations of state, initial opinions and communication shocks.

Our main methodological assumptions are that (i) there is a force towards the use of higher performing rules (e.g., justified by evolution or reinforcement learning), and (ii) in this quest for higher performing rules, each individual can only examine (or compare, or get feedback about) a limited set of rules (e.g., the FJ class where each m_i belong to $[0, 1]$).

Formally, our analysis boils down to examining a rule choice game where, given the rules adopted by others, each player chooses the rule within the permitted set of rules that maximizes her own expected utility:[7] we are interested in the Nash equilibrium of this rule choice game.

To clarify the difference, note that a Bayesian optimizes his/her opinion-formation rule for each t and every possible realization of $(x_i, z_i^1, .., z_i^{t-1})$, while we restrict attention to a limited and fixed set of rules, applicable in all periods, parameterized by m_i and γ_i. These parameters are meant to capture some general features of opinion formation: specifically the *persistence* of initial opinions, and *speed of adjustment* of the current opinion.

[5]Throughout our analysis, we shall assume that all γ_i are strictly positive.

[6]In Expression (1), all neighbors contribute symmetrically. In our formal exposition we will actually allow for non-symmetric weights. Our main simplification is that we assume that relative weights across neighbors are fixed, not subject to optimization.

[7]Expected across different realizations of the state, initial opinions and communication shocks. m_i is thus "chosen" at an ex ante stage.

Our view is that these features probably do adjust to the broad economic environment agents face, but for each opinion-formation problem within a certain context, the actual sequence of opinions is mechanically generated given these features. For example, people may have one rule for all political opinions, but a different set of rules about all questions pertaining to where to go for vacation.

It is precisely this fact that the rule is very simple and that it applies across many different problems that makes our third route cognitively less demanding than the Bayesian route. While we agree that choosing m_i and γ_i optimally is a difficult problem which in principle requires knowledge of the structure of the model, there is no reason why the standard justification of Nash Equilibrium as a resting point of an (un-modeled) learning/evolutionary process would not apply here. Moreover, one of our most important results is that DG rules, and indeed all rules that put too little weight (m_i) on initial opinions, are dominated, suggesting a strong force away from DG even if agents find it difficult to find the exact optimal value of m_i.

Our main results and the logic behind them are as follows.

Fragility. We start by observing that DG has the undesirable property that the variance of every decision-maker's belief grows without bound in the presence of any noise in communication (Proposition 1).

This points to a specific sense in which DG is fragile: when $\gamma_i > 0$, agent i puts less than one hundred percent weight on his most recent belief, so the weight that agent i directly puts on his own initial signal is going to 0. Without noise in transmission, this is at least partly offset by the weight agent i puts on reports from others, which themselves contain agent i's initial signal. This is why the influence of initial signals on current beliefs does not dissipate. In other words, the agent holds on to his own signal only through the feed-back from others. The problem is that when transmission is noisy, you only get the feedback at the cost of some extra noise in every round. Given that the initial signals enter only at the beginning and the noise keeps coming in every period, it is no wonder the noise comes to dominate. This contrasts with DG in a noiseless environment, which has been shown by Golub-Jackson (2010) to have the attractive property that, under some restrictions on network structure and weights on neighbors, learning converges to perfect information aggregation (indeed within the FJ class, DG is the only rule with this property).

This intuition suggests that allowing each person's initial signal to come in with some weight every period would provide a countervailing force, which is what FJ does: under FJ rules, every decision-maker's average belief as well as its variance converges (Proposition 2). To see exactly why, note that under FJ, the prime mover of beliefs is a change in the beliefs of an agent's neighbors, who in turn are reacting to changes in the beliefs of their own neighbors. Potentially, a single change in the belief of one agent can unleash a sequence of changes in the beliefs of others, resulting in echo effects that end up altering everyone's beliefs. However, with (FJ) rules, because agent's always put some weight on their own initial signals, these echo effects are on average dampened at each round and this dampening guarantees convergence. In fact we show that convergence requires only one agent to put positive weight m_i on his initial signal.

However this convergence result does not imply a discontinuity between DG and FJ rules. We show that if the weight on initial signals in FJ is small for *all* players, the variance of the long term outcome will tend to be very large (see Proposition 3).

Incentives. We next examine incentives. The above observations suggest that the presence of transmission errors tends to favor the use of FJ-type rules with *significant* weight on the own signal. In fact Proposition 4 shows that rules where m_i is too low are dominated. One step of the argument is obvious: essentially, if every other agent chooses DG, then choosing FJ is the only way to stop the variance from blowing up. Moreover if some players use FJ, then those who stick with DG become followers: their initial opinions disappear from current opinions in the long-run, because they face players that constantly feed in their own initial signals. Long-run opinions are a weighted average of the signals of those who are putting positive weight on their signals. If an individual is currently putting zero weight on his own signal ($m_i = 0$), increasing that weight slightly always reduces the ex ante variance of his final opinion (because an additional independent signal always reduces ex ante variance).

In fact we can go a step further. Restricting attention to the simpler case where errors in transmission are modeled as a systematic bias drawn at the start of the interaction, long-run outcomes are independent of γ_i and we can characterize the equilibrium where all the players are non-cooperatively choosing the weight m_i to put on their own initial signal (Proposition 5). We show that the equilibrium always involves putting too little weight on one's own signal relative to the social optimum (Proposition 6). In equilibrium each player is trying too hard to free ride on the collective wisdom of the others, not fully taking into account the fact that setting m_i too low increases the correlation across opinions.[8]

We then return to the question of rule choice when there is also idiosyncratic noise in transmission. In the presence of idiosyncratic noise, FJ rules with large γ_i generate a lot of bouncing around because i is reacting every period to the latest reports from others and each of those comes with a different piece of noise in every period. i can limit the churn by putting some weight on past beliefs (i.e., a smaller γ_i), which is what Proposition 7 addresses.

Extensions. In the penultimate section of the paper we examine other potential sources of errors or shocks. Our first exercise shows that nothing essential changes in our analysis when the information transmitted is deliberately slanted in any direction, though there is a further shift towards reliance on one's own initial signal. A similar observation obtains when preferences are heterogenous and players have biased perceptions of others' preference.

The next sub-section shows that another key difference between FJ and DG rules comes from the way they deal with uncertainty over the exact communication protocol – for example the fact that not everyone may speak in every period. We show by example that the outcome from using DG rules is sensitive to who speaks when, even in the absence of noise, whereas under FJ rules, expected opinions are always independent of the communication protocol, whether or not there are errors in communication. Nevertheless, the long-run opinions under DG remain weighted averages of initial opinions, so the variance of long-run opinions induced by these kinds of shocks remains bounded, absent transmission errors.

In a similar vein, we examine the effect of uncertainty about the precision of the initial signals and show that, in the absence of transmission errors, this *does not* undermine the performance of DG-type rules. As a matter of fact, in a set-up where each participant only knows the precision of own initial signal, perfect information aggregation can be achieved under DG, by choosing γ_i that is suitably scaled to the precision. This observation delineates the key role played by transmission shocks in our analysis, as opposed to other sources of shocks.

We next turn to the possibility of coarse communication–say each party only reports their current best guess about which of two actions is prefer-

[8]Player i takes into account the correlation in opinions when setting own m_i optimally, but she does not fully take into account the consequence for player j, in particular, the positive correlation between the sources (including j's own signal) that directly contribute to j's opinion independently of i's mediation, and player i's opinion which also contributes to j's opinion and partially contains these same sources.

able. In this setting, the class of potentially "natural" rules is more limited: they include the infection models, studied in Jackson (2008) among (many) others, and the related class of models studied by Ellison and Fudenberg (1993, 1995). In this environment, systematic error in interpreting guesses by neighbors makes the long-run outcome from a DG-like rule entirely insensitive to the actual state of the world (Frick et al. (2019) report a related result for a proto-Bayesian rule), but this is not true for FJ-type rules

In the next sub-section, we move away from the linear aggregation rule assumption. We introduce a class of non-linear aggregation rules that remain in the spirit of DG rules, and show that the non-linearity may actually exacerbate the long-run drift in beliefs, suggesting the possibility that linear rules may actually be the best case scenario for DeGroot rules.

We end this section with a discussion of non-stationary rules and where and why they may not always be appropriate.

Literature review. DeGroot (1974) has had a significant impact in the economics literature, possibly due to its simplicity and Bayesian flavor. Friedkin and Johnsen (1990) provides a well-known alternative model, where interpersonal influence does not necessarily generate long-run consensus.[9]

In economics, our paper is related to and inspired by the recent upsurge of interest in the social learning with less than fully Bayesian agents. Eyster and Rabin (2010), Sethi and Yildiz (2012, 2016, 2019), Jadbabie et al. (2012) and Gentzkow et al. (2018), among others, explore the implications of applying Bayes rule when the underlying information structure is misspecified, as does the previously mentioned paper by Frick et al. (2019).

Finally it relates to the literature on Bayesian social learning (see above on why this approach imposes a heavy cognitive burden on agents). Acemoglu et al (2011) study the case where every agent updates their opinion and communicates it only once. They provide conditions on signals and network structure under which information is perfectly aggregated as the network grows to be very large.[10] More recent work, in which agents repeatedly communicate (like in the model we analyze) include Mossel et al. 2015 who derive necessary conditions on the network structure under which Bayesian learning yields consensus and perfect information aggregation.[11]

[9] For example, in subsequent work, Friedkin and Johnsen (1999, page 3) write, referring to the work of DeGroot and other precursors: "These initial formulations described the formation of group consensus, but did not provide an adequate account of settled patterns of disagreement".

[10] Bayesian models of social learning go back to Banerjee 1992, and Bhikchandani, Hirshleifer and Welch 1992 but the network structure they study is extremely special.

[11] They build on Rosenberg et al. 2009 and the literature on "Agreeing to Disagree"

2 Basic Model

2.1 Transmission on the network

We consider a finite network with n agents, assume noisy transmission/reception of information and define a simple class of rules that players may use to update their opinions.

Formally, at any date t, each agent i in the network has an opinion that can be represented as a real number.[12] We consider a class of updating rules due to Friedkin and Johnsen (1990) (henceforth FJ), in which player i's current opinion y_i^t is a convex combination of his initial opinion x_i, his most recent opinion y_i^{t-1} and some summary perception z_i^{t-1} of his neighbors' opinions. Formally, this can be written as

$$y_i^t = (1-\gamma_i)y_i^{t-1} + \gamma_i(m_i x_i + (1-m_i)z_i^{t-1}) \tag{FJ}$$

where

$$z_i^t = A_i.y^t + \varepsilon_i^t \tag{2}$$

where y^t is the vector of all opinions at t, A_i is a row vector whose jth element A_{ij} is such that $\sum_j A_{ij} = 1$ and ε_i^t represents an error in transmission or reception. z_i^t is meant to be some average of the opinions of i's neighbors (denoted N_i), so the presumption is that $A_{ij} > 0$ for $j \in N_i$. This average is then modified by some noise in transmission or reception.

When $m_i = 0$, the rule corresponds to the well-studied DeGroot rule (DG). When $m_i > 0$, the updating process works like DG, but the perception of other's opinions is adjusted using the decision-maker's own initial opinion as a perpetual seed. This perpetual use of the initial opinion in the updating process gives FJ a non-Bayesian flavor, since for a Bayesian, their prior (i.e., the seed) is already integrated into y_i^{t-1} and therefore there is no reason to go back to it.[13,14]

that goes back to Aumann 1976

[12] This opinion can be interpreted as a point-belief about some underlying state, which will eventually be used to undertake an action.

[13] In fact, as mentioned already, the one obvious attraction of *DG* has been its quasi-Bayesian flavor. If y_i^{t-1} is viewed as a summary statistic of past signals, and z_i^{t-1} as a new signal, then the linear weight γ_i can be seen as the optimal weighting strategy of a Bayesian aiming to reduce the variance of his or her opinion. Of course, over time, a Bayesian would typically not keep that parameter constant, as the relative informative content of their own current opinion and that of others will in general not be constant.

[14] Note that although the expression (FJ) encompasses the DG rule, we shall refer to FJ as a rule for which $m_i > 0$.

To avoid technical difficulties once we give agents discretion in choosing their updating rule, we set $\underline{\gamma} > 0$ arbitrarily small and restrict attention to FJ rules where $\gamma_i \geq \underline{\gamma}$. We also assume that the matrix A of the A_i's is connected in the sense that for some positive integer k, $A^k_{ij} > 0$ for all i, j. In other words everyone is within a finite number of steps of the rest.

Finally, before proceeding, it is useful to define a simplified version of the rule FJ, where $\gamma_i = 1$. We refer to it at SFJ:

$$y_i^t = m_i x_i + (1 - m_i) z_i^{t-1} \tag{SFJ}$$

One can think of SFJ as a process that works like FJ, except that agents do not attempt to smooth out variations in their own opinion. In the absence of idiosyncratic shocks on the perception of the opinions of others (see details below), SFJ and FJ will generate identical long-run opinions.

Note that all the rules considered are stationary, in the sense that the weighting parameters m_i and γ_i do not vary over time. We are interested in these rules not only because they have been studied in the literature, but also because we see them as plausible ways by which agents might incorporate others' opinions into their current opinion. Of course, with some knowledge of the structure of the network, and the process by which information gets incorporated, an agent might want to adjust the weights over time. We shall discuss in Section 7.7 the risks that such elaborate adjustments be misguided, in particular when there is randomness over the dates at which communication takes place.

We have also imposed the assumption that everyone operates on the same time schedule: periods are so defined that everyone changes their opinion once every period and everyone else gets to observe that change of opinion before they adjust their opinion in the following period. We will discuss what happens if we relax this assumption in Section 7.3.

2.2 Errors in opinion sharing

The term ε_i^t is an important ingredient of our model, meant to capture some imperfection in transmission.[15] It defines a distortion in what each individual "hears" that aggregates all the different sources of errors. distortions

[15] There has been several recent attempts to introduce noisy **or biased** transmission in networks. In Jackson et al. (2019), information is coarse (0 or 1), and noise can either induce a mutation of the signal (from 0 to 1 or 1 to 0) or a break in the chain of transmission (information is not communicated to the next neighbor). In Frick et al. (2019), agents communicate through a choice of action $a \in \{0, 1\}$ correlated with an unknown underlying state, and they make a **systematic** error in interpreting these actions because they have an erroneous model of the preferences of others. See Section 7.5.

may result from each individual being imprecise in expressing his or her opinion, or from an error in hearing or interpretation.

We assume that the error term has two components:

$$\varepsilon_i^t = \xi_i + \nu_i^t.$$

The term ξ_i is a *persistent* component realized at the start of the process, that applies for the duration of the updating process.[16] The term ν_i^t is an *idiosyncratic* component drawn independently across agents and time. We interpret ξ_i as a systematic bias that slants how opinions of others are perceived. For convenience, we assume that all error terms are homogenous across players and unbiased (that is, $E\xi_i = E\nu_i^t = 0$).[17] We let $\varpi_i = var(\xi_i)$ and $\varpi_0 = var(\nu_i^t)$ and assume that:

$$\varpi_i = \varpi > 0$$

2.3 The objective function

There is an underlying state θ, and agents want their decision to be as close as possible to that underlying state, where the decision is normalized to be the same as the agent's long-run opinion. In other words, we visualize a process where agents exchange opinions a large number of times before the decision needs to be taken.

Given this private objective, we explore each agent's incentives to choose his updating rule within the class of FJ rules to maximize his objective on average across realizations of the underlying state of the world, the initial opinions and the transmission errors. The set of possible updating rules is extraordinary vast, so the limitation to FJ rules is of course a restriction. Our motivation is to examine the incentives of *mildly* sophisticated agents who have some limited discretion over how they update opinions. In particular we have in mind examining whether there are forces away from DG rules, and whether private and social incentives differ. We also have in mind that rules apply across problems, which is why we will evaluate their ex ante performance.[18]

[16] One interpretation is that each information aggregation problem is characterized by the realization of an initial opinion vector x and persistent bias vector ξ, and that agents face a distribution over problems.

[17] The assumption $E\nu_i^t = 0$ is without loss of generality. We shall come back to the case where $E\xi_i \neq 0$ is the Discussion Section.

[18] That is, on average over states, initial opinions and transmission errors.

Formally, we assume that the initial signals are given by

$$x_i = \theta + \delta_i$$

where the θ are drawn from some distribution $G(\theta)$ with mean zero and finite variance, δ_i, ξ_i and ν_{it} are random variables that are independent of each other for all i and t and are also independent of θ. We assume that noise terms δ_i are unbiased, with variance σ_i^2. For convenience, we mostly assume that $\sigma_i = 1$ for all i, but we do not actually need this assumption.[19]

For any t, each profile of updating rules (m, γ) generates at any date t, a distribution over date t opinions. We now define the expected loss (where the expectation is taken across realizations of θ, δ_i, ξ_i and η_{it}, for all i and t):

$$L_i^t = E(y_i^t - \theta)^2$$

Define $\delta = (\delta_1, ..., \delta_n)$, $\xi = (\xi_1, ..., \xi_n)$ and $\nu_s = (\nu_{1s}, ..., \nu_{ns})$ for all s. Now given the set of updating rules that we consider, it will become evident that

$$y_i^t = b_i^t \delta + c_i^t \xi + \sum_{s=1}^{t} d_{is}^t \nu_s + \theta$$

for some non-negative vectors b_i^t, c_i^t and $\{d_{is}^t\}_{s=1}^t$.[20] It follows that

$$L_i^t = E[b_i^t \delta + c_i^t \xi + \sum_{s=1}^{t} d_{is}^t \nu_s]^2$$

We define the limit loss $L_i = \lim_{t \nearrow \infty} L_i^t$.[21] We assume that each agent i aims at minimizing L_i. Now whenever L_i is finite, we can write it as

$$L_i = L_i^0 + V_i \text{ where } L_i^0 \equiv E(b_i \delta + c_i \xi)^2.$$

The term L_i^0 results from variations in initial opinions and the persistent component, while the term V_i results from the idiosyncratic components only. Note that the distribution over θ plays no role, so θ can be normalized to 0.

[19] We analyze the case with heterogenous variances only when this has pedagogical value.

[20] This is because θ enters additively in all opinions.

[21] Alternatively, one could define $L_i = \lim_{h \searrow 0} (1-h) \sum h^{t-1} L_i^t$, assuming that the agent makes a decision at a random large date in the future and that his preference over decisions is $u_i(a_i, \theta) = -(a_i - \theta)^2$.

L_i is well-defined for any vector m, γ so long as $m \neq 0$. As it will turn out, for $m = 0$, L_i is infinite. Note that each player can secure $L_i \leq var(\delta_i) = \sigma_i^2 = 1$ by ignoring everyone else's opinions ($m_i = 1$).

There are two reasons why it is useful to decompose the loss L_i. First, it allows us to separate the effect of persistent and idiosyncratic errors. Second, it turns out that when all γ's are small, the losses V_i are small, and L_i^0 is then the preponderant loss.

In the next Section we start by exploring the long-run properties of different learning rules within the FJ class. Then we turn to the optimal choice of learning rules.

3 Properties of learning rules

We are interested in long-run opinions: whether they converge to some limit opinion and if they do, what determines the variance of the limit opinion. In particular what part of it comes from the "signal"− the original seeds − and what part from the noise that gets added along the way?

3.1 Exploding dynamics under DG

Our first result shows that if all agents follow a DG rule, as long as there is any idiosyncratic component in the noise, the variance of long-run opinions diverges. Moreover, for almost all realizations of the persistent component, y_i^t must diverge in expectation over time for all i.

To show this we fix x and ξ and define $\overline{y}_i^t = E y_i^t$ and $V_i^t = var(y_i^t)$. We have:

Proposition 1: *Assume that $m_i = 0$ for all i. (i) If $\varpi_0 > 0$, then for all i and any fixed x, ξ, $\lim_t V_i^t = \infty$. (ii) For almost all realizations of the persistent components ξ, $\lim \left|\overline{y}_i^t\right| = \infty$ for all i* and *x.*

For example, the proposition shows that a bias in a single player's perception ξ_1 may be enough to drive up the opinions of all: if $\xi_1 > 0$, say, the bias creates a discrepancy with other's opinions, and each time others' opinions catch up, player 1 further raises his opinion compared to others, prompting another round of catching up, and eventually all opinions blow up.

We present here some intuition that explains why DG works well without noise and becomes fragile as soon as there is some noise. Let y^t denote the vector of opinions at t. Let Δ_n be the set of vectors of non-negative weights $p = \{p_i\}_i$ with $\sum p_i = 1$. For any i, we have $y_i^t = B_i y^{t-1} + \gamma_i \varepsilon_i^t$ with $B_i \in \Delta_n$. Because the network is connected, there is a strictly positive vector of weights $\pi \in \Delta_n$ such that $\pi.B = \pi$,[22] so

$$\pi.y^t = \pi.y^{t-1} + \sum_i \pi_i \gamma_i \varepsilon_i^t.$$

Without noise, the limit weighted opinion $\pi.y$ coincides with the weighted initial opinion $\pi.x$. This explains why in the absence of noise the influence of initial opinions never dissipates (and also why all initial opinions matter – as $\pi >> 0$): the direct contribution of i's initial signal to i's opinion vanishes, but it surfaces back from the influence of neighbors' opinions (which increasingly incorporate i's initial signal), settling at a limit weight equal to π_i.

With noise however, $\pi.y$ is a random walk, explaining why the influence of initial opinions vanishes and why the variance diverge. Besides, the random walk has a drift when $\sum_i \pi_i \gamma_i \xi_i \neq 0$, explaining why $\pi.\overline{y}$ then diverge.

3.2 Anchored dynamics under FJ.

Fixing again x and ξ, we now examine long-run dynamics under FJ. Define $\overline{y}^t = (\overline{y}_i^t)_i$ and $V^t = (V_i^t)_i$ as the vector of expected opinions and variances.

Proposition 2. *Assume at least one player, say i_0, updates according to FJ (with $m_{i_0} > 0$).Then, for any fixed x, ξ, $\overline{y}^t$ and V^t converge. Besides, the limit variance V does not vary with x and ξ, and the limit vector of expected opinions $\overline{y}$ does not depend on the signal x_i of any individual with $m_i = 0$*.

Proposition 2 shows that to avoid that all opinions drift, it is enough that there is one player who continues to put at least a minimum amount of weight on his own initial opinion in forming his opinion in every period. Proposition 2 also shows that when $m_i = 0$, the signal initially received by i has no influence on players' long-run opinions. A detailed proof is in the Appendix.

Before providing some intuition for the proof, let us consider a two-player example where we set $m_2 = 1$ and $m_1 = 0$. Then player 2 always keeps the same opinion ($y_2^t = x_2$ for all t) and

$$\begin{aligned}\overline{y}_1^t &= \gamma_1(x_2 + \xi_1) + (1-\gamma_1)\overline{y}_1^{t-1} + \\ &= (x_2 + \xi_1)\gamma_1(1 + (1-\gamma_1) + ... + (1-\gamma_1)^{k-1}) + (1-\gamma_1)^k \overline{y}_1^{t-k}\end{aligned}$$

[22] This is because when $\gamma_i > 0$ for all i, $B = (B_i)_i$ is an irreducible probability matrix.

implying that $\overline{y}_1^t$ converges to $x_2 + \xi_1$ as t grows large, independently of player 1's initial opinion. Player 2 serves as an anchor that prevents agent 1's opinion from drifting. Long-run opinions however only incorporate player 2's initial opinion.[23]

The general argument for convergence runs as follows. For any fixed x, ξ, the expected opinion evolves according to

$$\overline{y}^t = \Gamma X + B\overline{y}^{t-1} \text{ with } B = I - \Gamma + \Gamma(I - M)A$$

where $X_i = m_i x_i + (1 - m_i)\xi_i$, Γ and M are diagonal matrices with $\Gamma_{ii} = \gamma_i$ and $M_{ii} = m_i$. When $m_{i_0} > 0$ for some i_0, proving convergence is standard[24] and the limit expected opinion $\overline{y}$ is the unique solution of

$$\overline{y} = X + (I - M)A\overline{y} \tag{3}$$

Next, defining $\eta^t = y^t - \overline{y}^t$ and $w_{ij}^t = E\eta_i^t\eta_j^t$, we have

$$\eta_i^t = (1 - m_i)\gamma_i\nu_i^t + B_i\eta^{t-1}$$

implying an expression for the evolution of the covariance vector $w^t = (w_{ij}^t)$ of the form

$$w^t = \Lambda + \overline{B}w^{t-1},$$

where $\overline{B}_{ij}$ is the row vector $(\overline{B}_{ij,hk})_{hk}$ with $\overline{B}_{ij,hk} = B_{ih}B_{jk}$ and Λ is the column vector with $\Lambda_{ii} = (1 - m_i)^2\gamma_i^2\varpi_0$. Proving convergence to the solution of

$$w = \Lambda + \overline{B}w \tag{4}$$

is also standard.[25]

One immediate corollary of Equations (3) and (4) is that the loss terms L_i^0 do not depend on γ or on the magnitude of the idiosyncratic component ϖ_0, while the loss terms V_i are proportional to the idiosyncratic component

[23] More generally, up to noise terms, long-run opinions are determined by the opinions of agents for which $m_i > 0$.

[24] The key to convergence is whether $\sum_j B_{ij} < 1$ for all i. When this is the case, we say that B has the *contraction property*. When $m_i > 0$ for all i, this property trivially holds: $\sum_j B_{ij} = (1 - \gamma_i) + \gamma_i(1 - m_i)\sum_j A_{ij} < 1$ for all i. When $m_i > 0$ for only some players, we use the fact that the network is connected to conclude that for some large enough K, $C = B^K$ has the contraction property: with large enough K, then for any i, there are paths of length K that go through i_0 for which $m_{i_0} > 0$.

[25] This is by the same logic as Footnote 24. Among all the K-step paths that start in ij, there is at least one that goes through i_0k for some k implying that $\overline{B}^K$ has the contraction property.

ϖ_0 (and equal to 0 when $\varpi_0 = 0$). This also implies that when $\varpi_0 = 0$, the parameters γ have no effect on L_i, and we can focus on the weights m and the analysis of the rule SFJ.

3.3 Fragility under low m.

Although convergence is guaranteed when at least one player does not use DG, there is no discontinuity at the limit where *all* m_i get small: long-run opinions then become highly sensitive to the permanent component of the noise, and the variance induced by the idiosyncratic errors becomes very high. Formally, we have:

Proposition 3: Let $\overline{m} = \max m_i$. Then $L_i^0 \geq \frac{\varpi}{n} \frac{(1-\overline{m})^2}{\overline{m}^2}$ and $V_i \geq \frac{\varpi_0}{2n} \frac{\underline{\gamma}^2 (1-\overline{m})^2}{\overline{m}}$.

The proof is in the Appendix. The lower bound on V_i is obtained as a simple extension of the proof of Proposition 1. We provide here the key step enabling us to obtain the lower bound on L_i^0, as it highlights interesting properties of the FJ process.

The lower bound on L_i^0 is obtained by showing that for given x, ξ, long-run expected opinions are weighted average of *modified initial opinions*, defined, whenever $m_i > 0$, as

$$\widetilde{x}_i = x_i + (1 - m_i)\xi_i / m_i.$$

When $m_i > 0$ for all i, one can write (using previous notations) $X = M\widetilde{x}$, and (3) implies that each y_i is an average over modified initial opinions:

$$\overline{y} = M\widetilde{x} + (I - M)AM\widetilde{x} + ((I - M)A)^2 M\widetilde{x} + ... = P\widetilde{x} \qquad (5)$$

where $P = (I - (I - M)A)^{-1}M$ is a probability matrix (see Lemma 4 in appendix). Intuitively, in each period, $\widetilde{x}_i$ can be though of as the effective seed for individual i, and all long-run opinions are averages over effective seeds. For a fixed x_i, the variance of each $\widetilde{x}_i$ induced by the persistent component is bounded below by $\frac{\varpi(1-\overline{m})^2}{\overline{m}^2}$, so we obtain the desired lower bound.

The argument can be generalized to the case where a subset N^0 of agents follows DG ($m_i = 0$). Then long-run opinions become linear combinations of modified opinions of the agents *not in* N^0, and these modified opinions are

$$\widehat{x}_i = \widetilde{x}_i + (1 - m_i)R_i\xi^0 / m_i$$

where R_i is a positive vector that only depends on the structure of the network which captures the influence of agents in N^0 on i (see Lemma 5). In other words, long-run expected opinions only depend on the seeds $\widehat{x}_i$ that *agents not using DG* incorporate into their own opinions. Our conclusion regarding L_i^0 extends to this case.

The two-player case. With two players, assuming m_1 and m_2 strictly positive and $\gamma_1 = \gamma_2 = 1$ (both use SFJ), the model can be solved by directly substituting y_2^{t-2}, then y_1^{t-2}, and so on. Letting $\rho = (1 - m_1)(1 - m_2)$, we have:

$$\begin{aligned} y_1^t &= m_1\widetilde{x}_1 + (1 - m_1)\nu_1^t + (1 - m_1)y_2^{t-1} \\ &= m_1\widetilde{x}_1 + (1 - m_1)m_2\widetilde{x}_2 + (1 - m_1)\nu_1^t + \rho\nu_2^{t-1} + \rho y_1^{t-2} \end{aligned}$$

which further implies:

$$y_1^t = \sum_{0}^{K-1} \rho^k(m_1\widetilde{x}_1+(1-m_1)m_2\widetilde{x}_2+(1-m_1)\nu_1^{t-2k}+\rho\nu_2^{t-2k-1})+\rho^K y_1^{t-2K} \quad (6)$$

which in turn gives us (7) and (8) below for the limits $\overline{y}_1$ and V_1:

$$\overline{y}_1 = p_1\widetilde{x}_1 + (1 - p_1)\widetilde{x}_2 \text{ with } p_1 = m_1/(m_1 + (1 - m_1)m_2). \quad (7)$$

$$V_1 = \varpi_0 \frac{(1 - m_1)^2 + \rho^2}{1 - \rho^2} \quad (8)$$

This example confirms that $p_1 = 0$ when $m_1 = 0$ and it illustrates that when both m_1 and m_2 get close to 0, $1 - \rho \simeq m_1 + m_2$, and the variance of opinion V_1 induced by the *idiosyncratic* noise gets arbitrarily high, approximately equal to $\varpi_0/(m_1 + m_2)$.

3.4 Comments

(a) **On anchoring, influence and consensus:** DG and FJ generate a very different dynamic of opinions. Permanently putting weight on one's initial opinion is equivalent to putting a weight on the opinion of an individual that never changes opinion: it anchors one's opinion, preventing too much drift. As a result, it also anchors the opinions of one's neighbors, hence, the opinions of everyone in the (connected) network.

The channel through which each player influences long-run opinions also differs substantially. In the absence of noise, and for a given network structure, relative influence in DG depends on relative speed of adjustment γ.

More precisely, let $\rho \in \Delta_n$ be the vector such that $\rho.A = \rho$. When the γ_i's are identical across players, long-run opinions all converge to $\rho.x$, so ρ_i defines i's influence as determined by the network structure. When the γ_i's differ, long-run opinions all converge to $\pi.x$ where $\pi \in \Delta_n$ and

$$\pi_i/\pi_k = (\rho_i/\gamma_i)/(\rho_k/\gamma_k), \tag{9}$$

which explains how both the network and speeds of adjustment determine influence.[26]

In contrast, under FJ, only the m_i's (and the structure of the network) affect the expected long-run opinions $\overline{y}$. The speeds of adjustment γ have no effect on expected long-run opinion, they only affect the variance induced by idiosyncratic noise.

Regarding influence under FJ, it can be shown that at the limit where all m_i's are very small, all long-run expected opinions are close to another and close to $p.\widetilde{x}$ where all p_i are proportional to $m_i\rho_i$, that is:

$$p_i/p_k = m_i\rho_i/(m_k\rho_k),$$

Thus, close to the limit, m_i plays the same role as $1/\gamma_i$ does in DG and consensus obtains. As the m_i's go up however, consensus disappears: players "agree to disagree".

(b) **On the fragility of DG:** There is something inherently fragile about the long-run evolution of opinions under DG. Since individuals don't put any weight on their own initial signal after the first period, the direct route for that signal to stay relevant is through the weight put on their own previous period's opinion. This source clearly has dwindling importance over time. This gets compensated by the growing weight on the indirect route–each individual i adjusts his or her opinion based on the opinions of their neighbors, and these are in turn influenced by i's past opinions and through those, by i's initial signal. In DG without transmission errors, the second force at least partly offsets the loss due to the force–but this is no longer true when there is any transmission error because of the cumulative effect of noise that comes with the feedback from others.

(c) **On the source of change in opinion**: One way to assess the difference between DG and SFJ is to express them in terms of changes of

[26] To see why (9) holds, observe that, up to a multiplicative constant, π is the unique solution of $\pi.B = \pi$ where $B = I - \Gamma + \Gamma A$. Now observe that $\pi = \rho\Gamma^{-1}$ is one such solution.

opinions and opinion spreads. Defining the change of opinion $Y_i^t = y_i^t - y_i^{t-1}$, the change in perception of neighbors' opinions $Z_i^t = z_i^t - z_i^{t-1}$, and the spread between own and neighbors' opinions $D_i^t = z_i^t - y_i^t$, we have the following expressions:

$$Y_i^t = \gamma_i D_i^{t-1} \tag{DG}$$

$$Y_i^t = (1 - m_i) Z_i^t \tag{SFJ}$$

Under DG, one changes one's opinion whenever there is a difference between that opinion and the opinions of one's neighbors: any difference generates an adjustment, which is why the evolution is so sensitive to transmission errors. Errors are eventually incorporated into the opinions of all the players, and repeated errors tend to cumulate and generate a general drift in opinions. The force towards consensus is too strong.

At the opposite extreme, under SFJ, players only incorporate *changes* in the opinions of others. So, in the case where the transmission error is always the same, ξ_1 will generate a *one time* change on 1's opinion, but it won't by itself generate any further changes for player 1. Of course, this initial change of opinion will trigger a sequence of further changes – it will be partially incorporated in player 2's opinion, and therefore come back to player 1 again. But, when $m_i > 0$ for at least one player, the knock-on effect will be smaller than the initial impact and will get even smaller over time. Hence over all it won't blow up. If all m_i are small however, these indirect effects are not dampened enough, and the consequence is a high sensitivity of the final opinion to the magnitude of the errors.

(d) **On talking and listening.** We have so far introduced noise in the reception of opinions. Other sources of noise are also plausible: for example, not everyone needs to express their opinions to their neighbors every period, potentially generating some randomness in the communication protocol. Or there may be randomness in whether expressed opinions are actually heard or processed. We will argue in Section 7.7 that these add to the fragility of DG but leave SFJ largely unaffected.

4 Choosing the rule

Equipped with these insights about the properties of different rules, we now return to the main question of this paper: what rule will people choose and what rule should they choose (and are these the same)? Recall that we already specified the objective function of any individual i, which is to minimize the loss function.

$$L_i = E(y_i - \theta)^2$$

To fix ideas it is worth starting with the case where there is no noise. In this case

$$y_i^t = (1-\gamma_i)y_i^{t-1} + \gamma_i(m_i x_i + (1-m_i)A_i y^{t-1})$$

As t becomes large, this system of equations must converge to limiting values of y_i, given by the system of equations

$$y_i = m_i x_i + (1-m_i)A_i y$$

for all i. This implies that

$$y_i = \sum_j b_{ij}x_j = \sum b_{ij}(\theta + \delta_j) = \theta + \sum_j b_{ij}\delta_j$$

where $\beta_{ij} \geq 0$ and $\sum_j \beta_{ij} = 1$. Therefore

$$L_i = E(\sum_j b_{ij}\delta_j)^2$$

Under our assumption that the variance of the δ_j's are identical, it is evident that across all possible probability vectors $(b_{ij})_j$, L_i is minimized by choosing $b_{ij} = \frac{1}{n}$ so that all the y_i would be equal to $\sum_j \frac{1}{n}x_j$. However b_{ij} are endogenously determined by the underlying rules that the players adopt, so there is no guarantee that these weights will be implemented as a result of a rule that results from individual choice. In particular, if all the y_i are going to be the same then they must all satisfy

$$y_i = m_i x_i + (1-m_i)y_i \text{ or equivalently, } m_i y_i = m_i x_i.$$

Since the realizations of x_i may differ, the equality of the y_i requires $m_i = 0$ for all i. There is thus no way to get to $b_{ij} = \frac{1}{n}$ unless $m_i = 0$ for all i. In other words, within the class of rules we consider, DG rules are the only ones that even offer the possibility of reaching the lowest feasible L_i.[27]

[27] In fact as observed in the seminal paper by De Marzo et al. (2003), for generic networks, for any finite n, even DG rules will not implement these weights, though for large n the outcomes generated by DG rules will approximately minimize L_i, for a large class of networks (Golub and Jackson (2010)).

However as soon as there is some noise, we already saw that the outcome generated by any DG rule drifts very far from minimizing L_i. The loss grows without bound. Indeed from the point of view of the individual decision maker it would be better to ignore everyone else than to follow DG. In fact all strategies that put too little weight on their own seed (recall DG puts zero weight) are dominated from the point of view of the individual decision-maker, as well as being socially suboptimal.

Proposition 4: *Let $\underline{m} = \varpi/(1+\varpi)$. Any (m_i, γ_i) with $m_i < \underline{m}$ is dominated by $(\underline{m}, \gamma_i)$, from the individual and social point of view.*

Regarding the choice of the individually optimal rule, Proposition 4 builds on two ideas. First, if all other players use DG, then for agent i, any $m_i > 0$ is preferable to DG because everyone's opinion drifts off indefinitely if $m_i = 0$, as we saw above. Second, if some players use FJ (with $m_j > 0$), then initial opinions of these players x_j (plus any persistent noise in their reception of the signal) totally determines the long run outcome and the seeds of all the players that use *DG* do not get any weight – they end up as pure followers. This is not desirable for the same reason why, in the absence of noise, the ideal rule puts strictly positive weight on all the seeds. Hence the lower bound on m_i.

To see why this is also true of the socially optimal rule, i.e. the rule that minimizes $\sum_i L_i$, we observe that when $m_i = 0$, the only effect of information transmission by i to his neighbors is to introduce i's perception errors into the network. When i raises m_i above 0, he raises the quality of the information he transmits, while reducing the damaging echo effect that low m_i generates.

The next Proposition provides further characterization of the privately optimal choice of m_i. To simplify exposition, we focus on the case where the idiosyncratic component is null ($\varpi_0 = 0$), so the outcome does not depend on γ. For the purpose of explaining how m_i is affected by own and others quality of initial signals and transmission errors, we allow here for heterogenous disturbances. Recall σ_k^2 is the variance of k's initial opinion and ϖ_k the variance of $k's$ persistent component. Also let $W_k = \sigma_k^2 + \varpi_k(\frac{1-m_k}{m_k})^2$. We have:

Proposition 5. *Player i's optimal choice m_i satisfies:*

$$\frac{m_i}{1-m_i} = \frac{\varpi_i + (1-\lambda_i)^2 \sum_{k \neq i} (r_{ik})^2 W_k}{\sigma_i^2 (1-\lambda_i)}$$

where λ_i and $r_i = (r_{ik})_{k \neq i} \in \Delta_{N-1}$ only depend on A and m_{-i}.

We see from this that m_i's response shifts up when the variance of his own signal (σ_i^2) goes down or that of anyone else (σ_k^2) goes up. It also shifts up when the variance of the error term goes up. It further implies that the best response is a continuous function (which we know maps into a compact set $[\underline{m}, 1]$), so existence of an equilibrium is guaranteed.

Finally, although we are unable to fully characterize the social optimum, the following Proposition clarifies the relationship between the private and social optima.

Proposition 6: *At any Nash equilibrium, any player would increase aggregate social welfare by increasing m_i further.*

The next subsection fully studies a simpler environment (two players), where we explore the equilibrium, the social optimum and the relation between the two in greater detail. We provide below some intuition for Proposition 6.

A convenient way to express player j's opinion is by seeing it as an average between sources different from i that are unmediated by i's opinion, and player i's opinion (see Lemma 6 in Appendix):

$$\overline{y}_j = (1 - \mu_{ji}) Q_j^i \widetilde{x}_{-i} + \mu_{ji} \overline{y}_i$$

where $\mu_{ji} \in (0, 1)$ and Q_j^i is a probability vector. Since we consider the influence of $\widetilde{x}_{-i}$ through channels that are unmediated by i, μ_{ji} and Q_j^i are independent of m_i.

The expression above permits us to separate the loss L_j into three terms:

$$L_j = (1 - \mu_{ji})^2 var(Q_j^i \widetilde{x}_{-i}) + \mu_{ji} L_i + 2(1 - \mu_{ji}) \mu_{ji} Cov(Q_j^i \widetilde{x}_{-i}, \overline{y}_i). \quad (10)$$

When m_i is raised above $i's$ private optimum, there is no effect on the first term because it only depends on sources other than i. There is a second-order effect on the second term (because we start at $i's$ private optimum). The last term is what creates a discrepancy between private and social incentives.

This last term captures the correlation between sources other i and the opinion of i. The correlation is positive for two reasons:

(i) Any report made by j incorporates $\widetilde{x}_j$ and eventually reaches i who also incorporates it in $\overline{y}_i$, and $\overline{y}_i$ in turn affects j ($\mu_{ji} > 0$). This is an *echo effect*.

(ii) Whenever there are paths that go from k to j without going through i and others that go from k to i without going through j, $\widetilde{x}_k$ contributes to

j's opinion through both the unmediated channel (so $Q^i_{jk} > 0$) and through i's opinion ($\mu_{ji} > 0$). This a *confounding effect*: when j hears $\overline{y}_i$, she/he cannot separate the seeds $\widetilde{x}_i$ and $\widetilde{x}_k$ (for $k \neq i$) which both contribute to $\overline{y}_i$.

When m_i increases, the influence of each $k \neq i$ on i's opinion is reduced, and the correlation between $\overline{y}_i$ and $\widetilde{x}_k$ is also reduced. Overall, starting at a Nash equilibrium, L_j goes down when m_i is raised.

5 Equilibrium and Efficiency in Simple networks.

To derive more specific insights into the relation between equilibrium and efficiency in the rule choice game, we now turn to two specific examples: a two-player network and a network in the form of large circle, both cases where closed-form expressions are easy to compute. We focus on persistent errors, unless mentioned otherwise.

5.1 Two-player case.

Social optimum. With two players we have

$$y_i = m_i \widetilde{x}_i + (1 - m_i) y_j, \tag{11}$$

which yields $y_i = p_i \widetilde{x}_i + (1 - p_i) \widetilde{x}_j$ with $p_i = m_i/(m_i + (1 - m_i)m_j)$ (see (7)). This further yields

$$L_1 = I(p_1) + (p_1)^2 \mathcal{X}(m_1) + (1 - p_1)^2 \mathcal{X}(m_2)$$

where $I(p) = p^2 + (1 - p)^2$ is the variance of long run opinion in the absence of transmission noise (minimized at $p = 1/2$),[28] and $\mathcal{X}(m) = \varpi \frac{(1-m)^2}{m^2}$ measures the variance of cumulated error term. The total social loss is $L_1 + L_2$.

It is easy to check that minimizing the social loss requires setting identical values for m_1 and m_2. When both players use the same rule ($m = m_1 = m_2$), the loss is:

$$L = I(\frac{1}{2-m})(1 + \mathcal{X}(m)) = \frac{1}{2}(1 + \frac{m^2}{(2-m)^2})(1 + \varpi \frac{(1-m)^2}{m^2})$$

Given that initial opinions are equally informative, optimal information aggregation in the absence of noise by both players (which amounts to minimizing $I(p)$) would require setting $p_i = p_j = 1/2$. This is not feasible, but

[28] This is under our assumption that the variance σ_i^2 is equal to 1 for all i.

if m is small enough, p_i and p_j are both close to 1/2 and $I(p)$ is potentially close to its minimum. When m goes up from close to zero, the contribution of initial opinions to long-run opinions become asymmetric ($p_i, p_j > 1/2$), and this pushes $I(p)$ up. Indeed as the expression for $I(\frac{1}{2-m})$ makes clear, the logic of standard information aggregation gives the players a joint incentive to reduce m. However as the second term in the expression for L, $(1 + \varpi\frac{(1-m)^2}{m^2})$, makes evident, there is also a cost to lowering m. When $\varpi > 0$ and m is small, communication errors are hugely amplified. Welfare is maximized for an m^{**} that optimally trades off these two effects and the socially efficient weight m^{**} (which minimizes L) can be significantly different from 0 even when ϖ is small (for $\varpi = 0.0001$, $m^{**} = 0.13$ and for $\varpi = 0.001$, $m^{**} = 0.21$).[29]

Nash Equilibrium. We now assume that individuals choose their rules non-cooperatively and see what this does to the choice of rules.

For very low m_2, player 1 should choose m_1 close to m_2 for information aggregation purposes, but this would generate very high cumulated error, and player 1 is better off ignoring player 2 (m_1 close 1). For higher m_2, information aggregation is the main issue, and getting p_1 close to 1/2 requires choosing $m_1 < m_2$. A similar best response curve obtains for player 2.

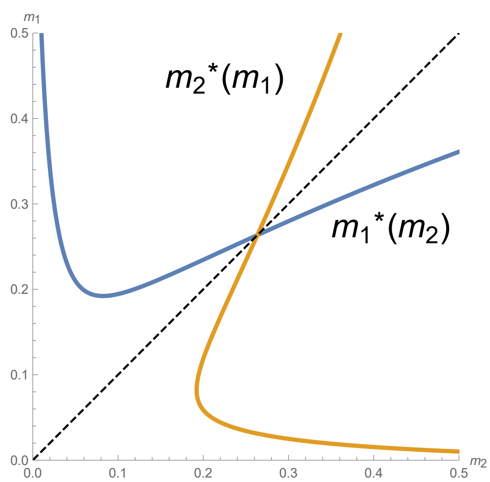

Figure 1: Best responses, $\varpi = 0.01$

The point where curves cross defines the equilibrium weights (m_1^*, m_2^*).

[29]For ϖ arbitrarily small, $2L \simeq 1 + m^2/4 + \varpi/m^2$, so $m^{**} \simeq (4\varpi)^{1/4}$.

Private versus social incentives. Computing the social optimum and Nash equilibrium values as a function of ϖ, we obtain the following Figure.

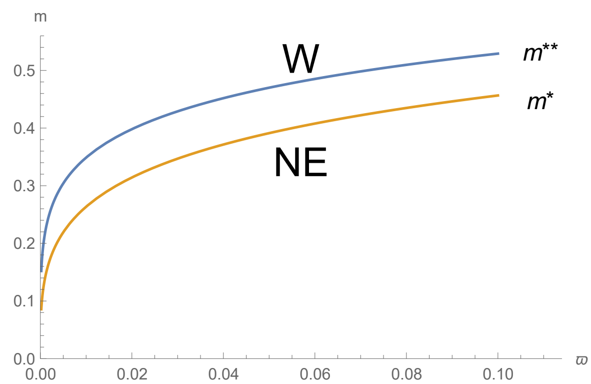

Figure 2: Equilibrium and socially optimal weights m^* and m^{**}

Equilibrium weights are below socially optimal weights, as expected. As explained earlier, the discrepancy between social and private incentives arises from the covariance term $Cov(\widetilde{x}_2 y_1)$ (see (10)) and the observation that this term strictly decreases with m_1, which can be checked directly as, using (11), we have:

$$Cov(\widetilde{x}_2 y_1) = \frac{(1-m_1)m_2}{m_1 + (1-m_1)m_2} var(\widetilde{x}_2).$$

Finally, we observe that, under SFJ, adding idiosyncratic noise to the persistent noise makes the incentive to increase m_i even stronger. Using (6) we get:[30]

$$L_1 = I(p_1) + ((p_1)^2 \mathcal{X}(m_1) + (1-p_1)^2 \mathcal{X}(m_2))(1 + \frac{1-\rho}{1+\rho}\frac{\varpi_0}{\omega}).$$

With idiosyncratic transmission errors, the variance of the error term is thus amplified proportionally. Efficient and Equilibrium weights on one's own signal both go up.

Information aggregation and welfare loss. The loss in welfare is significant relative to a benchmark case where players would observe (with transmission noise) the initial opinion of the other player(s) in the network and perfectly aggregate this (these) signal(s) with their own opinion. Under

[30]Recall $\rho = 1 - (m_1 + (1-m_1)m_2)$.

this benchmark, a persistent noise of magnitude $\varpi = 0.05$ would yields a total loss $L = 1.024$. Under FJ, the minimum loss rises up to 1.164. The equilibrium loss is even higher, 1.187. Figure 3 summarizes how welfare levels compare between the Nash equilibrium, the social optimum and the benchmark case as ϖ varies.

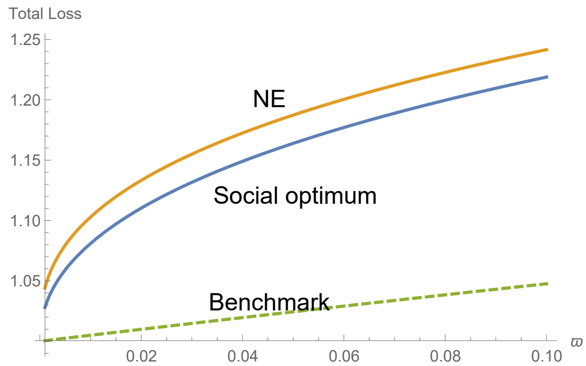

Figure 3: Losses

Even when transmission errors are small, equilibrium weights may be high, implying significant dispersion in opinions and welfare losses in the long-run.

5.2 Large circle case.

Social optimum Next we consider a large circle where information transmission is directed and one-sided: player i communicates to player $i+1$, who communicates to $i+2$, and so on.[31] Long-run opinions satisfy

$$y_i = m_i \widetilde{x}_i + (1 - m_i) y_{i-1}.$$

Hence if player i chooses m_i and all other players choose m, we have

$$y_i = m_i \widetilde{x}_i + (1 - m_i)(Z + (1 - m)^{n-1} y_i) \text{ where } Z = m \sum_{k=0}^{n-2} (1 - m)^k \widetilde{x}_{i-1-k}. \tag{12}$$

One can use this expression to derive y_i and its variance $J(m)$ when all choose the same m and n is set arbitrarily large:

$$J(m) = \frac{m}{2 - m}(1 + \mathcal{X}(m)),$$

[31] Player $n + 1$ coincides with player 1.

which is minimized at some m^{**} for which $J'(m^{**}) = 0$.[32] As in the two-player case, there is trade-off between improving information-aggregation (which calls for reducing all m_i close to 0) and reducing the amplification of communication errors (which calls for increasing all m_i).

Private and social incentives. We first check directly that private and social incentives coincide. One may use (12) to write:

$$L_i = (m_i)^2(1+\mathcal{X}(m_i))+(1-m_i)^2 J(m) = (m_i)^2+(1-m_i)^2(\varpi+J(m)). \quad (13)$$

At a symmetric Nash equilibrium m^*, private incentives require $m_i = (1 - m_i)(\varpi + J(m^*))$ with $m_i = m^*$. Since $J(m) = m^2 + (1 - m)^2(\varpi + J(m))$, this implies $J'(m^*) = 0$, so m^* is also a social optimum.[33]

To connect this result with the intuition provided earlier on the source of discrepancy between private and social incentives, consider $j = i+k$, that is, j is k communication steps away from i. We have

$$y_j = m\sum_{s=0}^{k-1}(1-m)^s\widetilde{x}_{j-s} + \mu_{ji}y_i \text{ where } \mu_{ji} = (1-m)^k$$

As explained earlier, the magnitude of the terms $\mu_{ji}Cov(\widetilde{x}_{j-s}, y_i)$ is key. When n is large, either k is large and $\mu_{ji} = (1 - m)^k$ is negligible, or $n - k$ is large and then $Cov(\widetilde{x}_{j-s}, y_i)$ is small (because $j - s$ is at least $n - k$ communication steps away from i). So as the circle gets very large, private and social incentives coincide.

Information aggregation and welfare loss. The loss in welfare is again significant relative to the benchmark case where players would observe (with transmission noise) the initial opinion of each other players in the network and perfectly aggregate these signals.[34] Under this benchmark, each player's loss would be close to 0.[35] Under FJ, the loss remains bounded away from

[32] m^{**} solves $\frac{m^2}{1-m} = \varpi$, which further implies $J(m^{**}) = m^{**}$.

[33] Note that the social optimum is symmetric: if $\underline{L}$ is the minimum loss that a player experiences at the social optimum, then $\underline{L} \geq \phi(\underline{L}) = \min_m (m)^2 + (1-m)^2(\varpi + \underline{L})$, which implies $\underline{L} \geq \lim_n \phi^n(0)$. $\phi(L) = (\varpi + L)/(1 + \varpi + L)$, so ϕ is a contraction and since $\phi(J(m^{**})) = J(m^{**})$, $\lim_n \phi^n(0) = J(m^{**})$.

[34] For i, this corresponds to getting the opinion $z_j = x_j + \varepsilon_i$ if j is a neighbor, and $z_k = x_k + \varepsilon_i + \varepsilon_j$ if k is not a neighbor of i but a neighbor of j, and so on.

[35] Even with transmission errors, the variance of the opinion of a neighbor at k steps is $v_k = 1 + k\varpi$. Since it grows linearly with k, the optimal weighting of these opinions would lead to an opinion with variance $(\sum 1/v_k)^{-1}$, which goes to 0 with k.

0: m^* is of the order of $\varpi^{1/2}$ when ϖ is very small, and $N = 1/J(m^*)$ is a measure of the quality of the aggregation of information: it represents the number of signals that are eventually aggregated into the information of each player. For example, with $\varpi = 0.05$, $m^* = 0.2$, $J(m^*) = 0.2$, and a player's long-run information is comparable to her having received only five independent signals (hence only four additional signals), out of the infinite pool that is available when the circle is arbitrarily large. We draw N as a function of ϖ.

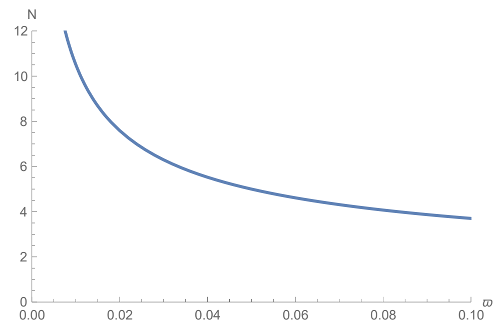

Figure 4: Information aggregation: large circle

6 Choosing among a richer class of rules

In Section 4, we examined incentives to modify the weight m_i. We now turn to the other sets of weights, the γ_i. A potential issue with FJ where γ_i is large is that long-run opinions are sensitive to idiosyncratic noise in transmission, and more generally to temporary changes in other's opinions. Choosing a lower γ_i slows down these reactions, hence opinions are only mildly affected by temporary shocks on perception and temporary variations in others' opinions. The next Proposition examines the effect of γ on the variance V_i induced by the idiosyncratic errors, as well as incentives for an individual to choose a low γ_i:

Proposition 7: *Fix $\underline{m}$. We have:*
(i)There exists c such that for any $\gamma > 0$ and $m \geq \underline{m}$, $V_i \leq c \max \gamma_j$.
(ii) For any $\gamma_{-i} > 0$, there exists c such that for all $m \geq \underline{m}$, $V_i \leq c\gamma_i$.

The proof is in Appendix B. Item (i) shows that when all γ_i are small, all V_i are small. Item (ii) shows that by choosing γ_i very small, a player can get rid of the additional variance induced by the idiosyncratic noise.

While the incentive is clear, a technical issue potentially arises if players wish to set γ_i arbitrarily small, as $\gamma_i > 0$ is an open interval.[36] We address this issue in the Appendix by showing that when all γ's are restricted to be above some lower bound $\underline{\gamma}$, any player i can secure a loss V_i no larger than $1/ \mid \log \underline{\gamma} \mid$ by choosing $\gamma_i = \underline{\gamma}$.[37] So if $\underline{\gamma}$ is small, V_i must be small in equilibrium. This also implies that investigating the properties of the game without idiosyncratic noise is a good enough approximation when $\underline{\gamma}$ is small.[38]

Also observe that the incentive to set γ_i arbitrarily small obviously depends on the assumption that players only care about long-run opinions. If players also cared about opinions at shorter horizons, then they would have incentives to increase γ_i to more quickly absorb information from the opinions of others: the trade-off is between increasing the rate of convergence (which is desirable when the relevant horizon is shorter) and increasing the variance induced by idiosyncratic noise (which is not desirable).

7 Extensions and interpretations

In this section we discuss extensions of and possible variations upon our base model, with the view to understand why different rules lead to different degrees of information aggregation in different settings. The general point is that long-run dispersion of opinion remains part of the answer and indeed there are reasons to expect that *adding the new elements* exacerbates this property.

7.1 Biased persistent noise

We have so far assumed that the persistent noise is drawn from a distribution that is mean zero. One can however imagine settings where it is more reasonable to assume that the persistent noise is biased, centered on ξ_i^0 for player i, for example because some individuals are biased in what they report (for whatever reason). That could for example be because they are truly biased and therefore try to sway opinion in the direction of their bias, or because they believe that others are biased and try to correct for it.

[36] Note that the limit opinion-formation process where γ_i tends to 0 is *not* the process where $\gamma_i = 0$ (under which no change in opinion would occur).

[37] $1/ \mid \log \underline{\gamma} \mid$ is small number when $\underline{\gamma}$ is small.

[38] In particular, if (m^*, γ^*) is an equilibrium of the game, then m^* is an $\varepsilon-$equilibrium of the game with no idiosyncratic noise, with ε comparable to $1/ \mid \log \underline{\gamma} \mid$.

In any case it makes sense to consider a variant of the updating rule FJ in which the agent can shift the opinion he incorporates by a constant c_i, so as to try to undo the systematic biases in his perception or perception of others:

$$y_i^t = (1-\gamma_i)y_i^{t-1} + \gamma_i(c_i + m_i x_i + (1-m_i)z_i^{t-1}) \qquad \text{(FJc)}$$

Suppose that in all other respects, the model is as before. For any (m, γ, c), this shift does not affect the variance of opinions resulting from idiosyncratic noise, but it shifts all long-run opinions. Regarding expected long-run opinions, theses shifts imply as before that $\overline{y}_i$ is a linear combination P_i of the modified opinions $\widetilde{x}_j$ where

$$\widetilde{x}_j = x_j + \frac{(1-m_j)\xi_j + c_j}{m_j}$$

The linear combination P_i is independent of c, so for any fixed (m, γ), each i can in principle set c_i to fully offset the systematic bias in transmission and this turns out to be optimal. [39] If c_i cannot be adjusted (e.g., $c_i = 0$ for all i), then the bias ξ_i^0 amounts to an increase in the variance of own error term ϖ_i, which, as Proposition 5 explains, generates further incentives to increase m_i: intuitively, when ξ_i^0 increases, opinions of others become a less accurate estimate of θ, and i prefers to put more weight on his/her own.

7.2 Heterogenous preferences

Assume that preferences of player i are quadratic (i.e., $u(a, \theta_i) = -(a-\theta_i)^2$) but vary in their relation to the common component θ:

$$\theta_i = \theta + b_i \qquad (14)$$

and that x_i is a noisy estimate of one's preferred point, that is,

$$x_i = \theta_i + \delta_i \qquad (15)$$

Define $Y_i^t = y_i^t - b_i$, $X_i = x_i - b_i$ and $\beta_i = (b_j - b_i)_j$. The "debiased" opinions Y_i^t evolve according

$$Y_i^t = (1-\gamma_i)Y_i^{t-1} + \gamma_i(c_i + m_i X_i + (1-m_i)(z_i^{t-1} + A_i\beta_i))$$

[39] Letting L_i^0 denote the loss when all ξ_i are centered on 0, and $C_i = (1-m_i)\xi_i^0 + c_i$. We have $L_i = L_i^0 + (P_i C)^2$, and L_i is minimized for $C_i = -\sum_{j\neq i} P_{i,j}C_j/P_{i,i}$. The equilibrium loss thus coincides with L_i^0, and $c_i = -(1-m_i)\xi_i$ for all i is an equilibrium.

and the problem becomes formally equivalent to the homogenous preference case with a persistent transmission term $A_i\beta_i$ added. If the biases b are fixed and if players can adjust c_i optimally, then like in the previous case, in equilibrium players can offset the bias by setting

$$c_i = -(1 - m_i)A_i\beta_i.$$

and the analysis is formally equivalent to the homogenous preference case, and the issue we raised (in particular, the fragility of long-run opinions to transmission errors) apply. In contrast, if players are unable to adjust c_i (e.g., $c_i = 0$), then the term $A_i\beta_i$ is akin to a systematic bias ξ_i^0, which, as explained in the previous subsection, generates incentives to further increase m_i.

Finally, consider the intermediate case where players can adjust c_i, but biases are not fixed and players can only adjust c_i on average across realizations of the β's. Said differently, across problems, there are variations in the heterogeneity, and players are unable to tune c_i to each realization of the heterogeneity. Then the problem is formally equivalent to one where preferences are homogenous and a persistent transmission term $A_i\beta_i$ is added.[40]

The general take-away should be that there are many potential sources of errors which will favor the choice of FJ over DG rules. To illustrate with one final example, assume that i misperceives other's preferences. He perceives $\widehat{\beta}_i$ instead of β_i and erroneously sets $c_i = -A_i\widehat{\beta}_i(1-m_i)$. Then the difference $(1 - m_i)A_i(\beta_i - \widehat{\beta}_i)$ is akin to an additional (independent) source of persistent bias/noise in transmission.

7.3 Other communication protocols

We have followed the standard approach to modeling communication in this literature, with each player communicating with all his neighbors at every date.[41] We now consider an extension where each round of communication is one-sided and, at any date t, each agent i only hears from a subset $N_i^t \subset N_i$ of his neighbors but there exists K such that each player hears from all his neighbors at least once every K periods.[42] Imperfect communication is

[40]One difference with the case examined in the basic model however is that the β_i's are correlated: with two players, $A_i\beta_i = b_j - b_i = -A_j\beta_j$. Nevertheless, so long as there still exists a persistent noise term ξ_i (with all ξ_i drawn independently of the β_i's), Proposition 3 applies, as for each realization of β, all $m_i < \underline{m}$ are dominated.

[41]Banerjee et al. (2019) introduce the idea of a Generalized DeGroot model where not everyone starts with a signal and therefore does not participate in the communication till they get a signal. They show that this partially weaken the "wisdom of crowds".

[42]That is, for all $t : \cup_{s=1,..,K} N_i^{t+s-1} = N_i$.

modeled as before, through the addition of an error term ε_i^t that slants what i hears. Together these give us

$$z_{i,j}^t = y_j^{t-1} + \varepsilon_i^t \text{ if } j \in N_i^t$$
$$z_{i,j}^t = z_{i,j}^{t-1} \text{ if } j \in N_i \backslash N_i^t$$

where $z_{i,j}^t$ is i's current perception of j's opinion, based on the last time he has heard from j. Player i uses these perceptions to construct an average over neighbor's opinions

$$z_i^t = A_i Z_i^t$$

where $Z_i^t = (z_{i,j}^t)_j$ is the vector of i's perceptions and A_i defines as before how i averages neighbors' opinions.[43] We continue to assume FJ updating.

For fixed x, ξ, define $\overline{y}_i^t = Ey_i^t$, $Y_i^t = (\overline{y}_i^{t-k})_{k=0,..,K}$, the column vector of i's past recent opinions, and $Y^t = (Y_i^t)_i$. One can write $Y^t = X + BY^{t-1}$. Y^t converges for standard reasons, to some uniquely defined Y. Consider now the vector $\overline{y}$ solution to

$$\overline{y}_i = m_i x_i + (1 - m_i) A_i (\overline{y} + \xi_i)$$

and let $\overline{Y}_i = (\overline{y}_i, ..., \overline{y}_i)$ and $\overline{Y} = (\overline{Y}_i)_i$. By construction, under this profile of opinions, it does not matter when i heard from j because opinions do not change. $\overline{Y}$ thus solves $Y = X + BY$ and it coincides with Y. The limit expected opinion vector under FJ is thus independent of the communication protocol.[44]

This robustness contrasts with what happens when players use DG rules. As we explain in the Appendix with a simple example, changes in the protocol and in particular, the frequencies with which players communicate amount to changes in the values of γ_i (when you hear less often from others, your opinion changes more slowly, effectively reducing γ_i), and even when communication is noiseless, changes in γ_i modify long-run opinions.

To illustrate this, consider the two-player case *with noiseless communication*. Under DG, for i and $j \neq i$, $y_i^t = (1 - \gamma_i) y_i^{t-1} + \gamma_i y_j^{t-1}$, so for any α, β such that $\beta\gamma_2 = \alpha\gamma_1$,

$$\begin{aligned} \alpha y_1^t + \beta y_2^t &= ((1 - \gamma_1)\alpha + \beta\gamma_2) y_1^{t-1} + ((1 - \gamma_2)\beta + \alpha\gamma_1) y_2^{t-1} \\ &= \alpha y_1^{t-1} + \beta y_2^{t-1} = \alpha x_1 + \beta x_2 \end{aligned}$$

[43] We abuse previous notations here, using the restriction of vector A_i to i's neighbors (A_i was previously defined over all players, with weight 0 on non-neighbors).

[44] So long as the condition in footnote 42 holds.

Since $y_1^t - y_2^t$ converge to 0,[45] y_i^t converges to the common long-run opinion

$$y = \frac{\alpha x_1 + \beta x_2}{\alpha + \beta} = \frac{\gamma_2 x_1 + \gamma_1 x_2}{\gamma_1 + \gamma_2}$$

Thus when γ_1/γ_2 rises, long run opinions get closer to player 2's initial opinion (see also (9) in Section 3.4).

Thus, even in the absence of transmission errors, variations in the communication protocol induce additional variance in long-run opinions which can be mitigated by the use of FJ rules by all players. Nevertheless, in the absence of transmission errors, long-run opinions under DG remain averages over initial opinions, so the fragility we have highlighted is not as severe: the variance induced by variations in the protocol remains bounded even when $m_i = 0$.

7.4 Uncertainty over the precision of initial signals.

We examine here another variation of the model, assuming that the precision of initial signals is a random variable. We will argue that in the absence of transmission errors, this type of shock does not affect the performance of DG and therefore, unlike in the case of the previous examples of variations on the classic DG setting, there is no incentive for players to use the instrument m_i.

Formally, assume that each the speed of adjustment γ_i as a linear function of the variance of signal, that is, $\gamma_i = \mu_i \sigma_i^2$. Then for well-suited coefficients $\mu^* = (\mu_i^*)_i$ information aggregation is perfect, which further implies that this particular μ^* is also a Nash Equilibrium of the game where each chooses μ_i.

To see why, recall that under DG, the consensual long-run opinion is a weighted average of initial opinions, with weights proportional to ρ_i/γ_i (see (9)). So if the μ_i's are proportional to ρ_i, the weights become proportional to ρ_i/γ_i, hence proportional to $1/\sigma_i^2$, implying that perfect aggregation obtains for each vector of realization $(\sigma_1, ..., \sigma_n)$.

7.5 Coarse communication

In the social learning literature, it is common to focus on cases where the choice problem is about whether action 1 or action 0 should be taken, and the

[45] The usual contraction argument works: $y_2^t - y_1^t = (1 - (\gamma_2 + \gamma_1))(y_2^{t-1} - y_1^{t-1})$ so $| y_2^t - y_1^t | \leq k \mid y_2^{t-1} - y_1^{t-1} \mid$ for $k < 1$

information being aggregated is which of the two is being recommended by others. Frick et al. (2019) explore this class of models under the assumption that the agents may have erroneous priors, which is related to our emphasis on the role of errors in transmission in social learning.

We introduce this into our model by assuming that $\theta_i = \theta + b_i$ characterizes i's preference (as in (14)) but that the optimal action a_i^* is 1 when $\theta_i > \theta^*$, 0 otherwise. Agent i knows the bias b_i and does not know θ perfectly. He has an initial opinion $x_i = \theta + \delta_i$ and aggregates opinions of others to sharpen his assessment of θ. Assume the b_i's are drawn from identical distribution F with full support on $\mathcal{R}$. Call $q = h(\theta)$ the fraction of agents that would choose $a = 0$ if their opinion was θ, and let $\phi(q) \equiv h^{-1}(q)$.[46]

A player with current opinion y_i^t about θ reports $a_i^t = 1$ to neighbors if $y_i^t + b_i > \theta^*$ and $a_i^t = 0$ otherwise. From a vector of reports, he computes the fraction f_i^t that report 0, and uses this as an input to make an inference about others' opinions. A plausible rule is

$$z_i^t = \phi(f_i^{t-1}) + \xi_i$$

where ξ_i is a persistent bias in making inferences.[47] He next incorporates z_i^t using FJ to generate a new opinion y_i^{t+1}.[48]

In the special case where the number of agents is large, each player hears from all other and all agents are subject to a perception bias $\xi_i = \xi > 0$, DG rules generate a dynamic that induces all players to report 1 independently of the state of the world.[49] Assume a fraction at most equal to $f > 0$ reports 0. Each makes an inference z_i at least equal to $\phi(f) + \xi$ regarding neighbors' opinions, so eventually, under DG, each player of type b_i may only report 0

[46] $h(\theta) = \Pr(\theta + b_i < \theta^*) = F(\theta^* - \theta)$. Choosing an F that arises from a density with full and unbounded support ensures that h is strictly decreasing from $\mathcal{R}$ to $(0, 1)$.

[47] In the terminology of Frick et al. (2018), ξ_i could stem from an erroneous prior $\widehat{F} \neq F$. Indeed, define $\widehat{h}(\theta) = \widehat{F}(\theta^* - \theta)$ and $\widehat{\phi} = \widehat{h}^{-1}$. If agents use $\widehat{\phi}$ to make inferences, i.e., $z_i^t = \widehat{\phi}(f_i^t)$, then the difference $\widehat{\phi}(f_i^t) - \phi(f_i^t)$ is a systematic bias in making inferences.

[48] Note that the loss function is no longer quadratic, but once one defines utilities $u(a, \theta, \theta_i)$, one can define loss functions, hence, further, express long-run expected losses as a function of the profile of updating rule. The optimal action given θ is $\sigma^*(\theta, \theta_i) = \arg\max u(a, \theta, \theta_i)$. Players form opinions y_i^t and report $a_i^t = \sigma^*(y_i^t, \theta_i)$. We assume that, eventually, if they take a decision at t, they mechanically use y_i^t and choose $a_i^t = \sigma^*(y_i^t, \theta_i)$.The loss associated with a decision taken at t is

$$L_i^t = E(u_i^*(\theta, \theta_i) - u(\sigma^*(y_i^t, \theta_i), \theta_i))$$

where $u_i^*(\theta, \theta_i) = u(\sigma^*(\theta, \theta_i), \theta_i)$, and agent i cares about minimizing the long-run expected loss $L_i = \lim_{t \nearrow \infty} L_i^t$.

[49] The assumption can be weakened to heterogenous biases ξ_i at least equal to $\xi > 0$.

if $b_i + \phi(f) + \xi < \theta^*$. Under the large number approximation, a fraction at most equal to $f' = h(\phi(f) + \xi) < f$ report 0, hence this fraction of agents reporting 0 eventually vanishes.

In contrast, under FJ with m_i sufficiently large, long-run opinions remain anchored on initial opinions, and opinions remain bounded (and correlated with the underlying state θ). For example, when all signals x_i coincide (say, $x_i = x = \theta + \delta$), the long-run opinion must solve:

$$h(m(x + \frac{(1-m)\xi}{m})) + (1-m)\phi(f)) = f$$

hence

$$\phi(f) = x + \frac{(1-m)\xi}{m}.$$

The trade-off is thus similar to the one in our basic model. Raising m reduces fragility with respect to transmission noise, dampening the echo term $\frac{(1-m)\xi}{m}$. However, it creates heterogeneity in agents beliefs when initial opinions x_i differ.[50]

Said differently, with agents who constantly seed in their own initial opinion x_i, the drift in opinions remains bounded. Information aggregation is not perfect because of the positive weight m_i, opinions remain dispersed in the long run, but they remain correlated with the underlying state.

Frick et al. (2019) obtain a fragility result similar to the one obtained above under DG. They consider players who naively apply Bayesian updating to their erroneous priors. Like DG, Bayesian updating incorporates a strong forces towards consensus, which eventually makes both processes (DG and Bayesian updating) fragile to errors.

FJ processes can be seen as a potential fix to the fragility of DG or Bayesian processes: by allowing for heterogenous opinions or beliefs and by triggering updates based on variations in others opinions (rather than discrepancies between others' and own opinions), they end up being more robust, not subject to this particular form of fragility.

[50]With a large number of neighbors, and heterogenous initial opinions, the long-run fraction f must solve:

$$Eh(m(\theta + \delta_i + \frac{(1-m)\xi}{m}) + (1-m)\phi(f)) = f$$

If h is locally linear around, this yields $\phi(f) = \theta + \frac{(1-m)\xi}{m}$: aggregation of private signals allows a perfect inference θ, up to the persistent echo effect $\frac{(1-m)\xi}{m}$.

7.6 Non-linear aggregation rules

We consider next an extension to non-linear updating rules. In DG, opinions adjust as a function of the spread between of own and others' opinions, and the adjustment is linear in the spread. We examine a two-player example where the adjustment is linear for player 1, and non-linear for player 2.

Formally, denote by $\Delta_i^t = y_j^t - y_i^t$ the spread of opinion between j and i, and assume that

$$y_i^t = y_i^{t-1} + \gamma_i(\phi_i(\Delta_i^{t-1}) + \varepsilon_i^t)$$

where

$$\phi_1(\Delta) = \Delta \text{ and } \phi_2(\Delta) = \Delta - d\rho(\Delta)$$

with $\rho(\Delta) = 1 - \exp -\Delta^2$. In other words, player 1 adopts the standard linear DG rule, while player 2 adopts a rule in the spirit of DG but less sensitive to bigger variations in Δ than a linear rule (for small Δ, $\phi_2(\Delta) \simeq \Delta - d\Delta^2$).

Choose α and β such that $\alpha + \beta = 1$ and $\alpha\gamma_1 = \beta\gamma_2$. Next define

$$Y^t = \alpha y_1^t + \beta y_2^t \text{ and } \Delta_1^t = y_2^t - y_1^t.$$

Letting $\varepsilon^t = \alpha\gamma_1\varepsilon_1^t + \beta\gamma_2\varepsilon_2^t$, we have:

$$\begin{aligned} Y^t &= Y^{t-1} + \alpha\gamma_1\phi_1(\Delta_1^{t-1}) + \beta\gamma_2\phi_2(-\Delta_1^{t-1}) + \varepsilon^t \\ &= Y^{t-1} - d\beta\gamma_2\rho(\Delta_1^{t-1}) + \varepsilon^t \end{aligned}$$

In other words, when both players use the linear DG, Y^t is a random walk. When players do not both use the linear DG rule, and one player uses an adjustment that is more conservative for large spread, then Y^t is a random walk with a negative drift. The drift is determined by $\rho(\Delta_1^{t-1})$, so it is vanishing if Δ_1^t tends to 0, but, for any t, Δ^t is actually bounded away from 0 with positive probability,[51] which implies that Y^t diverge.

7.7 Non-stationary weights.

The updating processes that we consider have stationary weights. Agents do not attempt to exploit the possibility that early reports possibly reveal

[51] Indeed, the evolution of Δ_1^t is determined by

$$\Delta_1^{t+1} = (1 - (\gamma_1 + \gamma_2))\Delta_1^t - d\beta\gamma_2\rho(\Delta_1^t) + \eta^t$$

where $\eta^t = \gamma_2\varepsilon_2^t - \gamma_1\varepsilon_1^t$. This implies that the spread Δ_1^t tends to revert to 0, but the noise term η^t keeps it up bounded away from 0 with positive probability. Hence the negative drift for Y^t.

more information than latter reports: later reports from neighbors may incorporate information that one has oneself transmitted to the network, and therefore should have lesser impact on own opinion.

As a matter of fact, with two players, one could imagine a process in which (i) player 1 combines the first report he gets with own opinion, yielding $y_1 = m_1 x_1 + (1 - m_1)(x_2 + \varepsilon)$, and then ignores any further reports from player 2; and (ii) player 2 follows DG. With m_1 set appropriately, such a process would permit player 1 to almost perfectly aggregate information and player 2 to benefit from that information aggregation performed by player 1.

There are however important issues with such time-dependent processes. In particular, it is not obvious how one extends these to larger networks since they require that each person knows his or her role in the network. They are also sensitive to the timing with which information gets transmitted or heard. With some randomness in the process of transmission, it could for example be that the first report y_2 that player 1 hears already incorporates player 1's own signal (because after a while y_2 starts being a mixture between x_2 and x_1), and as a result, player 1 should put more weight on the opinions of others. But of course, in events where $y_2 = x_2$, this increase in weight makes information aggregation worse.

To illustrate this strategic difficulty in a simple model with noisy transmission, assume that time is continuous, communication is one-sided (either 1->2 or 2->1), with each player getting opportunities to communicate at random dates. The processes generating such opportunities are assumed to be two independent Poisson process with (identical) parameter λ. Also assume that a report, once sent, gets to the other with probability p. Consider the time-dependant rule where each person communicates own current opinion, and current opinion coincides with their initial opinion if one has not received any report ($y_i = x_i$), and otherwise coincides with $y_i = m_i x_i + (1 - m_i) z_i^f$ where z_i^f is the perception of the first report received. Even if perceptions are almost correct (i.e. perceptions almost coincide with the other's current opinion), the noise induced by the communication channel generates uncertainty about who updates first, hence variance in the final opinion for all m_i. For example, in events where player 1 already sent a report and receives one from player 2, it matters whether player 2 received the report that 1 sent and incorporated it into her opinion, or whether player 2 failed to receive the report, in which case what player 1 gets is player 2's initial opinion.

In contrast, the time-independent FJ is not sensitive to that noise and achieves reasonably good information aggregation for many values of $m = m_1 = m_2$. FJ rules conveniently address a key issue in networks: whether what I hear already incorporates some of what I said.

8 Concluding remarks

We end the paper with a discussion of issues that we have not dealt with, and which may provide fruitful directions for future research.

One premise of our model is that everyone has a well-defined initial signal.[52]. However the analysis here would be essentially unchanged if some players did not have an initial opinion to feed the network and were thus setting $m_i = 0$ for the entire process. FJ would aggregate the initial opinions of those who have one.

In real life many of our opinions come from others and in ways that we are not necessarily aware of, and the existence of a well-defined "initial opinion" could be legitimately challenged. In other words, people may have a choice over the particular opinion they want to hold on to and refer back to (in other words, the one that gets the weight m_i).

To see why this might matter, consider a variation of our model where some players (N^{dg}) have initial opinions but use DG rule (or set m_i very low), while other agents (N^{fj}) have no initial opinions (or very unreliable ones). In this environment, there is a risk that the initial opinions of the *DG* players eventually disappear from the system, and soon be overwhelmed by noise in transmission. The FJ players could provide the system with the necessary memory, using the initial communication phase to build up an "initial opinion" based on the reports of their more knowledgeable DG neighbors, and then seed in perpetually that "initial opinion" into the network. In other words, in an environment where information is heterogeneous and weights m_i are set sub-optimally by some, there could be a value for some agent in adopting a more sophisticated strategy in which the "initial opinion" is temporarily updated until it becomes anchored. In other words, it may be optimal for some of the less informed to listen and not speak for a while as they build up their own "initial opinions" before joining the public conversation.

Another important assumption of our model is that the underlying state θ is fixed. In particular, there would be no reason to keep on seeding in the initial opinions if the underlying state drifts. However it may still be useful

[52]As mentioned earlier, Banerjee et al. (2019) introduce the idea of a Generalized DeGroot model where not everyone starts with a signal

to use a FJ type rules where the private seed is periodically updated by each player to reflect the private signals about θ that each one accumulates.

Finally, one interesting property of FJ type rules that we already emphasized is that one's opinions vary as a result of *variations* in others' opinions vary, rather than because of a difference between one's and others' opinions. In particular, players' opinions may differ in the long run. One could imagine applying a similar idea to beliefs about the state of world. With two states for example, one could let $y_i = \ln p_i/(1-p_i)$ measure the belief of i over the underlying state[53] and assume an updating process to y_i in the spirit of SFJ rule:

$$y_i^t - y_i^{t-1} = (1-m_i)(z_i^{t-1} - z_i^{t-2})$$

where

$$z_i^t - z_i^{t-2} = \frac{1}{N_i}\sum_{j\in N_i}(y_j^{t-1} - y_j^{t-2}) + \varepsilon_i^t$$

measures the perceived variations in others' opinions. This updating process allows for diverse beliefs in the population, and also transmission of information when some signals in the network induce variations in beliefs. These updating rules (with m_i set appropriately) could turn out be more robust than Bayesian rules when the updating process is subject to noise or biases.

References

Acemoglu, Daron, Munther A. Dahleh, Ilan Lobel, Asuman Ozdaglar (2011), "Bayesian Learning in Social Networks," *The Review of Economic Studies*, 78 (4), 1201–1236.

Alatas, Vivi, Abhijit Banerjee, Arun G. Chandrasekhar, Rema Hanna, and Benjamin A. Olken (2016), "Network Structure and the Aggregation of Information: Theory and Evidence from Indonesia." *The American Economic Review* 106, no. 7: 1663-704.

Aumann, Robert (1976), "Agreeing to disagree," *The Annals of Statistics*, 4(6):1236–1239.

Bala, Venkatesh and Sanjeev Goyal (1998), "Learning from Neighbours," *The Review of Economic Studies*, 65 (3), 595–621.

[53] This is in the spirit of log-linear learning rules that use the logarithm of likelihood ratios, as in Molavi et al. (2018).

Banerjee, Abhijit V. (1992) "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics*, 107 (3), 797–817.

Banerjee, Abhijit, Emily Breza, Arun G. Chandrasekhar and Markus Mobius, 2019. "Naive Learning with Uninformed Agents," NBER Working Papers 25497.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch (1992), "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy,* 100 (5), 992-1026.

Compte, Olivier and Andy Postlewaite (2018). Ignorance and Uncertainty (Econometric Society Monographs). Cambridge: Cambridge University Press.

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

DeMarzo, Peter M., Dimitri Vayanos, Jeffrey Zwiebel (2003), "Persuasion Bias, Social Influence, and Unidimensional Opinions," *The Quarterly Journal of Economics*, 118 (3), 2003, 909–968.

Dasaratha, Krishna, Benjamin Golub and Nir Hak (2020),"Learning from Neighbors About a Changing State," SSRN working paper n° 3097505.

Ellison, Glenn, and Drew Fudenberg (1993), "Rules of Thumb for Social Learning," *Journal of Political Economy* 101 (4), 1993, 612–43.

Ellison, Glenn, and Drew Fudenberg (1995), "Word-of-Mouth Communication and Social Learning," *The Quarterly Journal of Economics* 110 (1), 93–125.

Eyster, Erik, and Matthew Rabin (2010), "Naive Herding in Rich-Information Settings." *American Economic Journal: Microeconomics*, 2 (4), 221-43.

Friedkin, Noah E. and Eugene C. Johnsen (1990), "Social influence and opinions," *The Journal of Mathematical Sociology*, 15 (3-4), 193-206.

Friedkin, Noah E. and Eugene C. Johnsen (1999), "Social Influence Networks and Opinion Change," *Advances in Group Processes* 16: 1-29.(1999)

Gentzkow, Matthew, Michael Wong and Allen T. Zhang (2018), "Ideological Bias and Trust in Information Sources," Working paper.

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

Golub, Benjamin, and Matthew O. Jackson (2010), "Naive Learning in Social Networks and the Wisdom of Crowds." *American Economic Journal: Microeconomics*, 2 (1): 112-49.

Jackson, Matthew O., Suraj Malladi, and David McAdams (2019) "Learning through the Grapevine: the Impact of Message Mutation, Transmission Failure, and Deliberate Bias" Working paper

Levy, Gilat, and Ronny Razin (2015) "Correlation Neglect, Voting Behavior, and Information Aggregation." *The American Economic Review* 105 (4), 1634–645.

Mira Frick, Ryota Iijima, and Yuhta Ishii (2019), Misinterpreting Others and the Fragility of Social Learning, Cowles foundation paper n° 2160.

Molavi, P., Tahbaz-Salehi, A. and Jadbabaie, A. (2018), A Theory of Non-Bayesian Social Learning. *Econometrica*, 86: 445-490.

Mossel, Elchanan, Allan Sly, and Omer Tamuz (2015), "Strategic Learning and the Topology of Social Networks," *Econometrica* 83, no. 5: 1755-794.

Roberts, Ben, and Dirk P. Kroese (2007), "Estimating the Number of s-t Paths in a Graph," *Journal of Graph Algorithms and Applications,* 11 (1): 195–214.

Rosenberg, D., E. Solan, and N. Vieille (2009), "Informational Externalities and Emergence of Consensus," *Games and Economic Behavior*, 66 (2), 979–994.

Sethi, Rajiv, and Muhamet Yildiz (2012) "Public Disagreement," *American Economic Journal: Microeconomics*, 4 (3): 57-95.

Sethi, Rajiv, and Muhamet Yildiz (2016). "Communication with unknown perspectives," Econometrica 84 (6), 2029–69.

Sethi, Rajiv and Yildiz, Muhamet (2019), "Culture and Communication," SSRN working paper n° 3263560.

Appendix.

Notations. Define M as the $N \times N$ diagonal matrix where $M_{ii} = m_i$ (and $M_{ij} = 0$ for $j \neq i$). For any fixed vectors of signals x and systematic bias ξ, we let

$$X = Mx + (I - M)\xi$$

and, whenever $m_i > 0$, we let $\widetilde{x}_i = x_i + \xi_i(1 - m_i)/m_i$ denote the modified initial opinion.

Next define $B_{ij} = (1 - m_i)A_{ij}$ and the $N \times N$ matrix $B = (I - M)A$. Also define the (N^2) vector Λ with $\Lambda_{ij} = 0$ if $i \neq j$, $\Lambda_{ii} = (1 - m_i)^2 \varpi_0$ and $\overline{B}$ the $(N^2 \times N^2)$ matrix where $\overline{B}_{ij}$ is the row vector $(\overline{B}_{ij,hk})_{hk}$ with $\overline{B}_{ij,hk} = B_{ih} B_{jk}$.

For any fixed (x, ξ), we define the expected opinion at t, $\overline{y}_i^t = E y_i^t$ and the vector of expected opinions $\overline{y}^t = (\overline{y}_i^t)_i$. We further define $\eta^t = y^t - \overline{y}^t$, $w_{ij}^t = E \eta_i^t \eta_j^t$ and the vector of covariances $w^t = (w_{ij}^t)_{ij}$.

Finally, we shall say that P is a *probability matrix* if and only if $\sum_j P_{ij} = 1$ for all i. Note that A is a probability matrix.

Evolution of expected opinions and covariances. Under SFJ, the evolution of opinions and expected opinions (given x, ξ) follows

$$y^t = X + (I - M)\nu + B y^{t-1} \tag{16}$$

$$\overline{y}^t = X + B \overline{y}^{t-1}, \tag{17}$$

from which we obtain:

$$\eta^t = (I - M)\nu^t + B \eta^{t-1}$$

Since the ν_i^t are independent random variables, the evolution of the vector of covariances follows:

$$w^t = \Lambda + \overline{B} w^{t-1} \tag{18}$$

In the general case (FJ rather than SFJ), the evolution is defined similarly, with $X_i = \gamma_i(m_i x_i + (1 - m_i)\xi_i)$ and $B_{ij} = (1 - \gamma_i) I_{ij} + \gamma_i (1 - m_i) A_{ij}$ and $\Lambda_{ii} = (\gamma_i (1 - m_i))^2 \varpi_0$.

Paths. For any K, any $K-$sequence $q = (i_1, ..., i_K)$ and any probability matrix $D = (D_{ij})_{ij}$, we let $\pi^D(q) \equiv \prod_{k=1}^{K-1} D_{i_k, i_{k+1}}$, and for any set of sequences Q, we abuse notations and let $\pi^D(Q) = \sum_{q \in Q} \pi^D(q)$. We define a $K-$*path* as a $K-$sequence q for which $\pi^D(q) > 0$. For any i, j is a K-neighbor of i if there exists a $K-$path ending in j, and we denote by $\overline{N}_i^D$ the set of individuals that are $K-$neighbors of i for some K, under D.

Assumption 1: $\overline{N}_i^A = N$ *for all* i.

We denote by $Q_{i,j}^K$ the set of paths of length K from i to j, and Q_i^K the set of paths of length K that start from i. $Q_i^K = \cup_j Q_{i,j}^K$ and by construction, for any i, j

$$A_{ij}^K \equiv \pi^A(Q_{i,j}^K) \text{ and } \sum_{j \in N} A_{ij}^K = \pi^A(Q_i^K) = 1 \tag{19}$$

We also extend the notion of sequences and paths to pairs $ij \in N^2$ (rather than individuals). For any sequence of pairs $\overline{q} = (i_1 j_1, ..., i_K j_K)$ (or equivalently, any pair of sequences $\overline{q} = (q^1, q^2) = ((i_1, ..., i_K), (j_1, ..., j_K)))$ and any matrix $D = (d_{ij})_{ij}$, and we let $\overline{\pi}^D(\overline{q}) = \pi^D(q^1)\pi^D(q^2)$. We define a path $\overline{q}$ as a sequence such that $\overline{\pi}^A(\overline{q}) > 0$.

Proof of Proposition 1: Let y^t denote the vector of opinions at t. Let Δ_n be the set of vectors of non-negative weights $p = \{p_i\}_i$ with $\sum p_i = 1$. We have $y_i^t = B_i y^{t-1} + \gamma_i \varepsilon_i^t$ with $B_i \in \Delta_n$. So for any $p \in \Delta_n$, there exists $q \in \Delta_n$ such that:[54]

$$p.y^t = q.y^{t-1} + \sum_i p_i \gamma_i \varepsilon_i^t. \tag{20}$$

Define $\underline{V}^t = \min_{p \in \Delta_n} var(p.y^t)$. We have $V_i^t \geq \underline{V}^t$ and since $\gamma_i \geq \underline{\gamma}$ for all i, Equality (20) implies $\underline{V}^t \geq \underline{V}^{t-1} + \frac{1}{n}\underline{\gamma}^2 E(\varepsilon_i^t)^2$, hence the divergence.

Next let $\Gamma = (\gamma_i \xi_i)_i$. In matrix form, we have $\overline{y}^t = B\overline{y}^{t-1} + \Gamma$, which implies:

$$\overline{y}^t = \sum_k B^k \Gamma + B^t x$$

Since the network is connected, for some large enough k, B^k is a strictly positive probability matrix. Let π be the stationary distribution ($\pi B = \pi$). Consider a realization ξ such that $\pi.\xi \neq 0$, say $\pi.\xi > 0$. For k large enough, each row of B^k is close to π, implying that for k large enough, all $B^k\Gamma$ are positive and bounded away from 0, which proves the divergence of $\overline{y}^t$.■

Before proving Proposition 2, we start with two standard results.

Lemma 1: *Consider any non-negative matrix* $C = (c_{ij})_{ij}$ *such that* $\mu = \min_i(1 - \sum_j c_{ij}) > 0$. *Then* $I - C$ *has an inverse* $H \equiv \sum_{k \geq 0} C^k$, *and for any* X^0 *and* Y^0, $Y^t = X^0 + CY^{t-1}$ *converges to* HX^0.

Lemma 2: *Under Assumption 1, if* $m_{i_0} > 0$, *then for* K *large enough,* $C = B^K$ *and* $\overline{C} = \overline{B}^K$ *both satisfy the condition of Lemma 1, and* $I - B$ *and* $I - \overline{B}$ *have an inverse.*

[54] $B_{ii} = 1 - \gamma_i$ and $B_{ij} = \gamma_i A_{ij}$. $q_i = \sum_i p_j B_{ji} = p_i(1 - \gamma_i) + \sum_j \gamma_j p_j A_{ji}$.

Proof of Proposition 2: We iteratively substitute in (17) to get:

$$\overline{y}^t = X^0 + C\overline{y}^{t-K}$$

where $X^0 = DX$ with $D \equiv I + B + ... + B^{K-1}$, and $C = B^K$. By Lemma 2, Lemma 1 applies to C, so convergence of $\overline{y}^t$ to $\overline{y}$ is ensured, and $I - B$ has an inverse, which we denote H. We have $\overline{y} = HX$, hence the conclusion that $\overline{y}$ does not depend on x_i when $m_i = 0$ (since X does not depend on x_i when $m_i = 0$).

Regarding the covariance vector, we iteratively substitute in (18) to get

$$w^t = \Lambda^0 + \overline{C}w^{t-K}$$

where $\Lambda^0 = \overline{D}\Lambda$ with $\overline{D} = I + \overline{B} + ... + \overline{B}^{K-1}$ and $\overline{C} = \overline{B}^K$. By Lemma 2, Lemma 1 applies to $\overline{C}$, so convergence of w^t to w is ensured, and $I - \overline{B}$ has an inverse which we denote $\overline{H}$. We have $w = \overline{H}\Lambda$, which is thus independent of initial opinions.■

Before proving Proposition 3, we report standard results (Lemma 3 and Corollary 3 below) enabling us to show that long-opinions are weighted average of suitably defined modified opinions (Lemma 4 and 5). Let 1_N denote the column vector of dimension N for which all elements are equal to 1.

Lemma 3: *Let A^0 be a non-negative $N^0 \times N^0$ matrix and A^1 a non-negative $N^0 \times N^1$ matrix. Assume $I - A^0$ has an inverse and $A^0 1_{N^0} + A^1.1_{N^1} = 1_{N^0}$. Then $P = (I - A^0)^{-1}A^1$ is a $N^0 \times N^1$ probability matrix, i.e., $P1_{N_1} = 1_{N_0}$.*

We apply Lemma 3 to the case where $A^1 = M$ and $A^0 = B = (I - M)A$. By construction $A^0 1_N + A^1.1_N = 1_N$ holds, which gives the following immediate corollary:

Corollary 3: *Assume $m_{i_0} > 0$ and let $P = (I - B)^{-1}M$. Then P is a probability matrix.*

Lemma 4. *Assume $m_i > 0$ for all i. Then for each i, there exists* $P_i \in \Delta_n$ such that for all x, ξ, $\overline{y}_i = P_i.\widetilde{x}$.

Proof of Lemma 4: When $m_i > 0$ for all i, the condition of Proposition 2 applies. Let $H = (I - B)^{-1}$ and $P = HM$. (16) can be rewritten as:

$$\overline{y} = M\widetilde{x} + B\overline{y}$$

implying that $\overline{y} = P\widetilde{x}$ with $P = (I - B)^{-1}M$, and P is a probability matrix by Corollary 3.■

Lemma 4 can be generalized to the case where a subset N^0 of agents has $m_i = 0$. Call N^1 the set of agents with $m_i > 0$, and accordingly define the vectors of expected long-run opinions $\overline{y}^0$ and $\overline{y}^1$, and the vectors of persistent errors ξ^0 and ξ^1. We have:

Lemma 5. *Fix N^0. There exists R and Q (defined independently of m) such that, for any m, there exists a probability matrix P such that $\overline{y} = P\widehat{x} + Q\xi^0$ and $\widehat{x}_i = \widetilde{x}_i + (1 - m_i)R_i\xi^0/m_i$ for each $i \in N^1$.*

Proof of Proposition 3. The lower bound on L_i^0 follows immediately from Lemma 4 and 5. We focus here on loss V_i induced by the idiosyncratic shocks. Recall

$$\eta_i^t = \gamma_i(1 - m_i)\nu_i^t + (1 - \gamma_i)\eta_i^{t-1} + \gamma_i(1 - m_i)A_i\eta^t$$

This implies that for any $p \in \Delta_n$, there exists $q \in \Delta_n$ such that:

$$p.\eta^t = q.\eta^{t-1} + \sum_i \gamma_i(1 - m_i)p_i\nu_i^t \text{ and } \sum_i q_i \geq 1 - \underline{m} \tag{21}$$

Define $\underline{V}^t = \min_{p\in\Delta_n} var(p.\eta^t)$. Note that $V_i^t \geq \underline{V}^t$. Since $var(q.\eta^{t-1}) \geq (1 - \underline{m})^2\underline{V}^{t-1}$, Equality (21) implies $\underline{V}^t \geq (1 - \underline{m})^2\underline{V}^{t-1} + \frac{1}{n}\underline{\gamma}^2(1 - \underline{m})^2\varpi_0$, which yields the desired lower bound.■

Proof of Proposition 4. Let $\underline{m} = \varpi/(1 + \varpi)$. We show that DG and all strategies $m_i < \underline{m}$ are dominated by $\underline{m}$.

Assume first that all other players either use DG. Then, if player i uses DG as well, L_i^t diverges and by Corollary 2, for any $m_i > 0$, $y_i = \widehat{x}_i = x_i + (1 - m_i)(\xi_i + R_i\xi_{-i})/m_i$. The variance of y_i thus decreases strictly with m_i.

Now assume that at least one player j chooses $m_j > 0$. Then, long-run expected opinions converge to $\overline{y}$. Now define $Y_j^i \equiv E\overline{y}_jX_i$ and the N vector of covariances $Y^i \equiv (Y_j^i)_j$. Also define $\overline{Y}_{jk} = E\overline{y}_j\overline{y}_k$ and $\overline{Y}_{jk} = (\overline{Y}_{jk})_{jk}$ as the N^2 vector of covariances. From (17) we have:

$$Y^i = \Gamma^i + B^iY^i \text{ and}$$
$$\overline{Y} = \overline{\Gamma} + \overline{B}\,\overline{Y}$$

where $\Gamma_j^i = EX_jX_i$ and $\Gamma^i = (\Gamma_j^i)_j$, and $\overline{\Gamma}_{jk} = \Gamma_j^k + (1 - m_k)A_kY^j + (1 - m_j)A_jY^k$ and $\overline{\Gamma} = (\overline{\Gamma}_{jk})_{jk}$.

Given our independence assumptions, Γ^i has just one positive element, $\Gamma^i_i = EX_i^2$ and for any $m_i < \underline{m}$, Γ^i_i is strictly decreasing in m_i. Next observe that B^i only has non-negative elements, and that B^i is non-increasing in m_i. So Y^i is strictly decreasing in m_i for all $m_i < \underline{m}$. Applying the same argument to the vector $Y^k \equiv EyX_k$, and since $\Gamma^k = EXX_k$ does not vary with m_i, we obtain that Y^j is non-increasing in m_i.

It follows that all terms $\overline{\Gamma}_{jk}$ for $k \neq i$ and $j \neq i$ are non-increasing in m_i and all terms $\overline{\Gamma}_{ij}$ are strictly decreasing on the range $m_i < \underline{m}$. Since $\overline{B}$ is non-increasing in m_i, it follows again that for all $m_i < \underline{m}$, $\overline{Y}$ is non-increasing in m_i and $\overline{Y}_{ii}$ is strictly decreasing in m_i, and $\overline{Y}$ is non-increasing in m_i.

We now examine the effect of m_i on the vector of covariances w where $w_{jk} = \lim E(y^t_j - \overline{y}^t_j)(y^t_k - \overline{y}^t_k)$. Recall $w = \Lambda + \overline{B}w$. Since Λ and $\overline{B}$ are non-increasing in m_i and Λ_{ii} is strictly decreasing in m_i, w_{ii} strictly decreases with m_i, and w is non-increasing in m_i. Combining all steps, over the range $m_i < \underline{m}$, $L_i = \overline{Y}_{ii} + w_{ii}$ strictly decreases with m_i, and $\sum_k L_k$ also strictly decreases with m_i.■

From Lemma 4, $\overline{y}_i = P_i\widetilde{x}$ where P_i is the probability vector. We now characterize P_i and derive how each P_k varies with m_i. As a preliminary observation, we express $\overline{y}_{-i}$ as an average of $\widetilde{x}_{-i}$ and $\overline{y}_i$.

Lemma 6: *For each $k \neq i$, there exists μ_{ji} and a probability vector $Q^i_j \in \Delta_{N-1}$, each independent of m_i, such that*

$$\overline{y}_j = (1 - \mu_{ji})Q^i_j \widetilde{x}_{-i} + \mu_{ji}\overline{y}_i \tag{22}$$

The proof consists in using $\overline{y} = M\widetilde{x} + B\overline{y}$ to solve $\overline{y}_{-i}$ as an average over $\widetilde{x}_{-i}$ and $\overline{y}_i$. Since for each $\overline{y}_j$ with $j \neq i$, all coefficients in B_j are independent of m_i, the result follows. Details are in Appendix B. Note that Lemma 6 immediately implies

$$\frac{\partial P_j}{\partial m_i} = \mu_{ji}\frac{\partial P_i}{\partial m_i} \text{ and } P_j = (1 - \mu_{ji})\overline{Q}^i_j + \mu_{ji}P_i \tag{23}$$

where $\overline{Q}^i \in \Delta_N$ with $\overline{Q}^i_{jk} = Q^i_{jk}$ and $\overline{Q}^i_{ji} = 0$.

In the expression $\overline{y}_i = m_i\widetilde{x}_i + B_i\overline{y}$, we now substitute each $\overline{y}_k$ obtained in Lemma 6. This permits us to get an expression of P_i that makes explicit the dependence on m_i. Specifically, we have (see Appendix B):

Lemma 7: *There exists λ_i and $r_i = (r_{ik})_k \in \Delta_{N-1}$ that only depend on* A and m_{-i} such that

$$P_{ii} = \frac{m_i}{1 - \lambda_i + m_i\lambda_i} \text{ and } P_{ik} = \frac{(1 - m_i)(1 - \lambda_i)r_{ik}}{1 - \lambda_i + m_i\lambda_i} \tag{24}$$

We now express the loss L_i^0 as a function of P_i. Define $W_k = \sigma_k^2 + \varpi_k(\frac{1-m_k}{m_k})^2$, where σ_k^2 is the variance of k's initial opinion and ϖ_k the variance of $k's$ persistent component. We have:

$$L_i^0 = \sum_k (P_{ik})^2 W_k$$

Given the expression for P_i (see (24), optimization over m_i is simple, yielding Proposition 5 below. Next, Proposition 6 is also obtained as a simple corollary of (23) and (24). We prove here a more general version of Proposition 5, allowing here for heterogenous quality of signals and noise terms. That is, we let $W_k = \sigma_k^2 + \varpi_k(\frac{1-m_k}{m_k})^2$, where σ_k^2 is the variance of k's initial opinion and ϖ_k the variance of $k's$ persistent component. We have

Proposition 5: Let $c_i = \sum_{k \neq i}(r_{ik})^2 W_k$. *Player i's optimal choice m_i is uniquely defined and satisfies:*

$$\frac{m_i}{1 - m_i} = \frac{\varpi_i + (1 - \lambda_i)^2 c_i}{\sigma_i^2(1 - \lambda_i)}$$

Proof. Let $d_i = (\varpi_i + (1 - \lambda_i)^2 c_i)/(\sigma_i^2(1 - \lambda_i))$. Rewrite L_i^0 as

$$L_i^0 = \sigma_i^2(\frac{(m_i)^2 + (1 - m_i)^2(1 - \lambda_i)d_i}{(m_i + (1 - m_i)(1 - \lambda_i))^2})$$

Since d_i and λ_i do not depend on m_i, checking the first order condition yields $m_i/(1 - m_i) = d_i$, as desired.■

Proof of Proposition 6. We show that if $\frac{\partial L_i}{\partial m_i} = 0$, *then* $\frac{\partial L_j}{\partial m_i} < 0$ *for all* j. Using (23), and assuming no idiosyncratic noise, we rewrite

$$L_j = \sum_k ((1 - \mu_{ji})\overline{Q}^i_{jk} + \mu_{ji}P_{ik})^2 W_k$$

Since μ_{ji}, $\overline{Q}^i_{jk}$ and W_k are independent of m_i for $k \neq i$, and since $\overline{Q}^i_{ji} = 0$, we obtain:

$$\frac{\partial L_j}{\partial m_i} = \mu_{ji}(1-\mu_{ji}) \sum_{k \neq i} Q^i_{jk} \frac{\partial P_{ik}}{\partial m_i} W_k + (\mu_{ji})^2 \frac{\partial L_i}{\partial m_i}, \tag{25}$$

By Lemma 2, we further have $\frac{\partial P_{ik}}{\partial m_i} < 0$ *for all* $k \neq i$, which concludes the proof.■

Example with modified protocol of communication. We illustrate below how changing protocol amounts to changing the weights γ_i. We consider two players and assume that player 1 updates every period, while player 2 updates every other three periods. Then, at dates t where 2 updates, we have:

$$\begin{aligned}
y_1^t &= (1-\gamma_1)^3 y_1^{t-3} + (1-(1-\gamma_1)^3) y_2^{t-3} \\
y_2^t &= (1-\gamma_2) y_2^{t-3} + \gamma_2 y_1^{t-1} \\
&= (1-\gamma_2) y_2^{t-3} + \gamma_2((1-\gamma_1)^2 y_1^{t-3} + \gamma_2(1-(1-\gamma_1)^2) y_2^{t-3} \\
&= (1-\gamma_2(1-\gamma_1)^2) y_2^{t-3} + \gamma_2(1-\gamma_1)^2 y_1^{t-3}
\end{aligned}$$

So, the process evolves as if weights where $\gamma_1' = 1-(1-\gamma_1)^3 > \gamma_1$ and $\gamma_2' = \gamma_2(1-\gamma_1)^2 < \gamma_2$.

Appendix B (for on-line publication)

Proof of Lemma 1: Consider the matrix $H^t = (h^t_{ij})_{ij}$ defined recursively by $H^0 = I$ and $H^t = I + CH^{t-1}$. Let $z^t = \max_{ij} \left| h^t_{ij} - h^{t-1}_{ij} \right|$. We have $z^t \leq (1-\mu) z^{t-1}$, implying that H^t has a well-defined limit H, which satisfies $H \equiv \sum_{k\geq 0} C^k$. By construction, $(I-C)H = H(I-C) = I$, so $H = (I-C)^{-1}$. Similarly, defining $z^t = \max_i \left| Y^t_i - Y^{t-1}_i \right|$, we obtain that Y^t has a limit Y which satisfies $(I-C)Y = X^0$, implying $Y = HX^0$.■

Proof of Lemma 2: Call $Q_i^{K,i_0} \subset Q_i^K$ the set of paths of length K that start from i (to some j) and go through i_0. For any such path, $\pi^B(q) \leq (1-m_{i_0})\pi^A(q)$.[55] This implies

$$\sum_j C_{ij} \equiv \pi^B(Q_i^K) \leq (1-m_{i_0})\pi^A(Q_i^{K,i_0}) + \pi^A(Q_i^K \backslash Q_i^{K,i_0}) < 1$$

where the last inequality follows from (19) and Q_i^{K,i_0} non empty for K large enough.[56]

This implies that C satisfies the condition of Lemma 1, hence $I - C$ has an inverse. Let $D \equiv I + B + ... + B^{K-1}$ and $H = (I-C)^{-1}D$. We have

$$\sum_{k\geq 0} B^k = \sum_{k\geq 0} C^k D = H,$$

so $H(I-B) = (I-B)H = I$ and $I - B$ also has an inverse.

Regarding $\overline{C}$, the argument is similar. We work on paths $\overline{q}$ of pairs rather than paths q of individuals. Call $\overline{Q}_{ij}^K$ the set of paths $\overline{q} = (q^1, q^2)$ of length K that start from ij (to some hk), $\overline{Q}_i^{K,i_0}$ those for which q^1 goes through i_0. We have

$$\sum_{hk} \overline{C}_{ij,hk} \equiv \overline{\pi}^B(\overline{Q}_{ij}^K) \leq (1-m_{i_0})\overline{\pi}^A(\overline{Q}_i^{K,i_0}) + \overline{\pi}^A(\overline{Q}_i^K \backslash \overline{Q}_i^{K,i_0}) < 1$$

hence $\overline{C}$ satisfies the condition of Lemma 1, $I - \overline{C}$ has an inverse, and so does $I - \overline{B}$.■

Proof of Lemma 3: Let $q = P1_{N^1} - 1_{N^0}$. $P = A^1 + A^0 P$ so $P_{ij} = A^1_{ij} + \sum_{k\in N^0} A^0_{ik} P_{kj}$. Since $\sum_{j\in N^1} A^1_{ij} = 1 - \sum_{j\in N^0} A^0_{ij}$ we have

$$q_i = \sum_{j\in N^1} P_{ij} - 1 = \sum_{k\in N^0, j\in N^1} A^0_{ik} P_{kj} - \sum_{j\in N^0} A^0_{ij} = A^0_i q$$

[55] In the general case (FJ rather than SFJ), $\pi^B(q) \leq (1 - m_{i_0}\underline{\gamma})\pi^A(q)$.

[56] N is finite, so K can be chosen large enough that Q_i^{K,i_0} is non-empty for *all* i.

implying that $q = A^0 q$, hence, since $I - A^0$ *has an inverse,* $q = 0$.■

Proof of Lemma 5: Let $\widetilde{x}^1$ denote the vector of modified initial opinions of players in N^1, and M^1 the restriction of M to N^1. We have:

$$\overline{y}^0 = A^{00}\overline{y}^0 + A^{01}\overline{y}^1 + \xi^0 \tag{26}$$

$$\overline{y}^1 = M^1\widetilde{x}^1 + (I^1 - M^1)(A^{10}\overline{y}^0 + A^{11}\overline{y}^1) \tag{27}$$

Under A, for K large enough, all agents in N^0 have a K-neighbor in N^1, so $(A^{00})^K$ satisfies the condition of Lemma 1 and $I - A^{00}$ has an inverse, which we denote H^0. We thus have:

$$\overline{y}^0 = P^0\overline{y}^1 + H^0\xi^0 \tag{28}$$

where $P^0 \equiv H^0 A^{01}$ is a probability matrix (by Lemma 3 and because $A^{01}.1_{N^1} + A^{00}.1_{N^0} = 1_{N^0}$).[57]

Substituting $\overline{y}^0$ in (27), and letting $R = A^{10}H^0$ and $\widehat{x}_i = \widetilde{x}_i + (1 - m_i)R_i\xi^0/m_i$, we get

$$\overline{y}^1 = M^1\widehat{x} + (I^1 - M^1)\widehat{A}\overline{y}^1 \text{ where } \widehat{A} \equiv A^{11} + A^{10}P^0$$

Since P^0 is a probability matrix, so is $\widehat{A}$, and $C^1 = (I^1 - M^1)\widehat{A}$ therefore satisfies the condition of Lemma 1 (as all $m_i > 0$ for $i \in N^1$). Letting $H^1 = (I^1 - C^1)^{-1}$, we get $\overline{y}^1 = P^1\widehat{x}$ where $P^1 = H^1M^1$. Again, P^1 is a probability matrix because $\widehat{A}$ is a probability matrix and because $P^1 = M^1 + (I^1 - M^1)\widehat{A}P^1$. Substituting $\overline{y}^1$ in (28) we finally get $\overline{y}^0 = P^0P^1\widehat{x} + H^0\xi^0$ and $\overline{y}^1 = P^1\widehat{x}$, which concludes the proof.■

Proof of Lemma 6. Let $\widehat{m}_{ji} = 1-(1-m_j)(1-a_{ji})$. Define M^i and $\widehat{M^i}$ as $(N-1)\times(N-1)$ diagonal matrices where $\widehat{M^i}_{jj} = \widehat{m}_{ji}$ and $M^i_{jj} = m_j$. Let $\widehat{A^i}_{jk} = \frac{(1-m_j)a_{jk}}{1-\widehat{m}_{ji}}$ defined for all j, k different from i, and $g_{ji} = \frac{(1-m_j)a_{ji}}{\widehat{m}_{ji}}$. $\widehat{A^i}$ is a probability matrix. Also let $\widehat{X}_j = (1-g_{ji})\widetilde{x}_j + g_{ji}y_i$. By construction $y_j = \widehat{m}_{ji}\widehat{X}_j + (1-\widehat{m}_{ji})\sum_{k\neq i}\widehat{A^i}_{jk}y_k$, which in matrix form gives

$$y_{-i} = \widehat{M^i}\widehat{X}_{-i} + \widehat{B^i}y_{-i}$$

where $\widehat{B^i} = (I - \widehat{M^i})\widehat{A^i}$, which in turn yields $y_{-i} = R^i\widehat{X}_{-i}$ where $R^i \equiv (I - \widehat{B^i})^{-1}\widehat{M^i}$ is a probability matrix (by Lemma 1). Letting $\mu_{ji} = \sum_{k\neq i} R^i_{jk}g_{ki}$ and $Q^i_{jk} = R^i_{jk}(1-g_{ki})/(1-\mu_{ji})$, we obtain the desired expression for y_j,

[57] Indeed, for any $i \in N^0$, $\sum_{j\in N^0} A^{00}_{ij} + \sum_{j\in N^1} A^{01}_{ij} = \sum_{j\in N} A_{ij} = 1$

and Q^i is by construction a probability matrix. Note that $M^i, \widehat{M^i}$ and $\widehat{A^i}$ depend on A and m_{-i} only, so the same is true for Q^i and μ_{ji} for all j.$\blacksquare$

Proof of Lemma 7. Using $\overline{y}_j = P_j\widetilde{x}$ and (23) we get

$$\overline{y}_i = m_i\widetilde{x}_i + (1-m_i)\sum_{j\neq i} A_{ij}((1-\mu_{ji})\overline{Q}^i_j + \mu_{ji}P_i)\widetilde{x}$$

Letting $\lambda_i = \sum_j A_{ij}\mu_{ji}$, and $r_{ik} = \sum_{j\neq i} A_{ij}((1-\mu_{ji})Q^i_{jk}/(1-\lambda_i)$, and using $\overline{y}_i = \sum_k P_{ik}\widetilde{x}_{ik}$, we get the desired expressions.[58] Since μ_{ji} and Q^i depend only on A and m_{-i}, the same is true for λ_i and r_{ik}.$\blacksquare$

Proof of Proposition 7. *In addition to item (i) and (ii), we shall prove the following statement:* (iii) *If the lower bound $\underline{\gamma}$ on the choice set is sufficiently low and $\gamma_i = \underline{\gamma}$, $V_i \leq 1/\mid \log \underline{\gamma} \mid$ for all $m \geq \underline{m}$ and γ within the choice set.*

Let $\overline{\gamma} = \max \gamma_i$ and recall:

$$w_{ij} = \sum_{h,k} B_{ih}B_{jk}w_{hk} + \Lambda_{ij} \tag{29}$$

where $\Lambda_{ij} = 0$ if $i \neq j$ and $\Lambda_{ii} = (1-m_i)^2(\gamma_i)^2\varpi_0$, and $B_{ii} = 1-\gamma_i$, $B_{ij} = \gamma_i A_{ij}(1-m_i)$.

The proof starts by proving item (i), that is, computing a uniform upper bound on all w_{ij} of the form (see step 1)

$$w_{ij} \leq c\overline{\gamma} \tag{30}$$

To prove (ii), we define $\widehat{w} = (w_{ij})_j$ as the vector of co-variances involving i, and show that there exists a matrix C for which $\sum_k C_{jk} \leq 1$ for all j and such that

$$\widehat{w} \leq (1-\underline{m})C\widehat{w} + \Gamma \tag{31}$$

where $\Gamma_j \leq dp_{ij}$ for some d, with $p_{ij} = \gamma_i/(\gamma_i+\gamma_j)$. This in turn implies that $\max_j w_{ij} \leq \max_j \Gamma_i/\underline{m}$, which will prove (ii) (see step 3).

Finally, to prove (iii), we consider two cases. Either $\overline{\gamma}$ is "small" and (30) applies, or we can separate individuals into a subgroup J where all have a small γ_j, and the rest of them with significantly larger γ_j. In the later case, we redefine $\widehat{w} = (w_{jk})_{j\in J,k}$ as the vector of co-variances involving

[58] For example, focusing on the contribution of $\widetilde{x}_i$, and since $\overline{Q}^i_{ji} = 0$, we get $P_{ii} = m_i + (1-m_i)\sum_{j\neq i} A_{ij}\mu_{ji}P_{ii}$.

some $j \in J$, and obtain inequality (31) with $\Gamma_{jk} \leq dp_{jk}$ for $k \notin J$ and $\Gamma_{jk} \leq d\gamma_j$ for $k \in J$, for some d. By definition of J, all γ_j and p_{jk} are small, and all Γ_{jk} are thus small, which will prove (iii). Details are below.

Step 1 (item (i)) $w_{ij} \leq c\overline{\gamma}$ with $c = \varpi_0/\underline{m}$.

Let $\overline{V} = \max_i w_{ii}$ and $\overline{w} = \max_{i,j\neq i} w_{ij}$ and $\overline{w} = \max w_i$. For all $j \neq i$, w_{ij} is a weighted average between all $w_{h,k}$ and 0, so $w_{ij} < \max(\overline{w}, \overline{V})$, hence $\overline{w} < \max(\overline{w}, \overline{V})$, which thus implies $\overline{w} \leq \overline{V}$. Consider i that achieves $\overline{V}$. Since $\sum_{h,k} B_{ih}B_{ik} = (1-\gamma_i m_i)^2$, we have:

$$\overline{V} = w_{ii} \leq (1-\gamma_i m_i)^2 \overline{V} + \gamma_i^2 (1-m_i)^2 \varpi_0 \text{ hence}$$
$$\overline{V} \leq \frac{\gamma_i(1-m_i)^2}{m_i}\varpi_0 \leq \frac{\varpi_0\overline{\gamma}}{\underline{m}}$$

Step 2. Let $p_{ij} = \gamma_i/(\gamma_i+\gamma_j)$ and $\overline{v} = 2(c\overline{\gamma}+\omega_0)$. We have:

$$w_{ii} \leq \gamma_i p_{ii}\overline{v} + (1-\underline{m})\sum_k A_{ik}w_{ik} \tag{32}$$
$$w_{ij} \leq \gamma_j p_{ij}\overline{v} + (1-\underline{m})(p_{ij}\sum_k A_{ik}w_{kj} + p_{ji}\sum_k A_{jk}w_{ik}) \tag{33}$$

These inequalities are obtained by solving for w_{ij} in equation (29), that is, we write

$$(1-B_{ii}B_{jj})w_{ij} = \Gamma_{ij} + \sum_{k\neq i} B_{ii}B_{jk}w_{ik} + \sum_{k\neq i} B_{jj}B_{ik}w_{kj} + \sum_{k\neq i, h\neq j} B_{jk}B_{ih}w_{kj}.$$

Observing that $2B_{ii}B_{ik}/(1-B_{ii}B_{jj}) \leq (1-m_i)A_{jk}$, $B_{ii}B_{jk}/(1-B_{ii}B_{jj}) \leq (1-m_j)p_{ji}A_{jk}$, and $B_{jk}B_{ih}/(1-B_{ii}B_{jj}) \leq 2\gamma_j p_{ij}A_{jk}A_{ih}$ and $\Gamma_{ii}/(1-B_{ii}B_{jj}) \leq \gamma_i\omega_0$ yields (32-33).

Step 3 (item (ii))**.** It is immediate from (32-33) that (31) holds with $C_{jk} \equiv A_{jk}$ and $\Gamma_j = p_{ij}\gamma_j\overline{v} + p_{ij}c\overline{\gamma} \leq p_{ij}\overline{\gamma}(\overline{v}+c) \leq d\gamma_i$ for all j, for some d, which permits to conclude that $\widehat{w} \leq d\gamma_i/\underline{m}$.

Step 4 (item (iii))**.** Let $\varepsilon = \frac{1}{K|Log\underline{\gamma}|}$ with $K = 5\varpi_0/\underline{m}^2$ and set $\gamma_i = \underline{\gamma}$. Let us reorder individuals by increasing order of γ_j. Consider first the case where $\gamma_{j+1} \leq \gamma_j/\varepsilon$ for all $j = 1, ..., N-1$. Then $\overline{\gamma} < \underline{\gamma}/\varepsilon^{N-1}$, and for $\underline{\gamma}$ small enough, $\underline{\gamma}/\varepsilon^{N-1} < \varepsilon$, so $V_i \leq c\varepsilon < 1/\mid Log\underline{\gamma}\mid$.

Otherwise, there exists j_0 such that $\gamma_j \leq \underline{\gamma}/\varepsilon^{j_0-1}$ for all $j \in J$, and $\gamma_k > \gamma_j/\varepsilon$ for all $k \notin J$ and $j \in J$. It is immediate from (32-33) that (31) holds with Γ such that, for any $j \in J$,

$$\Gamma_{jk} = \gamma_j \overline{v} \text{ if } k \in J \text{ and}$$
$$\Gamma_{jk} = \gamma_j \overline{v} + p_{jk} \sum_{h \notin J} A_{jh} w_{hk} \text{ if } k \notin J$$

By definition of J, for all $j \in J$, $\gamma_j \leq \underline{\gamma}/\varepsilon^{N-1} < \varepsilon$ and for all $k \notin J$, $p_{jk} \leq \varepsilon$, which further that all Γ_{jk} are bounded by $\varepsilon(\overline{v} + c) \leq \underline{m}/ \mid Log\underline{\gamma} \mid$, which concludes the proof.■