A simple Bayesian heuristic for social learning and groupthink

Gilat Levy and Ronny Razin, LSE

Abstract: In this paper we analyze a simple Bayesian heuristic for learning from others’ posteriors and show its applicability to communication in groups and in social networks. The heuristic corresponds to rational Bayesian updating when individuals have conditionally independent information. When agents suffer from correlation neglect they also use the Heuristic. We show that communication in groups can lead to more polarized or less polarized beliefs for the group compared with those of the individuals. Applying the heuristic for social learning in networks, we show how consensus is in itself dynamic, and can shift as a result of repeated communication.

1 Introduction

In this paper we analyze a new heuristic of communication which is based on Bayesian updating. The heuristic is based on the assumption that individuals share a common prior, transmit the whole distribution of their belief to one another and update solely on this information. As we show, this heuristic is very useful to study communication in groups. First, if agents share a common prior and if they believe that their information sources are independent, the heuristic is consistent with Bayesian updating. The heuristic can also be extended to environments in which agents entertain the possibility of some correlation. Second, the heuristic is independent of the information structures that agents are exposed to, and thus allows for a more general analysis than models which take the information structures as primitive assumptions. Third, it can be used to better understand the effects of correlation neglect on group decision making. Finally, it is easy to work with in models of group communication in general and in social networks in particular.

In particular, consider a group, \( N = \{1, 2, ..., n\} \), who all share a common prior \( h(\omega) \) about a state of the world \( \omega \). Each individual is exposed to his own information structure, which generates information he uses to update his posterior, \( h^i(\omega) \), using Bayes Rule. Assume that individuals share their posterior distribution with one another, but not the details of their information structures. There exists a simple Bayesian updating rule which is solely a function of the common prior and the vector of posteriors, as long as individuals believe that their information structures provide conditionally independent signals. This rule results in a distribution function that is proportional to \( \prod_{i \in N} h^i(\omega) / (h(\omega))^{n-1} \). This formula is shown in Sobel (2014), but is also used in specific binary environments by Duffie and Manzo (2010) and Smith and Sorensen (2000) in a rational model, and by Levy and Razin (2014) and Eyster and Rabin (forthcoming) in behavioural models.

In this paper we study this heuristic and apply it to group decision problems. First we analyze the monotonicity properties of the heuristic. As the heuristic depends on distribution functions we need to reformulate this problem in terms of MLR shifts. We use the notion of Unimodal
Likelihood Ratio and show that shifts in individual posteriors translate to shifts of the group posterior.

We next analyze the robustness of the heuristic once we relax the assumption of conditional independence. We use the notion of point mutual information to formalise how bounds on conditional independence of information sources in the group translate directly to bounds on the group final posterior.

For individuals who are not aware of, or neglect, possible correlation in the information structures of the group members, we use the heuristic to derive results on groupthink and polarization as well as on social learning in networks. The literature on groupthink has illustrated how correlation neglect can lead to more polarization of the group opinion compared with the original individual opinions (e.g., Glaeser and Sunstein 2009). We use the Bayesian heuristic to show that this may depend on the original individual opinions and that if these are centered for example around one particular value (a unimodal distribution), the group can become more confident in this value and hence hold less polarized opinions.

When applied to networks, social learning becomes complicated, as individuals exchange information repeatedly with those they are connected with, and at some point, perhaps unknowingly, may receive information which they have already been exposed to. Again, this may lead to (dynamic) correlation neglect. Acemoglu et al (2014) have taken the route of full rationality in which every piece of information that goes through the system is tagged so individuals are not confused and can account for correlation between the different pieces of information and information redundancy. On the other hand, Jackson and Golub (2012) and De Marzo, Vayanos and Zwiebel (2003), among others, use the De Groot Heuristic (De Groot 1967) to consider environments in which individuals cannot fully account for correlation in the information they receive. In either case, the literature is concerned with whether a consensus is reached and how it depends on the structure of the network.1

We use the heuristic in a network environment. We show that the heuristic is simple to use and analyze in networks. In particular, one can use the Perron–Frobenius Theorem to calculate the limit of posteriors in strongly connected and aperiodic network matrices. Similarly to Jackson and Golub (2012) these posteriors converge to each other but have distinctive features. One key difference with the De Groot heuristic is that the Bayesian heuristic is based on rational foundations.2 This implies that by using the heuristic we can focus purely on the effects of correlation neglect without other implicitly imposed behavioral biases. Another difference is that in our model even if the group starts with the same beliefs, posteriors keep changing until the group converges to a degenerate distribution. In contrast, in the De Groot heuristic (in which people average beliefs of others), if a group starts with the same expectation about the state of the world, opinions don’t change anymore.

---

1See Mobius and Rosenblat (2014) for a recent survey of the social learning literature with a particular focus on networks.

2The De Groot one can be rationalized -in the initial phases of communication- for the particular information structure of the normal distribution.
2 The Model

A group of individuals, \( N = \{1, 2, ..., n\} \) share a common prior \( h(\omega) \) about a state of the world \( \omega \in \Omega \). We assume that \( h(\omega) \) is a (proper) density function. Each agent \( i \) is exposed to some information structure, \( (S^i, h^i(s, \omega)) \) where \( S \) is the set of possible signals and \( h^i(s, \omega) \) is the joint density on \( S \times \Omega \). Let \( \Lambda \) denote the set of feasible information structures \( (S^i, h^i(s, \omega)) \). We assume that for any \( i \neq j \), \( h^i(s, \omega) \) and \( h^j(s, \omega) \) are conditionally independent and that all agents know this. Each agent \( i \) knows \( (S^j, h^j(s, \omega)) \) but not \( (S^j, h^j(s, \omega)) \) for \( j \neq i \) (but he knows that \( (S^j, h^j(s, \omega)) \) is in \( \Lambda \)). Given an observation of some \( s \in S^i \) the agent has a posterior density \( h^i(w|s) \) where with some abuse of notation we will sometimes suppress \( s \) and denote by \( h^i(w) \).

We are interested in how an individual with a posterior \( h^i(w) \) updates his beliefs when he learns that the posteriors of \( j \neq i \) are \( h^j(w) \). That is, we are interested in solutions that do not depend on any beliefs on \( \Lambda \) that agent \( i \) might have about \( (S^j, h^j(s, \omega)) \), as long as they have some beliefs in this set. Below we will sometimes write \( h^i(s|\omega) \equiv \frac{h^i(s,\omega)}{h(\omega)} \). It is easy to show the following (see also Sobel 2014):

**Proposition 1:** After observing their opponent’s posteriors, all agents have the same posterior, with support \( \cap_{i \in N} \text{Supp}(h^i(\omega)) \) and a density given by,

\[
\eta(\omega) \propto \frac{\prod_{i \in N} h^i(\omega)}{(h(\omega))^{n-1}}
\]

\( \forall \omega \in \cap_{i \in N} \text{Supp}(h^i(\omega)) \).

**Proof:** We show that for any belief that agent \( i \) has in \( \Lambda \) about \( (S^j, h^j(s, \omega)) \), \( \eta(\omega) \) will be the posterior. Suppose \( i \) received a signal \( s_i \) and believes that \( j \) received a signal \( s_j \) from information structure \( (S^j, h^j(s, \omega)) \). He would then, using Bayes rule, update to

\[
\frac{h^i(\omega)h^j(s_j|\omega)}{\int h^i(v)h^j(s_j|v)dv} = \frac{h(\omega)h^i(s_i|\omega)h^j(s_j|\omega)}{\int h(v)h^i(s_i|v)h^j(s_j|v)dv}.
\]

But this is equivalent to the result stated as:

\[
\frac{h(\omega)h^i(s_i|\omega)h^j(s_j|\omega)}{\int h(v)h^i(s_i|v)h^j(s_j|v)dv} = \frac{\int h(\omega)h^i(s_i|\omega)h^j(s_j|\omega)\omega}{\int h(v)h^i(s_i|\omega)h^j(s_j|\omega)\omega}d\omega
\]

\[
= \frac{\int h(\omega)h^i(s_i|\omega)h^j(s_j|\omega)\omega}{\int h(v)h^i(s_i|\omega)h^j(s_j|\omega)\omega}d\omega
\]

\[
= \frac{h^i(\omega)h^j(\omega)}{h(\omega)} / \left( \int_{v \in \Omega} \frac{h^1(v)h^2(v)}{h(v)} dv \right).\]

In other words, for any conjectured information structure and a particular signal, the posterior would be equivalent.
The simple Bayesian posterior updating is attractive for applications relating to communication in groups. It is based on transmission of beliefs as opposed to signals.\(^3\) Moreover, it holds even if individuals have very little information about each others’ signal structure. Posteriors are not only sufficient statistics, but updating also does not require knowledge of the particular information structures of other players. The simple Bayesian rule above is belief-free in the sense that it holds for any beliefs that individuals may have about the information and signal structure of their group members. The only requirement -from a rational behaviour point of view- is that individuals believe that their own signals and those of their group members are conditionally independent and belong to the (general) set of possible information structures described above.

Finally, an attractive feature of the model is its simplicity. Combining posteriors in such a multiplicative way almost seems like a heuristics which individuals might follow even without any rational thinking.\(^4\) In the next Section we explore the applicability of this heuristic to environments in which agents are not rational. Beforehand, we show two attractive properties: monotonicity and robustness to correlated posteriors.

### 2.1 Monotonicity properties of the posterior

An attractive feature of this simple rule, is that the final posterior \(\eta(\omega)\) adopts the monotonicity properties of the individual posteriors. This is always the case with Bayesian updating. Here we simply formalise this to the case of learning from posteriors.

To this end, we use a weaker notion than the Monotone Likelihood Ratio. Consider a change in the posterior of agent \(i\) from \(h^i(\omega)\) to \(h^i(\omega')\). We say that \(h^i(\omega)\) and \(h^i(\omega')\) have a Unimodal Likelihood Ratio, denoted by \(h^i(\omega) \geq_{ULR} h^i(\omega')\), if \(E_{h^i(\omega)}(\omega) \geq E_{h^i(\omega')}(\omega)\) and there exists \(\hat{\omega}\) such that for all \(\omega\) and \(\omega'\) such that \(\omega' < \omega < \hat{\omega}\), or \(\hat{\omega} < \omega < \omega'\),

\[
\frac{h^i(\omega)}{h^i(\omega')} \leq \frac{h^i(\omega)}{h^i(\omega')}
\]

Note that the ULR relation implies second order stochastic dominance (See Hopkins and Korneinko 2003).

**Proposition 2:** If \(h^i \geq_{ULR} h^i\), then \(\eta' \geq_{ULR} \eta\).

**Proof:** Consider \(\omega\) and \(\omega'\) such that \(\omega' < \omega < \hat{\omega}\), or \(\hat{\omega} < \omega < \omega'\),

\[
\frac{\eta(\omega)}{\eta(\omega')} = \frac{h^i(\omega)h^2(\omega)}{h^i(\omega')} = \frac{h^i(\omega)h^2(\omega)}{h^i(\omega')} < \frac{h^i(\omega)h^2(\omega)}{h^i(\omega')}
\]

Thus, if one individual becomes more confident in the state \(\hat{\omega}\), the whole group becomes more confident in \(\hat{\omega}\). Trivially, if one individuals knows the truth, then he will convince all individuals.

---

\(^3\)To be sure, most credible communication is when individuals observe each other’s actions; this credibly reveals information about the other’s posterior. Repeated interaction and observations of actions would then reveal posteriors.

\(^4\)Note that if beliefs are transformed to log beliefs, as often used in applications in binary environments, then according to Bayesian updating, one should simply add log beliefs. See for example Duffie et al (2009) for the case of beliefs over a binary state space with signals that are Bernoulli trials.
2.2 Relaxing the independence assumption

In Proposition 1 we have assumed that the agents know that their information structures are conditionally independent. The result could extended if we relax this assumption.

To formalize this we use bounds on the pointwise mutual information which measures the degree of correlation between two sources of information. Let a \( f(x,y) \) be a joint probability distribution of two random variables \( x \) and \( y \), with marginal distributions \( f(x) \) and \( f(y) \). The point mutual information, or PMI at point \( (x, y) \) is \( \ln \frac{f(x,y)}{f(x)f(y)} \). Note that even though PMI may be negative or positive, or zero in the case of independent variables, its expected outcome over all joint events (i.e., the mutual information) is positive. Finally note that another way to express point mutual information is related to measures of entropy (see Cover and Thomas 1991).\(^5\)

Consider the same model as above, but relax the assumption about the independence between the two information structures. In particular assume that each individual believes that the PMI for any point \( s_i, s_j \), conditional on any \( \omega \), is bounded from below by \( a \leq 0 \) and from above by \( b \geq 0 \). Note that if \( a = b = 0 \) this is equivalent to independence. Let \( \alpha \equiv e^a \) and \( \beta \equiv e^b \).

**Proposition 3:** After observing agent \( j \)'s posterior, agent \( i \)'s updated density, \( \eta^i(\omega) \), satisfies

\[
\frac{\alpha}{\beta} \eta(\omega) \leq \eta^i(\omega) \leq \frac{\beta}{\alpha} \eta(\omega)
\]

\( \forall \omega \in \text{Supp}(h^1(\omega)) \cap \text{Supp}(h^2(\omega)). \)

**Proof:** Note that \( \alpha < \frac{h^i(s_i, s_j|\omega)}{h^i(s_i|\omega)h^i(s_j|\omega)} < \beta \) and therefore:

\[
\eta^i(\omega) = \frac{h^i(s_i, s_j|\omega)h(\omega)}{\int h(\omega)h^i(s_i, s_j|\omega)d\omega} \geq \frac{\alpha h^i(s_i|\omega)h^i(s_j|\omega)h(\omega)}{\int h(\omega)\beta h^i(s_i|\omega)h^i(s_j|\omega)d\omega} = \frac{\alpha}{\beta} \frac{h^i(s_i|\omega)h^i(s_j|\omega)h(\omega)}{h^i(s_i|\omega)h^i(s_j|\omega)h(\omega)} = \frac{\alpha}{\beta} \eta(\omega)
\]

\[
\eta^i(\omega) = \frac{h^i(s_i, s_j|\omega)h(\omega)}{\int h(\omega)h^i(s_i, s_j|\omega)d\omega} \leq \frac{\beta h^i(s_i|\omega)h^i(s_j|\omega)h(\omega)}{\int h(\omega)\alpha h^i(s_i|\omega)h^i(s_j|\omega)d\omega} = \frac{\beta}{\alpha} \frac{h^i(s_i|\omega)h^i(s_j|\omega)h(\omega)}{h^i(s_i|\omega)h^i(s_j|\omega)h(\omega)} = \frac{\beta}{\alpha} \eta(\omega).
\]

It is straightforward to extend the pointwise mutual information to \( n \) players.\(^6\)

3 Application I: Groupthink

Groupthink is the common term to describe failures of decision making at the group level, despite good information held by individuals that can be potentially aggregated towards a better decision.

We now consider the implications of using the Bayesian heuristic for groups of individuals who do not account for the correlation in their information structures (as in Glaeser and Sunstein

\(^5\)Another possibility is to use copulas and the Frechet-Hoeffding bounds (see Nelsen 2006). We find the point mutual information easier to apply.

\(^6\)Assume that individual believes that the mutual information for any point \( s_1, s_2, ..., s_n \), conditional on any \( \omega \), is bounded from below by \( a \leq 0 \) and from above by \( b \geq 0 \). In this case we can repeat the above in a straightforward manner.
thus, the support can only include the maximal points of
we have that,
arg max
h
i
= g
for all i. In other words, their information is, generically, fully correlated, but they are
not aware of it. This is the simplest exercise to start with.

For simplicity we assume also that all individuals share a uniform prior. Let
M
i
= \{\omega' | \omega' \in \arg \max_\omega g(\omega)\}. We consider the limit of \eta^n(\omega), after communication among \(n \to \infty\) individuals. It is easy to see that the support of the the limit distribution is a subset of \(M_g\), that is, the limit distribution is degenerate on a subset of the modes of \(g(\omega)\). To see why, note that for any two points \(\omega \) and \(\omega'\), we have that
\[
\frac{\eta^n(\omega)}{\eta^n(\omega')} = \left(\frac{g(\omega)}{g(\omega')}\right)^n \to \infty \text{ iff } g(\omega) > g(\omega') \text{ and otherwise to } 0
\]
thus, the support can only include the maximal points of \(g(\omega)\).

The communication in the group will result therefore in a limit posterior which accentuates
the modes of the original posterior.

The variance of the beliefs in the group may however be larger or smaller than that of the
individual prior to communication. We consider the following example to illustrate this:

**Example 1**: The Beta distribution: As an illustrating example, consider posteriors which
belong to the family of Beta distributions, i.e., individual \(i\) has a posterior,
\[
h^i(\omega) = \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} \omega^{\alpha_i-1}(1-\omega)^{\beta_i-1}.
\]

With this family it is easy to show that if individuals in a set \(N = \{1,2,\ldots,n\}\) are in the group
we have that,
\[
\eta(\omega) = \frac{\Gamma(\sum_{i\in N} \alpha_i + \sum_{i\in N} \beta_i - 2n + 2)}{\Gamma(\sum_{i\in N} \alpha_i - n + 1)\Gamma(\sum_{i\in N} \beta_i - n + 1)} \omega^{(\sum_{i\in N} \alpha_i - n + 1) - 1}(1-\omega)^{(\sum_{i\in N} \beta_i - n + 1) - 1},
\]
that is, \(\eta(\omega)\) is a beta distribution with parameters \(\sum_{i\in N} \alpha_i - n + 1\) and \(\sum_{i\in N} \beta_i - n + 1\)
when these are both positive numbers. Suppose then that \(\alpha_i = \beta_i = \alpha\) for all \(i \in N\). For any
distribution \(g\) let \(E_g\) and \(V_g\) be the expectation and variance of \(g\). Recall that for the a Beta
distribution \(g\) with parameters \(\alpha\) and \(\beta\) we have, \(E_g = \frac{\alpha}{\alpha + \beta}, V_g = \frac{\alpha \beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}\). We then have:

**Remark 1**: (i) \(E_g = E_g\); (ii) \(V_\eta < V_{h^i} \Leftrightarrow \alpha > 1\). (iii) If \(\alpha > 1\), as \(n\) grows large \(\eta\) converges
to a degenerate distribution on half. (iii) If \(\alpha < 1\), as \(n\) grows large \(\eta\) converges to a distribution
with support \(\{0,1\}\) and equal probabilities.

Note that if \(\alpha > 1\) the distribution \(h^i\) is unimodal and if \(\alpha < 1\) it is U-shaped. Intuitively, the
distribution \(\eta\) magnifies the modes of the distribution \(h^i\) so that if it is unimodal the result is a
more concentrated distribution with lower variance. If \(h^i\) is U-shaped, \(\eta\) will concentrate on the
two modes and increase the variance.

Thus, whether Groupthink results in beliefs with more or less variance than the individual ones
depends on the modality of the individual beliefs. In that sense, our analysis is closer to Sobel
who shows that in many environments there are few restrictions on the group’s decision coming from individual’s information and preferences. Here we show that these constraints can take a particular form. Glaeser and Sunstein (2009) on the other hand focus on larger polarization in groups which arises from correlation neglect, while here it may be the opposite, depending on the initial beliefs.

It is easy to extend this analysis to the case of heterogeneous posteriors. This can arise from a common prior and correlated information, non common priors, or independent information with common priors. Still, we continue to assume that agents behave as if they are rational and believe that their information is independent.

Assume then \( n \) individuals and a uniform prior. Assume that a share \( \alpha \) of the population has a prior \( h^1(\omega) \), and a share \( 1 - \alpha \) of the population has a prior \( h^2(\omega) \). Assume that both densities are bounded. Consider \( M = \{ \omega | \omega = \arg \max_{\omega'} h^1(\omega')^\alpha h^2(\omega')^{1-\alpha} \} \). Again we consider simultaneous (or public) group communication. We have:

**Proposition 4:** When \( n \to \infty \), the posterior of all individuals in the group is the same and is degenerate on a subset of \( M \).

Proof: To see that the posterior of the group will converge to the above, note that as above, for any finite \( n \), the group’s posterior will be

\[
\frac{h^1(\omega)^{\alpha n} h^2(\omega)^{(1-\alpha)n}}{\int h^1(v)^{\alpha n} h^2(v)^{(1-\alpha)n} dv},
\]

where for any \( \omega, \omega' \),

\[
\frac{\eta^n(\omega)}{\eta^n(\omega')} = \frac{h^1(\omega)^{\alpha n} h^2(\omega)^{(1-\alpha)n}}{h^1(\omega')^{\alpha n} h^2(\omega')^{(1-\alpha)n}} \to_{n \to \infty} \infty \text{ if } \omega \in M \text{ and } \omega' \notin M.
\]

What remains to show, is that there is convergence in distribution in the limit, which we show for the more general process considered in Theorem 1 below.

Finally note that if the initial posteriors are consistent with conditionally independent Bayesian posteriors, it is then also the case that the group converges to learn the truth, but not otherwise.

**Example 2:** The binary model: A concrete example to show the simplicity of the approach and its usefulness is the binary model.\(^7\) Suppose that the state of the world is either 0 or 1, and each player has beliefs \( h^i \in (0, 1) \) (degenerate on one point) that the state is 1. Suppose for example the prior is a half. Then, the posterior is simply \( \frac{h^i h_j}{h^i h_j + (1-h^i)(1-h_j)} \). It is easy to see that it satisfies attractive properties. First, it satisfies monotonicity as a player with beliefs above (below) a half will increase (decrease) the beliefs of all others. Also, it leads to group polarization. If all individuals in the group have beliefs above (below) a half, then the beliefs of the group will be higher (lower) than the beliefs of any individual. We explore the properties

\(^7\)See Eyster and Rabin 2014, Smith and Sorensen 2000, and Duffie and Manzo 2010, who use the Log of the beliefs.
of the binary framework in an application to segregation between groups in Levy and Razin (2014).

4 Application II: Social learning in networks

In this Section we consider how the Bayesian heuristic performs in networks. We will now show that computations of the limit beliefs in a network are related to the network matrix in a simple manner.

Consider a model in which \( n \) individuals are organised in a network. Let \( T \) denote the matrix of links where \( T_{ij} = 1 \) if there is a link between \( i \) and \( j \) and 0 otherwise. All individuals have a common uniform prior on the state \( \omega \in [0,1] \) and each holds an initial posterior \( h_0^j(\omega) \). We assume that the posteriors \( h_0^j(\omega) \) are full support, continuous and bounded. We assume that \( T \) is strongly connected (irreducible) and aperiodic.

The beliefs of individual \( i \) after the first round are

\[
h_i^1(\omega) = \frac{\prod_{j \in S_1(i)} h_0^j(\omega)}{\int \prod_{j \in S_1(i)} h_0^j(v) dv}
\]

where \( S_1(i) \) is the set of individuals that are linked to \( i \) directly.

And as we do above, easy to see that at any stage \( k \), beliefs are

\[
h_k^i(\omega) = \frac{\prod_{j \in S_k(i)} h_0^j(\omega)}{\int \prod_{j \in S_k(i)} h_0^j(v) dv}
\]

where \( S_k(i) \) is any \( j \) who is connected to \( i \) by a chain of \( k \) links. Note that the key observation here is that the order in which information was heard does not matter, as long as there is some connection so that information passes through.

The following Theorem characterises the limit beliefs:

**Theorem 1** (i) All individuals converge to the same posterior belief. (ii) There exists a vector \( \alpha = (\alpha_1, ..., \alpha_n) \) such that the limit posterior belief is degenerate on a subset of the modes of \( \Pi_{j \in N}(h_0^j(\omega))^{\alpha_j} \). (iii) The vector \( \alpha = (\alpha_1, ..., \alpha_n) \) is parallel to the Perron–Frobenius eigenvector.

Let us consider the following simple example just to illustrate the applicability of the approach and some of the arguments above.

**Example 3:** The star network: Consider three players. Player 1 is connected to both 2 and 3, while each of them is only connected to player 1 (a “star” network). Thus, the \( T \) matrix is (recall that every player is connected to herself as well):

\[
T = \begin{pmatrix}
1 & 1 & 1 \\
1 & 1 & 0 \\
1 & 0 & 1
\end{pmatrix}
\]
Suppose that the initial beliefs of the players are \( h_0^1, h_0^2 \) and \( h_0^3 \) for players 1, 2, 3 respectively. When communication is repeated infinitely, then the consensus in society converges to a distribution that puts equal weights on a subset of the modes of 
\[
\frac{h_1^i(\omega) \cdots h_1^j(\omega)}{\int h_0^i(v) \cdots h_0^j(v) dv}.
\]
This arises from simply calculating the eigenvectors of \( T \) which are \( v_1 = \sqrt{2} v_2, v_2 = v_3 \).

The example illustrates that the procedure is simple, and that the network structure matters, as when all players are connected, as shown above, the consensus will converge to a distribution that puts equal weights on a subset of the modes of 
\[
\frac{h_1^i(\omega) \cdots h_1^j(\omega)}{\int h_0^i(v) \cdots h_0^j(v) dv}.
\]
To see why, note that after the first period all players will have these posterior beliefs, and then we can apply Proposition 4. Note that the influence of the player depends here both on her position in the network (clearly player 1 has greater influence in that dimension), but also on the strength of her beliefs as a more confident player (with beliefs almost degenerate) will drag others in her direction.

**Proof of Theorem 1:** Note that the posterior of individual \( i \) at stage \( k \) is given by
\[
\frac{\Pi_i(h_0^i(\omega)) \sum_j T_{ij}^k}{\int \Pi_i(h_0^i) \sum_j T_{ij}^k}
\]
Where \( T^k = T \times T \times ... \times T \) \( k \) times.

If \( T \) is strongly connected (irreducible) and aperiodic, then it is primitive and we can use the Perron–Frobenius Theorem for primitive matrices. This theorem states:

1. There is a positive real number \( r > 1 \), which is an eigenvalue of \( T \) and any other eigenvalue \( \lambda \) satisfies \( |\lambda| < r \).
2. There exists a right eigenvector \( v \) of \( T \) with eigenvalue \( r \) such that all components of \( v \) are positive (respectively, there exists a positive left eigenvector \( w \)).
3. \( \lim T^k / r^k = P \) where \( P = vw^T \), where the left and right eigenvectors for \( T \) are normalized so that \( w^T v = 1 \). Moreover, the convergence is exponential.

Therefore we have that \( \sum_{i \in N} T_{ij}^k / r^k \to \sum_j P_{ji} \equiv \alpha_i \). We now show that this implies that all individuals’ beliefs converge to a degenerate distribution on a subset of the modes of \( \Pi j \in N(h_0^j(\omega))^{\alpha_i} \).

For simplicity assume that there is only one mode of \( \Pi j \in N(h_0^j(\omega))^{\alpha_i} \) and denote it by \( x^* \).

For any stage \( k \) let \( F_k(x) = \frac{\sum_j T_{ij}^k}{\int \Pi_i(h_0^i) \sum_j T_{ij}^k} \) be the cumulative distribution function of individual \( i \). Let \( F(x) \) be the distribution function for the degenerate random variable \( \tilde{x} = x^* \).

We show that \( \lim_{k \to \infty} F_k(x) = F(x) \) whenever \( F \) is continuous in \( x \), i.e., when \( x \neq x^* \).

Claim 1: Let \( x < x^* \), then \( \lim_{k \to \infty} F_k(x) = 0 = F(x) \).

Proof: Let \( x < x^* \) and let \( \varepsilon = |\Pi i \in N(h_0^i(\omega))^{\alpha_i} - \max \Pi i \in N(h_0^i(\omega))^{\alpha_i}| \). Also, let \( \tilde{k} \) be such that for all \( k > \tilde{k} \), \( |\Pi i \in N(h_0^i(\omega))^{\alpha_i} - \max \Pi i \in N(h_0^i(\omega))^{\alpha_i}| > \varepsilon / 2 \). Let \( \tilde{y}^m \in \arg \max \Pi i \in N(h_0^i(\omega))^{\alpha_i} \).
Remember that \( F_k(x) = \int_0^x \left( \frac{\sum_j T_{ij}^k}{\sum_i T_{ij}^k} \right)^k \, dv \)

\[
\leq \frac{x}{\int_0^1 \left( \frac{\sum_j T_{ij}^k}{\sum_i T_{ij}^k} \right)^k \, dv} = \frac{x}{\int_0^1 \left( \frac{\sum_j T_{ij}^k}{\sum_i T_{ij}^k} \right)^k \, dv}
\]

Note that by continuity, there must exist \( |x^* - y| > \delta > 0 \) such that \( |\Pi_{i\in N}(h_0^i(v)) - \Pi_{i\in N}(h_0^i(y))| > \frac{\delta}{K} \) for any \( v \in (x^* - \delta, x^* + \delta) \). Therefore there must be a \( k_0 \) such that for any \( k > k_0 \), \( |\Pi_{i\in N}(h_0^i(v)) - \Pi_{i\in N}(h_0^i(y))| > \frac{\delta}{K} \) for any \( v \in (x^* - \delta, x^* + \delta) \).

Therefore, \( F_k(x) \leq \frac{x}{\int_0^1 \left( \frac{\sum_j T_{ij}^k}{\sum_i T_{ij}^k} \right)^k \, dv} \to 0 \)

as \( K > 1 \) by boundedness of \( h_i(v) \).

Claim 2: Let \( x > x^* \), then \( \lim_{k\to\infty} F_k(x) = 1 = F(x) \).

Proof: The proof follows the proof of Claim 1 by focusing on \( 1 - F(x) \).

As we have shown that \( \lim_{k\to\infty} F_k(x) = 1 = F(x) \) for every \( x \) at which \( F \) is continuos, this implies convergence in distribution.

If we have a finite number of modes, \((x_1^*, \ldots, x_m^*)\). In this case we can repeat the above proof by showing that for any \( v \) and \( v' \) in \((x_1^*, x_{i+1}^*)\) we have \( F(v') - F(v) = 0 \).

We can now compare our model to the De Groot heuristic (see De Groot 1967 and the survey in Jackson 2012), in which individuals update by averaging the beliefs of themselves and others using weights that are described by some stochastic transition matrix. An important difference compared with the predictions of the De Groot heuristic is in the characteristics of the consensus reached compared to the set of initial beliefs. In the Bayesian heuristic, reaching a near-consensus among group members does not imply that beliefs do not change anymore. The type of consensus will shift as in Proposition 4, where agents with the same beliefs who face others with the same beliefs become more confident in the modes of these posterior beliefs. For example, in the fully connected network, all will have the same beliefs after the first stage, but will then keep on interacting and hence neglect the level of correlation in their information. They will converge then to the modes of the first period posterior. In the De Groot heuristic on the other hand, once all have the same beliefs, any averaging will result in the same beliefs and consensus in society will not change over time.

\*\*For Proposition 4, note that instead of \( \Pi_{i\in N}(h_0^i(v)) \sum_j T_{ij}^k \) we can write \( (h_0^i(v))^{\alpha_i} (h_0^i(v))^{(1-\alpha_i)^n} \) and the proof follows for \( n \to \infty \).
One last remaining question is whether individuals reach the correct beliefs, as for example fully Bayesian players who can identify where each piece of information comes from in the network (as in Acemoglu et al 2014). Our model allows to disentangle then the problems of repeated communication in a network, when individuals cannot track where the information had come from (and hence correlation) and rational updating. In the first period of communication, all act rationally when using our posterior (and when information is initially independent). This is different from the De Groot Heuristic for which this is true only for a particular information structure (such as the normal distribution, as in De Marzo et al 2003). Thus, in a one–off interaction where each indeed has independent information, information is fully aggregated (or in very large networks in which individuals would sample all beliefs in the population without meeting the same individual twice). But in a repeated interaction, as above, this would not be the case as individuals neglect the correlation in their information.

References

