Working Papers

Institute of Mathematical Economics

469

July 2012

# Opinion Dynamics under Conformity

Berno Buechel, Tim Hellmann and Stefan Klößner

IMW · Bielefeld University
Postfach 100131
33501 Bielefeld · Germany

email: imw@wiwi.uni-bielefeld.de
http://www.imw.uni-bielefeld.de/research/wp469.php
ISSN: 0931-6558

# Opinion Dynamics under Conformity

Berno Buechel
Economics Department
University of Hamburg
*berno.buechel@wiso.uni–hamburg.de*

Tim Hellmann
Institute of Mathematical Economics
Bielefeld University
*tim.hellmann@uni–bielefeld.de*

Stefan Klößner
Statistics and Econometrics
Saarland University
*s.kloessner@mx.uni–saarland.de*

July 12, 2012

**Abstract**

We present a model of opinion formation where individuals repeatedly engage in discussion and update their opinion in a social network similarly to the DeGroot model. Abstracting from the standard assumption that individuals always report their opinion truthfully, agents in our model interact strategically in the discussion such that their stated opinion can differ from their true opinion. The incentive to do so is induced by agents' preferences for conformity. Highly conforming agents will state an opinion which is close to their neighbors' while agents with low level of conformity may be honest or even overstate their opinion. We model opinion formation as a dynamic process and identify conditions for convergence to consensus. Studying the consensus in detail, we show that an agent's social influence on the consensus opinion is increasing in network centrality and decreasing in the level of conformity. Thus, lower conformity fosters opinion leadership. Moreover, assuming that the initial opinion is a noisy signal about some true state of the world, we consider the mean squared error of the consensus as an estimator for the true state of the world. We show that a society is "wise", i.e. the mean squared error is smaller, if players who are well informed are less conform, while uninformed players conform more with their neighbors.

Keywords: consensus, social networks, conformity, eigenvector centrality, wisdom of the crowds
JEL: C72, D83, D85, Z13

# 1 Introduction

If we could look into each other's heads, we would know what everybody thinks. Since this is not possible, we have to rely on what people say or do. Ignoring this issue, models of opinion formation have worked with the assumption that people do not misrepresent their opinion. They provide conditions for the emergence of consensus of opinions (DeGroot, 1974), identify opinion leaders (Friedkin, 1991), and even show that large societies can be "wise" in a well defined sense (Golub and Jackson, 2010). We challenge these results by incorporating the possibility that stated opinions differ from true opinions. This requires additional conditions to guarantee consensus, it affects who is an opinion leader, and it can undermine or foster the wisdom of societies, as we will show.

Individuals' opinions are important for several reasons. Majority opinions on political issues set the political course. The demand for a consumer product depends on the opinions about the quality of this product and about the integrity of its producing company. Moreover, opinions on the relative importance of issues decide upon the agenda of actions or on the allocation of a budget – be it within a company, within a government, or some other group of decision makers. Given the importance of opinions it is natural to ask where they come from. The cognitive process of combining pieces of information to an opinion might not be simple to understand. However, it seems beyond dispute that a person's social network, i.e. the people she talks and listens to, plays a central role in this process. Listening to opinions of others can simply be the source of new pieces of information which are combined with what has been known previously. A seminal model for this situation (i.e. the formation of opinions) is discussed by DeGroot (1974). In this so-called DeGroot model agents are assumed to update their opinion according to a weighted average of the current opinions. Thereby the weights of averaging are collected in an exogenously given learning matrix, which has the interpretation of a social network.

While the assumption of this form of naïve updating has been extensively discussed, motivated, and justified (Friedkin and Johnsen, 1990; DeMarzo et al., 2003), this is not true to the same extent for another crucial assumption of the DeGroot model framework: it is assumed that actors do not misrepresent their opinion; in other words, stated opinions are assumed to coincide with true opinions. This is certainly an important first step in analyzing opinion dynamics, but beyond that its justification is on shaky grounds. DeMarzo et al. (2003) argue that this assumption is problematic in contexts of persuasion, where actors have a material interest in influencing others' opinions. But even if there is no material incentive to persuade, people often misrepresent their opinions. In the famous study of Asch (1955), subjects wrongly judged the length of a stick after some other, allegedly neutral, participants had placed the same wrong judgment. Follow-up studies have shown that this effect is weaker if the subjects do not have to report their judgments publicly (Deutsch and Gerard, 1955). The authors argue that two forms of social influence can be observed in this study. While *informational social influence* describes the updating of opinions according to what others have said, *normative social influence* describes the behavior of stating an opinion that fits to the group norm.[1] Meanwhile, the concepts

[1]Deutsch and Gerard (1955, p. 629) further explain: "Commonly these two types of influence are found together. However, it is possible to conform behaviorally with the expectations of others and say things which one disbelieves but which agree with the beliefs of others. Also, it is possible that one will accept an opponent's beliefs as evidence about reality even though one has no motivation to agree with him, per se."

of informational and normative social influence have become a cornerstone in analyzing social influence, e.g. Ariely and Levav (2000, p. 279) call it the “primary paradigm”.

In terms of this paradigm, the DeGroot model of opinion formation is a model of informational social influence. It models how agents update their opinions according to a learning matrix (cf. Golub and Jackson, 2010). It is not a model of normative social influence, as long as stated opinions must not differ from true opinions. Models of normative social influence are also called models of conformity (e.g. Bernheim, 1994). They include a utility component that depends on the difference of the behavior of the focal actor and the behavior of some peer group or “reference group” (Hayakawa and Venieris, 1977). In particular, choosing a behavior that is different from the behavior of others might cause disutility, an effect that is called conformity (Jones, 1984).

It is argued that normative social influence is based on the level of identification with the peer group (Hogg and Abrams, 1988). Distinguishing between identification, non-identification and disidentification leads to three types of normative social influence: conformity, independence and counter-conformity/anti-conformity.[2] While instances of conformity are ubiquitous, examples for counter-conforming behavior are more difficult to find.[3] Hornsey et al. (2003) have run lab experiments where subjects could report their willingness to privately or publicly express and support their opinion. For subjects with a strong moral basis on the topic, the treatment of suggesting that a majority of the other subjects disagreed slightly increased the willingness to publicly express the opinion. As a special case of counter-conforming behavior one can consider exaggerating and overstating the own opinion to distinguish oneself from a group. A stylized fact on normative social influence is that people differ in the way and their degree of being influenced. The degree of conformity can be considered a personality trait, but it might also depend on the topic under discussion.

In this work, we present a model that incorporates both informational and normative social influence. The model consists of a sequence of discussion rounds. In each discussion round agents state an opinion depending on their true opinion and on their type, where types include conformity, counter-conformity, and honesty. From one discussion round to the next, learning takes place in the sense that agents update their opinion according to a learning matrix. In the special case where every agent is of the honest type, our model coincides with the classic DeGroot model which will serve as a benchmark throughout the paper.

We use this model to investigate *how opinions evolve given that actors may misrepresent their opinion in a conforming or counter-conforming way.* We first analyze the two-player case which illustrates that dynamics can diverge, converge or cycle. It turns out that a sufficiently conforming type will reach consensus with any other player. We then show more generally that excluding counter-conforming types is sufficient to guarantee convergence of opinions to consensus.

Assuming convergence, we then ask *how opinion leadership depends on conformity.* This research question is motivated by empirical research on identifying opinion leaders (cf. Katz and Lazarsfeld, 2005). A personality trait that seems strongly related to the

[2]Cf. table 8.2. in Hogg and Abrams (1988).

[3]Many findings on counter-conformity concern the decision on consumption goods. Especially concerning fashion and lifestyle products, a desire for “uniqueness,” that induces choices different from a majority of others, has been discussed (Snyder and Fromkin, 1980; Tian et al., 2001).

concept of conformity is called public individuation, i.e. the extent to which "people choose to act differently than others" (Maslach et al., 1985). This personality trait has been found to discriminate opinion leaders from followers (Chan and Misra, 1990). As any model following DeGroot (1974), our model provides an intuitive notion of opinion leadership. Opinion leadership or *power* of any agent can be measured by the influence of her initial opinion on the long-run (consensus) opinion of her group. As one of the main results, we show how power is determined by eigenvector centrality (Bonacich, 1972; Friedkin, 1991) with respect to the learning matrix and the distribution of conformity in the society. It turns out that conformity does not affect power if all agents have the same level of conformity. In a society with heterogeneous levels of conformity, power depends on each agent's conformity and centrality in relation to all others' conformity and centrality. Comparative statics reveal that an agents' power is decreasing in own level of conformity, increasing in other agents' level of conformity and increasing in network centrality. Thus, less conformity fosters opinion leadership, while a higher degree of conformity undermines it.

Finally, we consider a context where there is a true state of nature and the individuals' initial opinions are unbiased noisy signals which may differ with respect to signal precision (the inverse of the variance). The question is *how the misrepresentation of opinions affects the accuracy of information aggregation (the society's "wisdom").* Generically, it is the case that some agents are "too" powerful in comparison to their signal precision, while others are not powerful "enough." Using comparative statics we observe that for the goal of higher accuracy of the consensus opinion it would be helpful if people with a low signal precision (relative to their power) were more conforming, while people with a high signal precision (relative to their power) should be less conforming.

There is a wide branch of literature that is related to our work. Our model roots in the pioneer work of French (1956), Harary (1959), and DeGroot (1974). Friedkin and Johnsen (1990) provide a framework that subsumes former models as special cases. It refers not only to the pioneers mentioned above but also to the (sociological) literature on social influence and power (see also Friedkin, 1991). A particular feature of Friedkin (1991) is that opinions can be updated in every period not only according to the current profile of opinions but also according to the own initial opinion. Another variation of the classic model is to let agents only be affected by opinions that are not too different from the own opinion (Hegselmann and Krause, 2002). Moreover, Lorenz (2005) allows the learning matrix to vary over time and identifies general conditions for convergence. Under some conditions, convergence to consensus is also robust if updating is noisy, as Mueller-Frank (2011) shows. DeMarzo et al. (2003) allow the self confidence to vary over time. Besides their results, a particular contribution of DeMarzo et al. (2003) is the extensive discussion of the model: the authors justify the underlying rationality assumptions of the model and also provide several economic applications. Flache and Torenvlied (2004) study a variation of the classic model where actors anticipate the difference between own opinion and group decision ("frustration") and adapt learning weights ("salience") accordingly. Golub and Jackson (2010) discuss necessary and sufficient conditions for convergence in the classic model and provide conditions for the "wisdom of crowds" in the sense that the consensus opinion of a society comes arbitrarily close to the truth when letting the size of the society grow. Buechel et al. (2011) study the transmission of cultural traits from one generation to the next one and thereby introduce a generalization of the DeGroot model that also incorporates strategic interaction. While that model shares some properties of this present model, there are several crucial differences between the two: Buechel et al. (2011) consider a different updating rule and they restrict attention to a behavior that might be interpreted as counter-conformity.

Besides these highly related works, there are several contributions to similar research questions, but within different model frameworks. While their discussion is beyond the scope of this paper, we refer the reader to the following few prominent examples: models of social learning (Bikhchandani et al., 1992; Ellison and Fudenberg, 1993, 1995; Bala and Goyal, 1998, 2001), cooperative models of social influence (Grabisch and Rusinowska, 2010, 2011), and a model of strategic influence (Galeotti and Goyal, 2009). These models investigate social influence on a discrete choice of actions, such as the choice of one out of two technologies, as opposed to continuous opinions.

The rest of this paper is organized into four sections. In Section 2 we introduce the model. Before we present the main results (in Section 4), we discuss the two-player case (Section 3). Section 5 addresses the wisdom of a society and in Section 6 we conclude.

# 2 Model

We first present the set-up of opinion formation and then introduce misrepresentation of opinions.

## 2.1 Basic Setup

There is a set of agents/players $\mathcal{N} = \{1, 2, ..., n\}$ who interact with each other. A learning structure is given by a $n \times n$ row stochastic matrix $T$, i.e. $t_{ij} \geq 0$ for all $i, j \in \mathcal{N}$ and $\sum_{j=1}^{n} t_{ij} = 1$ for all $i \in \mathcal{N}$. This learning matrix represents the trust that the players put on the opinions of each other and it can be interpreted as a weighted and directed social network. We say that there is a directed path from $i$ to $j$ in this induced network, if there exists $i_1, ..., i_k \in \mathcal{N}$ such that $i_1 = i$ and $i_k = j$ and $t_{i_l i_{l+1}} > 0$ for all $l = 1, ..., k-1$ which is equivalent to $(T^k)_{ij} > 0$.

We study a dynamic model, where time is discrete $t = 0, 1, 2...$ and at the beginning each player has a predefined opinion $x_i(0)$ concerning some topic. The opinions of all players at time $t$ are collected in $x(t) \in \mathbb{R}^n$. In every period, players talk to each other and update their opinions according to the matrix $T$. In the classical DeGroot model players exchange opinions such that the opinions in period $t+1$ are formed by $x(t+1) = Tx(t) = T^{t+1}x(0)$ (DeGroot, 1974). The motivation for such a model is that players always report their true opinions and the next period's opinion of each player is formed as a weighted average of own and others' opinions according to the learning matrix $T$. Concerning the assumption of honesty in opinion formation, DeMarzo et al. (2003) note:

> For simplicity, we assume that agents report their beliefs truthfully. [...] We are thus ignoring issues of strategic communication.[4]

We relax this assumption. In particular, a player $i \in \mathcal{N}$ may choose to express some opinion $s_i(t) \in \mathbb{R}$ in period $t$.[5] We will call this the stated/expressed opinion.

[4] See DeMarzo et al. (2003), p. 3, footnote 9.

[5] In principle, it is possible to restrict the strategy space to some interval, say $[0, 1]$. However, assuming the strategy space $\mathbb{R}$, we do not have to deal with boundary conditions when calculating Nash equilibria. This makes the analysis more convenient.

A central assumption of our approach is that a player cannot observe the true opinions of the others but only their expressed opinions. Since each player knows her own true opinion $x_i(t)$, we get that player $i'$s next period's opinion is formed by $x_i(t+1) = t_{ii}x_i(t) + \sum_{j\neq i} t_{ij}s_j(t)$, where the weights $t_{ij}$ are the individual interaction weights, as in the classical model by DeGroot (1974). This holds for all players $i \in \mathcal{N}$ and, thus, the updating process becomes

$$x(t+1) = Dx(t) + (T - D)s(t), \tag{1}$$

where $D$ is the $n \times n$ diagonal matrix containing the diagonal of $T$.

## 2.2 How Players Misrepresent

There are various motives to misrepresent the own opinion. Not only strategic considerations of persuasion play a role, but also personality traits or emotional motives. We focus on two very basic features of human behavior: honesty and (non-)conformity. Stating a different opinion from one's true opinion (i.e. being dishonest) might cause discomfort (e.g. Festinger, 1957). Secondly, there is ample evidence that many people feel discomfort from stating an opinion that is different from their peer group's opinion (e.g. Asch, 1955; Jones, 1984). While certainly many people feel this type of normative social influence, this need not be true for all people—there are even some who enjoy stating an opinion that is far away from what others say (Hornsey et al., 2003).

In both cases, conformity and counter-conformity, a methodological challenge is involved. An individual's choice of behavior depends on the group norm, which is some aggregate of individual choices. Hence, we are in a situation of strategic interaction, which is best dealt with by using a game-theoretic approach. Therefore, we will model the choice of stated opinion in each round as a non-cooperative game that is played between actors.[6]

To formalize these ideas, suppose that $y_i'$ is a $1\times n$ vector with $y_{ij} \geq 0$ for all $j = 1, ..., n$ and $\sum_{j=1}^{n} y_{ij} = 1$ such that the subjective average of expressed opinions that player $i$ wants to conform with is given by $q_i(t) = y_i's(t)$. Since player $i'$s opinion should not matter in this average, we let $y_{ii} = 0$. Here, we assume that the subjective average $y_i'$ is given by the original learning matrix $T$, i.e. those players who player $i$ is connected with are also those players $i$ wants to conform/non-conform with.[7] In particular, we let $y_{ij} = t_{ij}/(1 - t_{ii})$ if $j \neq i$ and $y_{ii} = 0$. An interpretation of this assumption is that $q_i$ is the group norm as *perceived* by player $i$. Throughout the paper we will assume that $t_{ii} < 1$ for technical reasons.[8] When combining the incentive to conform with the reference group and the incentive to be honest, we assume that the utility of a player is additively separable into these two parts and that for each part the disutility takes a quadratic form. Thus, the

[6] While game-theoretic reasoning is often based on the assumptions of complete information and full rationality, those assumptions are not necessary in our game. We will have a unique and strict Nash equilibrium, which is also attractive in the sense that a sequence of boundedly rational adaptations of stated opinions would lead to this stable profile of behavioral choices.

[7] To study more general definitions of the relevant reference group $y_i$ is a task for further investigation.

[8] If $t_{ii} = 1$, then $y_{ii}$, and thus $q_i$, is not well-defined, i.e. for a player that does not learn from others and does not interact with others, the relevant peer group opinion is undefined. A simple way to fix this issue is to assume that those agents are honest, i.e. $s_i(t) = x_i(t)$. The case $t_{ii} = 1$ is not hard to solve because $x_i(t) = x_i(0)$ for any $t$. However, allowing for this case would considerably complicate the notation (because the matrix $T - D$ is not invertible then).

utility of some agent $i$ depends on the distance of true opinion to stated opinion as well as on the distance of stated opinion to perceived opinion in the following way:

$$u_i(s(t)|x_i(t)) := -\alpha_i \left(s_i(t) - x_i(t)\right)^2 - \beta_i \left(s_i(t) - q_i(t)\right)^2, \tag{2}$$

where $\alpha_i \in \mathbb{R}_{++}$ and $-\frac{\alpha_i}{2} < \beta_i \in \mathbb{R}$. For this utility representation $\alpha_i$ and $\beta_i$ display the relative importance of the preference for honesty and the preference for conformity. The assumption $\alpha_i > 0$ means that there is at least some incentive to report the opinion truthfully (i.e. some cost of dishonesty). Note that for $\beta_i > 0$ the second term is decreasing in distance between $s_i$ and $q_i$ incorporating discomfort of stating an opinion that differs from the benchmark opinion. This induces an incentive to conform. For $\beta_i < 0$ the second term is increasing in distance inducing an incentive to counter-conform. If $\beta_i = 0$, then a player has no incentive to misrepresent her opinion and thus will always report her opinion truthfully, this behavior can also be considered as non-conformity or independence. Further, we assume that $\beta_i > -\frac{\alpha_i}{2}$ for all $i \in \mathcal{N}$ mainly for technical reasons. This assumption restricts the extent of counter-conformity to a certain bound which seems weak enough to cover all reasonable cases and it guarantees that a utility maximum exists.[9] In some applications it may make sense to only allow for $\beta_i \geq 0$, since counter-conformity can be excluded and fully truthtelling can also be interpreted as non-conformity. However, we will see in the subsequent analysis that $\beta_i < 0$ will lead to an overshooting (or exaggerating) of own opinion and thus induces a behavior that is similar to attempts of convincing other players to the own opinion, and is, hence, also an interesting object to study.

The discussed assumptions determine the behavior of an agent in each period. We can think of one period of time as a discussion round. For each discussion round – which may be thought of as repeated exchange of opinions – players try to optimize against the expressed opinion of others. In particular, we assume that players find mutual best replies to each other's strategies, i.e. play a Nash equilibrium, at the end of each discussion round. We show in Proposition 1 that a Nash equilibrium exists and is unique for every opinion profile $x \in \mathbb{R}^n$. For a player $i \in \mathcal{N}$ with opinion $x_i(t)$ in period $t$ and other players' expressed opinions $s_{-i}(t) = (s_j)_{j \neq i}(t)$, the best reply is given by the first order condition, $s_i(t) = \frac{\alpha_i}{\alpha_i+\beta_i} x_i(t) + \frac{\beta_i}{\alpha_i+\beta_i} q_i(t)$. Defining $\delta_i := \frac{\beta_i}{\alpha_i+\beta_i}$, we get

$$s_i(t) = (1 - \delta_i)x_i(t) + \delta_i q_i(t),$$

i.e. the stated opinion is a combination of true opinion and perceived opinion. Note that with the assumptions $\alpha_i > 0$ and $\beta_i > -\frac{\alpha_i}{2}$ we have $\delta_i \in (-1, 1)$. Since $\delta_i$ fully describes the behavior of player $i$, there is no need to subsequently work with $\alpha_i$ and $\beta_i$. $\delta_i$ can directly be interpreted as the degree of conformity of player $i$, which we also call her type. This is illustrated in Figure 1 with a player who's true opinion is $x_i$, while the perceived group norm is $q_i$.[10] A conforming type, i.e. $\delta_i \in (0, 1)$, would state an opinion between the true opinion $x_i$ and perceived opinion $x_i$. A counter-conforming type, i.e. $\delta_i \in (-1, 0)$, would state an opinion that is more extreme than the true opinion $x_i$ (with respect to the perceived opinion $q_i$. Finally, an honest type, i.e. $\delta_i = 0$, would straight-forwardly state the true opinion, i.e. $s_i = x_i$.

[9] Otherwise the second term dominates the first and, hence, the utility maximizing choice would be to choose an infinitely far away opinion which would yield unreasonable behavior.

[10] $\delta_i \in (-1, +1)$ corresponds to the depicted interval because the limit cases are $\delta_i = -1$, which corresponds to $s_i = x_i - (q_i - x_i)$, and $\delta_i = 1$, which corresponds to $s_i = q_i$.

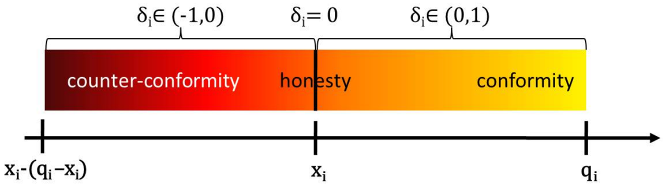



Figure 1: Type space: Conforming, counter-conforming, and honest types.

Note that the feasible set of stated opinions depends on the difference between $q_i$ and $x_i$. With our modeling assumptions, especially due to the quadratic disutilities, the stated opinion is proportional to the difference of $q_i$ and $x_i$ and the proportion is determined by the personality trait $\delta_i$.

Thus, we study a heterogeneous society of players who differ with respect to their degree of conformity and their position in the social network. Let $\Delta$ denote the $n \times n$ diagonal matrix with entries $\delta_i$ on the diagonal. In each period the players interact by stating some opinions. This constitutes a non-cooperative game, which is represented by the tuple $(\mathcal{N}, \mathbb{R}^n, u(\cdot|x(t)))$. Proposition 1 shows that there always exists a unique Nash equilibrium in this game and we show how it depends on the opinion profile in period $t$.

**Proposition 1.** *Let in period $t \in \mathbb{N}$ a non-cooperative game be given by $(\mathcal{N}, \mathbb{R}^n, u(\cdot|x(t)))$ such that $u$ satisfies* (2)*. Then there exists a unique Nash-equilibrium, given by $s^*(t) = [I - \Delta(I-D)^{-1}(T-D)]^{-1}(I-\Delta)x(t)$.*

## 2.3 Model Summary

Without modeling each discussion round explicitly, we simply assume that in each period the unique Nash equilibrium is played. Since opinions of period $t+1$ are formed by (1) and the Nash equilibrium of each period can be calculated as in Proposition 1, we conclude that the opinion profile in period $t+1$ depends on the opinion profile in period $t$ in the following way:

$$x(t+1) = Mx(t), \tag{3}$$

where $M := \Big[D + (T-D)[I-\Delta(I-D)^{-1}(T-D)]^{-1}(I-\Delta)\Big]$. Note that the transformation from $x(t)$ to $x(t+1)$, i.e. the matrix $M$, is independent of $x(t)$. Thus, the opinion dynamics under conformity is fully described by the power series $M^t$, since $x(t+1) = Mx(t) = M^2x(t-1) = ... = M^{t+1}x(0)$.[11] The relation to the classical DeGroot model becomes apparent in this expression when recalling $x(t+1) = Tx(t) = T^{t+1}x(0)$. In that light the misrepresentation of opinions leads to a transformation of the matrix $T$ into the matrix $M$.[12] If every player is honest, i.e. $\delta_i = 0$ for any $i \in \mathcal{N}$, then $M = T$ and, hence, we are back in the classical case of DeGroot. This special case will serve as a benchmark throughout the paper.

Let us illustrate the model introduced above by an example with three players.

[11]The simple linear structure of the best-replies is of course implied by our assumption of quadratic utility. This assumption enables us to focus on the power series $M^t$ and the structure of $M$ itself.

[12]Buechel et al. (2011) study a similar, but different, transformation of the DeGroot model.

**Example 1.** *Suppose there are three players. Player 1 (black) starts with an opinion* $x_1(0) = 0$, *Player 2 (red) and Player 3 (blue) have initial opinions of* $x_2(0) = 50$ *and* $x_3(0) = 100$. *Player 2 is an honest (truth-telling) player, i.e.* $\delta_2 = 0$, *Player 3 is a conforming player, i.e.* $\delta_3 = .5 > 0$, *and Player 1 is a counter-conforming player, i.e.* $\delta_1 = -.5 < 0$. *To illustrate the implications of the different degrees of conformity we let the players be in a symmetric network position. In particular, we let the interaction structure be given by* $T = \begin{pmatrix} .6 & .2 & .2 \\ .2 & .6 & .2 \\ .2 & .2 & .6 \end{pmatrix}$.

*The dynamics of opinions are displayed in Figure 2. The solid lines indicate the dynamics of true opinions* $x(t)$, *the dashed lines display the expressed opinions* $s(t)$, *and the dotted lines the perceived opinions* $q_i$ *(subjective average opinion of others).*

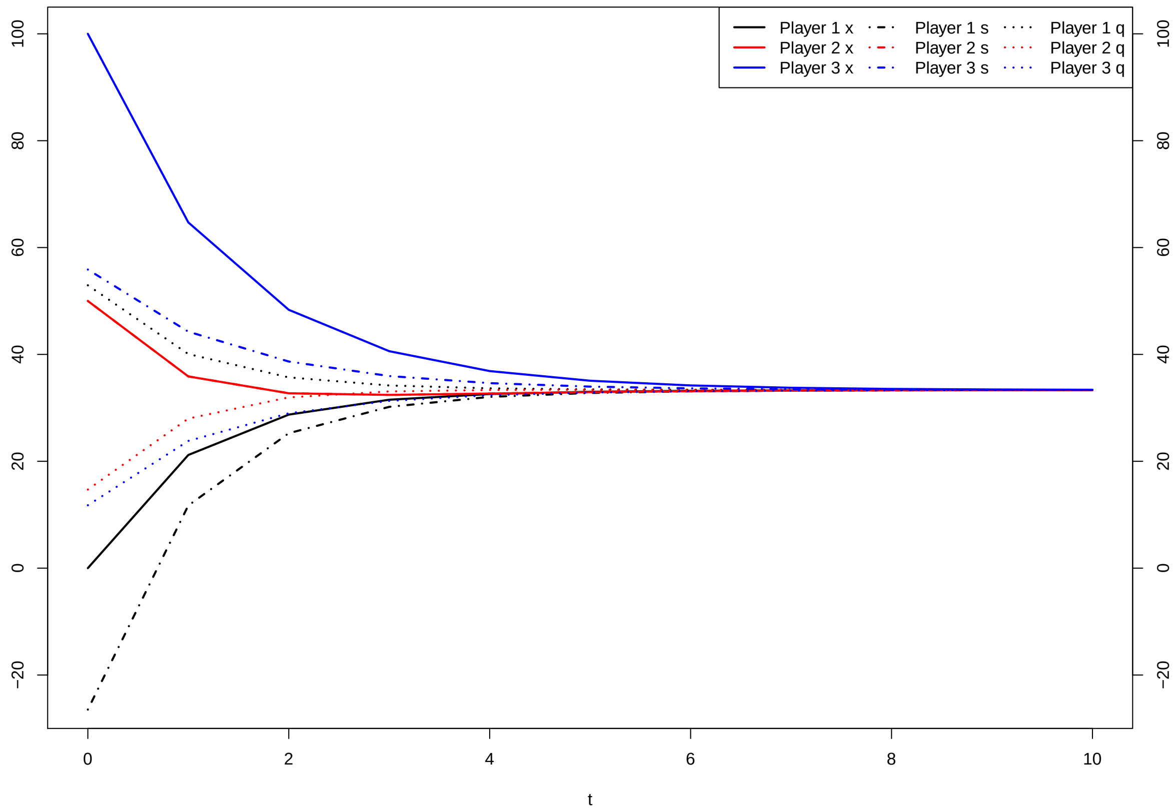



Figure 2: A three player example with one honest, one conforming, and one counter-conforming player.

*Since Player 2 (red) is honest, her stated opinion will always be equal to her true opinion. Therefore those functions (red dashed line and red solid line) coincide. Player 3 is a conforming player, she always expresses an opinion (dashed blue line) that is a convex combination of the perceived opinion of others (blue dotted line) and her true opinion (blue solid line). Player 1 is a counter-conforming player. With respect to the perceived average (black dotted line), she always expresses an opinion (black dashed line) that is more extreme than her true opinion (black solid line). This influences both Player 2* *and 3's perceived opinion and thus draws their true opinion more toward Player 1's true opinion. Such a behavior may, thus, also be interpreted as a convincing effort but as it is defined in our model, Player 2 simply enjoys (derives utility from) stating a more extreme opinion. Player 3, however, is a conforming player; her stated opinion draws the other players less into the direction of her true opinion.*

*The opinion dynamics are such that stated, true, and perceived opinions of each player become more and more similar. In the long–run, the dynamics of opinions converge to some consensus such that every player shares the same true opinion. Moreover, the perceived opinion $q(t)$, the stated opinion $s(t)$ and the true opinion $x(t)$ converge to the same point. To highlight the implications of the different degrees of conformity, let us compare this example with the DeGroot model, i.e. the case where every player is honest ($\delta_i = 0$, $i = 1, 2, 3$). In the case of DeGroot, due to the perfect symmetry of $T$, the long–run opinion would be* 50. *In this example, the long–run opinion is* 33.3 *which is caused by both the conformity of Player 3 (since she draws the others less towards her) and by the counter–conformity of Player 1 (since she draws the others more towards her). Thus, not only in the short–run a counter–conforming player is convincing, but her impact in the long–run is also larger than a conforming player.*

The dynamics of Example 1 highlights several features, the generality of which we discuss in the subsequent sections. While the opinion dynamics under conformity do not always converge, we provide sufficient conditions for convergence. Moreover, we elaborate on the properties of steady states and the properties of the consensus opinion. In particular we focus on the social influence of the players, i.e. the impact of each players initial opinion on consensus, as a function of network position and degree of conformity. We extend the setup to introduce the notion of wisdom, i.e. we examine the consensus with respect to information aggregation.

# 3 Two-Player Case

Let us begin with the analysis of the two player case. In this case closed form solutions are easy to obtain and, still, it is possible to observe several important properties of the opinion dynamics. (The $n$-player case is presented in Section 4.)

Let $n = 2$. Then we can write $T$ as

$$T = \begin{pmatrix} 1 - t_{12} & t_{12} \\ t_{21} & 1 - t_{21} \end{pmatrix}$$

with $t_{12}, t_{21} \in (0, 1)$. With only two players the relevant group average for one agent is simply the stated opinion of the other agent, i.e. $q_1 = s_2$ and $q_2 = s_1$. Plugging in the variables for $T$ into (3), yields

$$M = \begin{pmatrix} 1 - t_{12}\dfrac{1 - \delta_2}{1 - \delta_1\delta_2} & t_{12}\dfrac{1 - \delta_2}{1 - \delta_1\delta_2} \\ t_{21}\dfrac{1 - \delta_1}{1 - \delta_1\delta_2} & 1 - t_{21}\dfrac{1 - \delta_1}{1 - \delta_1\delta_2} \end{pmatrix}.$$

Recall that $x(t) = M^t x(0)$, i.e. the dynamics of opinions is determined by the power series of $M$ which depends on its eigenvalues. 1 is always an eigenvalue of $M$ because the

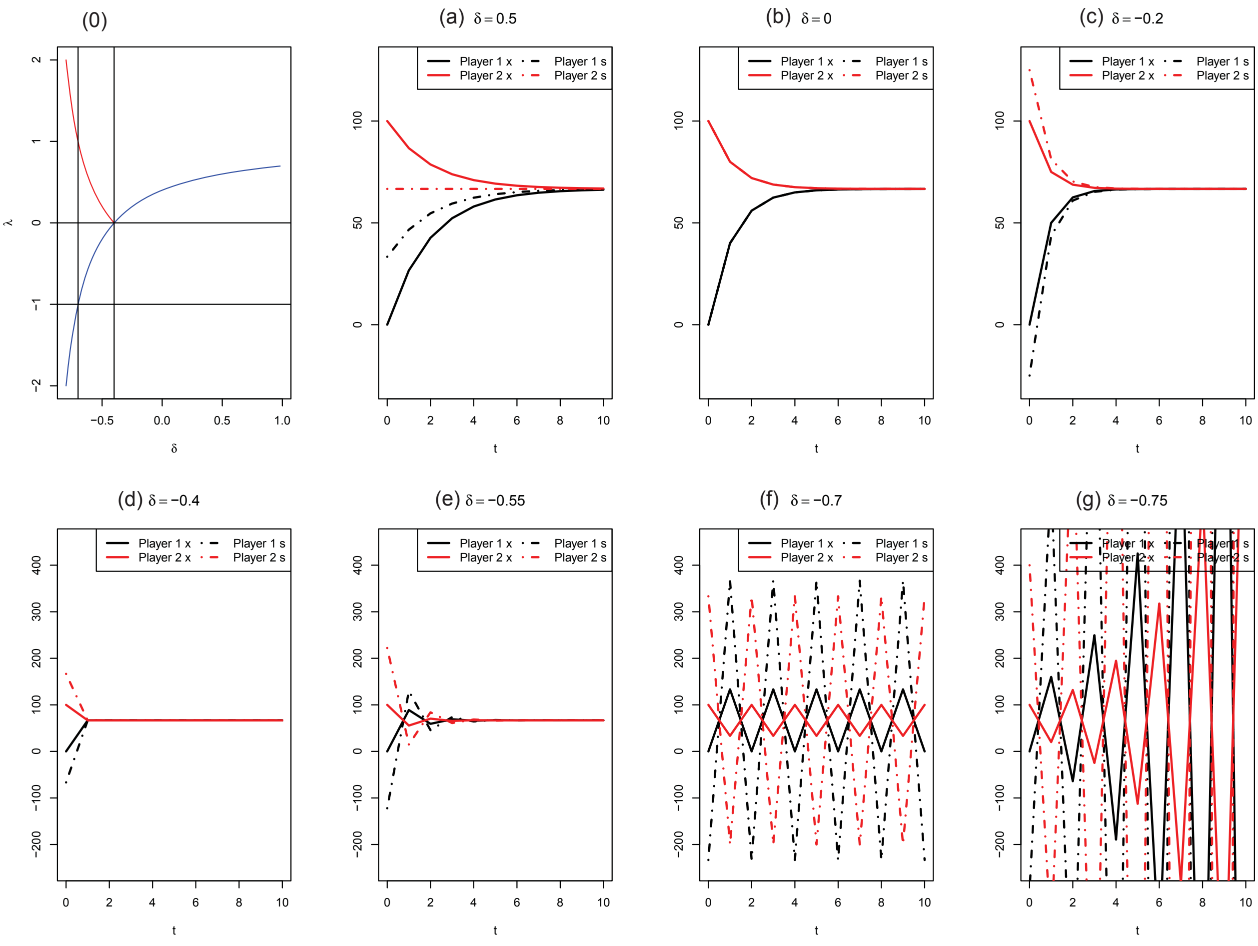



Figure 3: Seven cases of two players dynamics. Solid lines represent true opinions and dashed lines display stated opinions. (0) Shape of $\lambda$. (a) $\delta > 0$, conformity. (b) $\delta = 0$, honesty. (c) $\hat{\delta} < \delta < 0$, smooth convergence under counter–conformity. (d) $\delta = \hat{\delta}$, one-step convergence. (e) $\delta < \hat{\delta} < 0$, alternating dynamics with convergence under counter-conformity. (f) $\delta < \tilde{\delta}$, alternating dynamics ($\lambda = -1$). (g) $\delta < \tilde{\delta}$, divergence.

rows of $M$ sum up to one. Since the trace of a matrix equals its sum of eigenvalues, we can determine the second eigenvalue as $\lambda = 1 - t_{12}\frac{1-\delta_2}{1-\delta_1\delta_2} - t_{21}\frac{1-\delta_1}{1-\delta_1\delta_2}$ which is a real number. Since $\frac{1-\delta_i}{1-\delta_i\delta_j} > 0$ for $\delta_i, \delta_j \in (-1, +1)$, we have $\lambda < 1$, implying that $M$ is diagonalizable. Thus, we can write

$$M^t = V \begin{pmatrix} 1 & 0 \\ 0 & \lambda^t \end{pmatrix} V^{-1}, \tag{4}$$

where $V$ is the matrix of right-hand eigenvectors. In particular, $M^t$ converges if $|\lambda| < 1$ and, moreover, the smaller $|\lambda|$, the higher the speed of convergence.

To study the the effect of conformity/non-conformity on convergence and the speed of convergence, consider first the special case $\delta_1 = \delta_2 =: \delta$. Then $\lambda$ simplifies to

$$\lambda = 1 - \frac{1}{1+\delta}(t_{12} + t_{21}). \tag{5}$$

Suppose in addition that $t_{12} + t_{21} < 1$, i.e. the diagonals of $T$ have a higher value in sum than the off-diagonals. Then Figure 3 illustrates how the conformity parameter $\delta$ determines the second eigenvalue $\lambda$ and thereby the type of dynamics.[13] Part (0) depicts the shape of $\lambda$ as a function of $\delta$. $\lambda$ is increasing in $\delta$, and, moreover, $\lambda \overset{\delta\downarrow -1}{\to} -\infty$ and $\lambda \overset{\delta\uparrow +1}{\to} 1-\frac{1}{2}(t_{12}+t_{21})$. We have $\lambda \geq 0$ if and only if $\delta \geq (t_{12}+t_{21})-1 =: \hat{\delta}(\in (-1,1))$. Where $\lambda$ is negative, the absolute value $|\lambda|$ is depicted by the red curve. This absolute value is first decreasing then increasing in $\delta$. We distinguish 7 cases of different dynamics, labeled from (a) to (g). These cases are exhaustive in the sense that all possible convergence scenarios in the two player case with $\delta_1 = \delta_2$ and $t_{12}+t_{21} < 1$ are listed. Only case (a) shows dynamics under conformity; (b) to (f) are dynamics under counter-conformity. In cases (a) to (c), i.e. for $\lambda > 0 \Leftrightarrow \delta > \hat{\delta}$, we observe smooth convergence. Moreover, for these cases, convergence is slowed down by conformity. To see the reason, consider Player 2 in cases (a) and (b). If Player 1 conforms (case a), then Player 2 is less swayed to the center compared with the case (b) where Player 1 is honest.[14] Counter-conformity increases speed of convergence until the point where $\lambda = 0 \Leftrightarrow \delta = \hat{\delta}$ (case (d)). In that special case a steady state is reached immediately after one period of play, i.e. $x(1) = M^1 x(0) = M^\infty x(0)$, what we call one-step convergence. For $\lambda < 0 \Leftrightarrow \delta < \hat{\delta}$, i.e. cases (e) to (g), opinion dynamics alternate in the sense that the two players switch their relative positions. It depends on the degree of counter-conformity whether the alternating dynamics converge (like in case (e)), follow a two-cycle (case (f)), or diverge (like in case (g)). The condition for convergence is $|\lambda| < 1$, which becomes $\delta > \frac{1}{2}(t_{12}+t_{21}) - 1 =: \tilde{\delta}(< 0)$.[15]

In the two player case with $\delta_1 = \delta_2$ the following observations hold generally: For $\delta \leq \tilde{\delta}$ we have alternating dynamics that do not converge; for $\delta \in (\tilde{\delta}, \hat{\delta})$, we have alternating dynamics that converge; for $\delta \geq \hat{\delta}$, we have smooth dynamics that converge. Moreover, speed of convergence is increasing in conformity under alternating dynamics ($\delta < \hat{\delta}$) and decreasing in conformity under smooth dynamics ($\delta > \hat{\delta}$), while one-step convergence happens for $\delta = \hat{\delta}$.

Now, let us relax the assumption $\delta_1 = \delta_2$. It is still possible to have one-step convergence. The relevant condition, $\lambda = 0$, is equivalent to $t_{12}\frac{1-\delta_2}{1-\delta_1\delta_2} + t_{21}\frac{1-\delta_1}{1-\delta_1\delta_2} = 1$.

Since $\lambda < 1$ (see reason above), the necessary and sufficient condition for convergence of $M^t$ becomes $\lambda > -1$, which is equivalent to

$$t_{12}\frac{1-\delta_2}{1-\delta_1\delta_2} + t_{21}\frac{1-\delta_1}{1-\delta_1\delta_2} < 2. \qquad (6)$$

To interpret this condition in terms of individual conformity parameters, let us distinguish two cases:

(i) If $\delta_2 \leq \frac{2t_{21}+t_{12}-2}{2+t_{12}}$, then $M^t$ converges if and only if $\delta_1 > \frac{t_{12}(1-\delta_2)+t_{21}-2}{t_{21}-2\delta_2}$.

(ii) If $\delta_2 > \frac{2t_{21}+t_{12}-2}{2+t_{12}}$, then $M^t$ converges for any $\delta_1 \in (-1,+1)$.

[13] If the assumption $t_{12}+t_{21} < 1$ is violated, the picture and discussion only slightly change. Therefore, we omit the discussion of that case.

[14] Since players are resistant against their own misrepresentation, conformity of Player 1 does not increase her move to the opinion of Player 2.

[15] Another aspect that can be observed in this figure is that under convergence, i.e. in cases (a)-(e), it seems that the same opinion is approached in each case. We will show later on that this observation is not a coincidence and that it is induced by the setting $\delta_1 = \delta_2$.

As it can be checked, the threshold which defines the two cases is in the interval $(-1, \frac{1}{3})$ and given we are in case (i), the threshold for $\delta_1$ is below 1. Thus, if Player 2 has a relatively low degree of conformity (case (i)), then Player 1 must be sufficiently conforming in order to assure convergence. If we are in case (ii), where conformity of Player 2 is not too low, convergence is reached for any conformity of Player 1. In fact, $\delta_2 > \frac{1}{3}$ is sufficient to be in case (ii). Since similar arguments can be made by exchanging the players' labels, in the two-player case we always have convergence if there is a player with $\delta_i > \frac{1}{3}$. Thus, a sufficiently conforming player will reach consensus with any other player.

# 4 Opinion Dynamics

To study the dynamics of opinions of $n$ players, we first elaborate on the properties of steady states and the relation of true, perceived, and stated opinion. We then turn to providing conditions for convergence of opinions and finally determine where opinions converge to.

## 4.1 Steady States

The dynamics of stated opinions $s(t)$ can be derived from the dynamics of $x(t) = M^t x(0)$ since Proposition 1 determines $s(t)$ in dependence of $x(t)$. Also, the dynamics of $q(t)$ are determined by $s(t)$ since each perceived opinion $q_i(t)$ is exogenously defined as some weighted average of $s(t)$.

If it is the case that dynamics eventually settle down, we have $x(t+1) = x(t)$, which is only true if $Mx(t) = x(t)$. In general, we define $z \in \mathbb{R}^n$ to be a steady state of the opinion dynamics if $Mz = z$, i.e. if it is a (right-hand) eigenvector of $M$ corresponding to the eigenvalue 1. Considering the characteristic equation $\det(I - M) = 0$, we can rewrite its argument with use of (3) as follows:

$$I - M = \left[I - (T - D)\Delta(I - D)^{-1}\right]^{-1}(I - T), \tag{7}$$

as shown in the Appendix 6.2. Proposition 2 uses this expression to clarify the relation between steady states of the interaction structure $T$ and the law of motion of the opinion dynamics $M$ and the implications for the perceived, stated, and true opinions.

**Proposition 2** (Steady States). *1. The following are equivalent:*

*(a) $x$ is a steady state, i.e. $Mx = x$,*

*(b) $Tx = x$,*

*(c) perceived and true opinion coincide, i.e. $q = x$,*

*(d) perceived and stated opinion coincide, i.e. $q = s$.*

*2. If $s = x$, then $\delta_i(Tx - x)_i = 0$ for all agents $i \in \mathcal{N}$. If $\delta_i \neq 0$ for all agents $i \in \mathcal{N}$, then $s = x$ implies that $x$ is a steady state.*

The equivalence between $Mx = x$ and $Tx = x$ should not be misinterpreted. It does not mean that both dynamics $M^t x(0)$ and $T^t x(0)$ converge to the same vector of opinions. What this condition really means can best be seen when $T$ is irreducible, i.e. every player interacts (at least indirectly) with everybody else. Then, since $T$ is row stochastic, $Tx = x$ is equivalent to $x_i = x_j$ for all $i, j \in \mathcal{N}$. In this case, all those opinion profiles are steady states of $T$, where every player has the same opinion. We call this a consensus. Only consensus opinions can then be "steady states of $T$" (i.e. $Tx = x$) and hence of $M$. Thus, the opinion dynamics in our model (according to $M$) only lead to steady states that are also steady states of the special case with $\delta_i = 0$ for all $i$ (i.e. the classic DeGroot model), but they do in general not lead to the same vector of opinions when starting with some vector $x(0)$. Further, the equivalence between $Mx = x$ and $Tx = x$ implies that the rows of $M$ always sum up to 1. This is true since $T$ is row stochastic and hence $T\mathbb{1} = \mathbb{1}$ (where $\mathbb{1}$ is the vector of ones) and thus by the above result $M\mathbb{1} = \mathbb{1}$. Note however that, in contrast to $T$, $M$ may have negative entries.

Proposition 2 part 1 also shows that in a steady state true opinion, stated opinion and perceived opinion of any agent agree (since $x = q = s$). This is only true in a steady state. However, the fact that true opinion $x$ and stated opinion $s$ coincide is not sufficient for a steady state. The reason is simply that an honest agent ($\delta_i = 0$) always reports her opinion truthfully no matter of being in a steady state or not. In Part 2, however, we show that if agents are dishonest ($\delta_i \neq 0$ for all $i \in \mathcal{N}$), then all opinions are reported truthfully ($x = s$) only in a steady state.

In the following we study the power series $M^t$ and its limit because it determines the sequence of true opinions. Since $x(t) = M^t x(0)$ it is straightforward to see that the opinion dynamics $x(t)$ converges to a steady state (for any given initial opinion profile $x(0)$) if and only if $M^t$ converges. From Proposition 2 it follows that in this case also $q(t)$ and $s(t)$ converge. Note that we may also have convergence of opinions $x(t)$ if $M^t$ diverges. This can be most easily seen if every player starts with the same opinion (i.e. $x_i(0) = x_j(0)$ for all $i, j \in \mathcal{N}$). Then from Proposition 2 we get one period convergence of $x(t)$. This may also happen in the classical DeGroot model, i.e. such that $\delta_i = 0$.[16] However, in any case – whether or not $M^t$ converges – it is possible to show that in our model the true opinions $x(t)$ converge if and only if the stated opinions $s(t)$ converge, which is equivalent to convergence of perceived opinions $q(t)$ (see Appendix 6.3, Lemma 3). Moreover, all converge to the same limit. Thus, throughout the paper, we will restrict our analysis to the dynamics of true opinions $x(t)$.

## 4.2 The Structure of $M$

While some of the intuition gained in the two-player case will generalize, there are features of larger networks, that cannot appear between only two players. First, in the 2-player case, both players necessarily interact with one another since we assume that $t_{ii} < 1$ for all $i = 1, ..., n$. When considering the opinion dynamics with $n$-players, there can be players who are not influenced at all by one another, or where the influence is only one-way. Intuitively, even with our assumptions of conformity respectively non-conformity, the opinion of a given player should not matter for another player in the short or long run if there is no path in the network $T$ connecting both players, i.e. no (indirect) interaction takes place. We thus consider a partition of the player set $\mathcal{N}$ such that the players are ordered into groups which are determined by the interaction patterns, i.e. the paths in the network implied by $T$.

[16] See Berger (1981) for a necessary and sufficient condition for convergence of opinions in the DeGroot model.

**Definition 1.** *Let $\Pi(\mathcal{N},T)=\{\mathcal{C}_1,\mathcal{C}_2,...,\mathcal{C}_K,\mathcal{R}\}$ be a partition of $\mathcal{N}$ into $K(\geq 1)$ groups and the (possibly empty) rest of the world $\mathcal{R}$ such that:*

- *Each group $\mathcal{C}_k$ is strongly connected, i.e. $\forall i,j\in\mathcal{C}_k$ there exists $l\in\mathbb{N}$ such that $(T^l)_{ij}>0$.*
- *Each group $\mathcal{C}_k$ is closed, i.e. $\forall i\in\mathcal{C}_k$, $T_{ij}>0$ implies $j\in\mathcal{C}_k$.*
- *The (possibly empty) rest of the world consists of the players who do not belong to any group, i.e. $\mathcal{R}=\mathcal{N}\setminus\bigcup_{k=1}^{K}\mathcal{C}_k$.*

With a suitable renumeration, each matrix $T$ can be organized into blocks which correspond to the groups of the partition $\Pi(\mathcal{N},T)$:

$$T=\begin{pmatrix} T_{11} & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & T_{KK} & 0 \\ T_{\mathcal{R}1} & \cdots & \cdots & T_{\mathcal{R}K} & T_{\mathcal{R}\mathcal{R}} \end{pmatrix} \tag{8}$$

with $T_{kk}=T|_{\mathcal{C}_k}$, $T_{\mathcal{R}\mathcal{R}}=T|_{\mathcal{R}}$, and $T_{\mathcal{R}k}$ consisting of the rows of $T$ belonging to $\mathcal{R}$ and the columns of $T$ belonging to $\mathcal{C}_k$. This kind of organizing the players into groups and organizing the matrix into blocks is standard in the literature based on the DeGroot model (DeMarzo et al., 2003; Golub and Jackson, 2010).

Proposition 3 explicitly shows that $M$—and in fact $M^t$—has the same block structure as $T$. Moreover, it characterizes $M^t$.

**Proposition 3** (Blocks)**.** *Let $T$ be given as in* (8)*, i.e. organized into blocks according to the partition $\Pi(\mathcal{N},T)=\{\mathcal{C}_1,\mathcal{C}_2,...,\mathcal{C}_K,\mathcal{R}\}$. Then for every $t=1,2,...$ we have*

$$M^t=\begin{pmatrix} M_{11}^t & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & M_{KK}^t & 0 \\ (M^t)_{\mathcal{R}1} & \cdots & \cdots & (M^t)_{\mathcal{R}K} & M_{\mathcal{R}\mathcal{R}}^t \end{pmatrix}$$

*with*

$$M_{kk}^t=[I-\left(I-(T_{kk}-D_{kk})\Delta_{kk}(I-D_{kk})^{-1}\right)^{-1}(I-T_{kk})]^t$$

*for all $k=1,\ldots,K,\mathcal{R}$, and*

$$(M^t)_{\mathcal{R}k}=\sum_{l=0}^{t-1}M_{\mathcal{R}\mathcal{R}}^l M_{\mathcal{R}k}M_{kk}^{t-1-l},$$

*where $M_{\mathcal{R}k}=(I-(T_{\mathcal{R}\mathcal{R}}-D_{\mathcal{R}\mathcal{R}})\Delta_{\mathcal{R}\mathcal{R}}(I-D_{\mathcal{R}\mathcal{R}})^{-1})^{-1}T_{\mathcal{R}k}[(I-\Delta_{kk}(I-D_{kk})^{-1}(T_{kk}-D_{kk}))^{-1}(I-\Delta_{kk})]$ for all $k=1,\ldots,K$.*

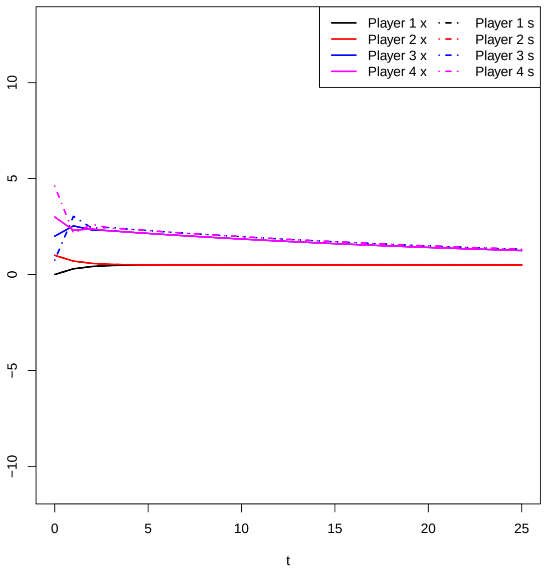
The proof of Proposition 3 is given in Appendix 6.4. Concerning the block structure of $M^t$ and considering that $x(t) = M^t x(0)$, Proposition 3 shows that the opinion dynamics of each group $\mathcal{C}_k$ can be studied independently. Only for players in $\mathcal{R}$ multiple groups may matter. The players in $\mathcal{R}$, on the other hand, do not affect the dynamics within groups.

More importantly, Proposition 3 provides an explicit expression for $M^t$ and thus for the sequence of true opinions (since $x(t) = M^t x(0)$). Let us now investigate the limit of this sequence.

## 4.3 Convergence

From Proposition 3 it becomes apparent that each closed and strongly connected group can be studied independently. Therefore, it is necessary for convergence of $M^t$ that for all groups $\mathcal{C}_k$ of $T$ the relevant blocks $M^t_{kk}$ converge for $t \to \infty$. To see that this is not sufficient, consider the following example:

**Example 2.** *Suppose there are four players such that* $T = \begin{bmatrix} .7 & .3 & 0 & 0 \\ .3 & .7 & 0 & 0 \\ .03 & .03 & .7 & .24 \\ .03 & .03 & .24 & .7 \end{bmatrix}$. *Thus players* 1 *and* 2 *form a closed and strongly connected group* $\mathcal{C}_1$, *while players* 3 *and* 4 *are the rest of the world* $\mathcal{R}$. *Let the conformity parameter* $\delta$ *be given by* $\delta = (0, 0, \delta_{ROTW}, \delta_{ROTW})$ *Figure 4 shows the opinion dynamics for the cases* $\delta_{ROTW} = -.7$ *and* $\delta_{ROTW} = -.92$. *While convergence within the closed and strongly connected group is guaranteed, the ROTW may cause divergence of* $M^t$ *for* $t \to \infty$.



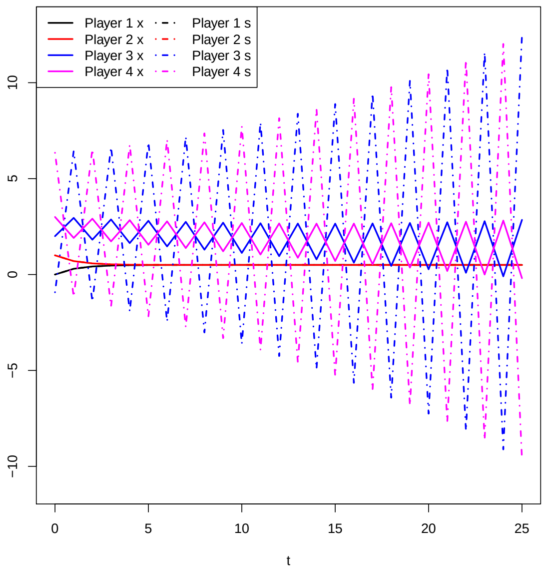




Figure 4: The opinion dynamics of Example 2 for (a) $\delta_{ROTW} = -.7$ and (b) $\delta_{ROTW} = -.92$.

Thus, convergence of all closed and connected groups $M^t_{kk}$ is not sufficient for convergence of $M^t$. In Proposition 4, we identify the additional condition on the ROTW such that $M^t$ converges.

**Proposition 4** (Convergence result 1)**.** *Let the block structure of $M$ be given as in Proposition 3. $M^t$ converges for $t \to \infty$ if and only if $M^t_{kk}$ converges for all $k = 1, \dots, K$ and $M^t_{\mathcal{R}\mathcal{R}}$ converges to* 0*.*

The proof of Proposition 4 is given in Appendix 6.5. Proposition 4 presents a necessary and sufficient condition for convergence of $M$ in terms of the block structure. In Example 2 the condition that $M_{\mathcal{R}\mathcal{R}}$ converges to 0 fails since strong counter-conformity of two players leads to eigenvalues with high absolute value to the extent that $|\lambda_{\mathcal{R}\mathcal{R}}| > 1$, for some eigenvalue of $M_{\mathcal{R}\mathcal{R}}$. A similar violation of the necessary condition for convergence occurs if counter–conformity of players in the closed and strongly connected groups is too high. Thus, one can derive the intuition that too high counter-conformity may cause non-convergence. The following result presents simple conditions on the degree of conformity and the interaction structure that ensure convergence of the opinion dynamics.

**Proposition 5** (Convergence result 2)**.** *$M^t$ converges for $t \to \infty$ if $\forall i \in \mathcal{N}$ we have $t_{ii} > 0$ and $\delta_i \geq 0$.*

The condition presented here is fairly weak. If we exclude counter-conformity ($\delta_i \geq 0$), and every individual has at least some self-confidence, then the opinion dynamics converge. Although all cases of conformity are covered by Proposition 5, it is important to emphasize that this condition is not necessary for convergence. Examples of convergence also with counter-conformity are given in Examples 1, 2 and in Section 3.

## 4.4 Long-run

For the remainder, we now assume that the power series $M^t$ converges. Although conformity is sufficient for convergence, we do not explicitly assume this. There may be some counter–conforming players in the society.

We are now left to address where opinions converge to (in the long-run) when starting with some opinion profile $x(0)$. The answer to this question depends on the learning matrix $T$ and the conformity parameter $\delta$. We are particularly interested in the influence of each player's initial opinion on the long–run opinion given her network position and her degree of conformity. By Proposition 3, we can again focus on the elements of the partition $\mathcal{C}_k \in \Pi(\mathcal{N}, T)$ separately such that we can characterize $\lim_{t\to\infty} M^t$ in terms of each $M^t_{kk}$ and the rest of the world $\mathcal{R} \in \Pi(\mathcal{N}, T)$.

**Theorem 1.** *Let $T$ and $M$ be organized as in Proposition 3. We denote by $w, v \in \mathbb{R}^n$ the vectors that fulfill the following: for each closed and strongly connected group $\mathcal{C}_k \in \Pi(\mathcal{N}, T)$, $w_{|\mathcal{C}_k}$ is the left unit eigenvector of $T_{kk}$ with $\sum_{i \in \mathcal{C}_k} w_i = 1$, while $v_{|\mathcal{C}_k}$ is left unit eigenvector of $M_{kk}$ with $\sum_{i \in \mathcal{C}_k} w_i = 1$. If $M^t$ converges for $t \to \infty$ to some matrix $M^\infty$, then the following holds:*

$$M^\infty = \begin{pmatrix} M^\infty_{11} & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & M^\infty_{KK} & 0 \\ M^\infty_{\mathcal{R}1} & \cdots & \cdots & M^\infty_{\mathcal{R}K} & 0 \end{pmatrix}$$

*with*

$$M^{\infty}_{kk} = \mathbb{1}_{\mathcal{C}_k} v'_{|\mathcal{C}_k} = \mathbb{1}_{\mathcal{C}_k} w'_{|\mathcal{C}_k} \frac{I - \Delta_{kk}}{\mathbb{1}'_{\mathcal{C}_k}(I - \Delta_{kk}) w_{|\mathcal{C}_k}}, \tag{9}$$

*and*

$$M^{\infty}_{\mathcal{R}k} = (I - T_{\mathcal{R}\mathcal{R}})^{-1} T_{\mathcal{R}k} M^{\infty}_{kk} \tag{10}$$

*for all* $k = 1, \ldots, K$.

Theorem 1, the proof of which can be found in Appendix 6.6, fully characterizes the long-run dynamics of (true) opinions given convergence since $x(\infty) = M^{\infty} x(0)$. For the interpretation of the result, we distinguish again between the closed and strongly connected groups $\mathcal{C}_k$ and the rest of the world $\mathcal{R}$.

We can first observe that the long–run opinions may differ across groups, but each closed and strongly connected group $\mathcal{C}_k$ reaches a consensus $c_k$ as each block $M^t_{kk}$ of $M^t$ converges to a matrix of rank 1. Each row of $M^{\infty}_{kk}$ is given by the left-hand unit eigenvector $v'_{|\mathcal{C}_k}$, implying

$$c_k := x_i(\infty) = x_j(\infty) = v'_{|\mathcal{C}_k} x(0)_{|\mathcal{C}_k} \tag{11}$$

for all agents $i, j$ in group $\mathcal{C}_k$. The left-hand normalized unit eigenvector $v'_{|\mathcal{C}_k}$ thus displays the extent to which the initial opinion of each player $i$ within group $\mathcal{C}_k$ matters for consensus. Moreover, $v'_{|\mathcal{C}_k}$ is dependent on $w'_{|\mathcal{C}_k}$, the left-hand unit eigenvector of $T_{kk}$, and the conformity parameters within the group, $\Delta_{kk}$. We delay the interpretation of this result to the next subsection.

The long-run opinion of a player in the ROTW is simply some weighted average of the long-run opinions $c_1, \ldots, c_k$ within the groups 1, ..., $K$. To see this, consider the matrix

$$\Gamma := (I - T_{\mathcal{R}\mathcal{R}})^{-1} (T_{\mathcal{R}1} \mathbb{1}_{|\mathcal{C}_1}, \ldots, T_{\mathcal{R}K} \mathbb{1}_{|\mathcal{C}_K})$$

which is easily seen to be row-stochastic and enables translating (10) into

$$x(\infty)_{|\mathcal{R}} = \Gamma c \tag{12}$$

with the $K$-dimensional vector $c = (c_1, \ldots, c_K)'$ combining the long–run opinions of the closed and strongly connected groups. Thus, the initial opinion of some player in the ROTW does not affect the long–run opinion profile $x(\infty)$ since the ROTW players end up with a weighted average of the consensus opinions of the closed and strongly connected groups, which in turn are dependent on the initial opinions within those groups. Moreover, the weights of averaging depend on $T$ but not on the conformity parameters $\delta$. Consequently, the long-run opinion of an agent in the ROTW neither depends on an initial opinion nor on the conformity parameter of any agent within the ROTW (including herself). Since each player in the ROTW may average differently between consent opinions of the closed and strongly connected groups, the players in the ROTW need not reach a consensus if there is more than just one closed and strongly connected group. Summing up, the important contribution of Theorem 1 lies in the characterization of $v$ as a function of $w$ and $\Delta$, as we will discuss next.

## 4.5 Interpretation (Opinion Leadership)

To simplify the discussion let us now restrict attention to one closed and strongly connected group by assuming that there is only one such group, i.e. $\Pi(\mathcal{N}, T) = \mathcal{N}$. For this purpose it is sufficient to assume that $T$ is strongly connected or, equivalently, that $\text{rk}(I - T) = n - 1$.

Theorem 1 implies that convergence leads to consensus within a group. Assuming $\text{rk}(I - T) = n - 1$ we get from (11) that $x(\infty) = \mathbb{1}v'x(0)$ and hence $x_j(\infty) = v'x(0) = \sum_{i \in \mathcal{N}} v_i x_i(0)$. Thus, an entry $v_i$ of $v$ determines the weight of the original opinion of agent $i$ on the long-run consensus opinion of her group. This is a very intuitive formalization of opinion leadership: $v$ measures the social influence of each player on the group.

Note that for $\delta_i = 0$ for all $i \in \mathcal{N}$, (9) yields $v = w$, i.e. opinion leadership is fully determined by the unit eigenvector of $T$. $w$ is a well-studied object in network science. It is known as eigenvector centrality or Katz Prestige (Bonacich, 1972; Friedkin, 1991). This index of centrality or power in a social network is recurrently defined via the columns of $T$ (formally via the rows of $T'$): An agent is powerful if she is important for agents who are powerful themselves.

When relaxing the assumption that every player is honest, then the following Corollary of Theorem 1 shows how opinion leadership is not only determined by eigenvector centrality, but also by the degree of conformity.

**Corollary 1.** *Let* $\text{rk}(I - T) = n - 1$. *Let $w$, respectively $v$, be the normalized left-hand unit eigenvectors of $T$, respectively $M$. Then we have for any $i \in \mathcal{N}$*

$$v_i = \frac{(1 - \delta_i)w_i}{\sum_{j \in \mathcal{N}} (1 - \delta_j) w_j}. \tag{13}$$

*Moreover,*

$$\frac{\partial v_i}{\partial \delta_k} = \frac{w_k}{\sum\limits_{j=1}^{n} w_j (1 - \delta_j)} \left( \frac{w_i (1 - \delta_i)}{\sum\limits_{j=1}^{n} w_j (1 - \delta_j)} - 1_{i=k} \right) = \frac{w_k}{\sum\limits_{j=1}^{n} w_j (1 - \delta_j)} \left( v_i - 1_{i=k} \right). \tag{14}$$

Opinion leadership (social influence) $v_i$ of some agent $i$ is determined by the *combination* of her network centrality in $T$ ($w_i$) and the individual conformity $\delta_i$ weighted by the sum of these values of all players as becomes apparent from (13). Thus, there is complementary relationship between network centrality and $1 - \delta_i$ (call it: the degree of non-conformity): Social influence becomes minimal $v_i \to 0$ if either $i'$s network centrality approaches zero or $i$ is fully conform ($\delta_i \to 1$). In the same sense, social influence is maximal if all other players' influence is minimal.

Taking the network $T$ as given, we can observe the comparative statics with respect to $\delta_i$. From (14) we get for all $i \in \mathcal{N}$ that opinion leadership is decreasing in "own" conformity $\delta_i$ and increasing in other players' conformity $\delta_k$, $k \neq i$, since $w_j \in [0, 1]$ and $1 - \delta_j \geq 0$ for all $j \in \mathcal{N}$. Thus, low own conformity fosters opinion leadership. The same is true if other players are more conforming. We may use (14) also to examine which player's influences decreases most in response to a marginal increase in her own conformity. From (14), we calculate that

$$\left| \frac{\partial v_i}{\partial \delta_i} \right| < \left| \frac{\partial v_j}{\partial \delta_j} \right| \quad \Leftrightarrow \quad w_j^2 (1 - \delta_j) - w_i^2 (1 - \delta_i) < (w_j - w_i) \left( \sum_{k=1}^{n} w_k (1 - \delta_k) \right). \tag{15}$$

Thus, if two players have the same network centrality, $w_i = w_j$ then by (15), $\left|\frac{\partial v_i}{\partial \delta_i}\right| < \left|\frac{\partial v_j}{\partial \delta_j}\right|$ if and only if $\delta_i < \delta_j$. In other words, the player with the already higher degree of conformity and thus lower influence looses even more influence in response to a marginal increase in conformity than a player with low conformity. Holding $\delta_i = \delta_j$, we get $\left|\frac{\partial v_i}{\partial \delta_i}\right| < \left|\frac{\partial v_j}{\partial \delta_j}\right|$ if and only if $w_i < w_j$, which implies that for two players with equal conformity the player with the higher network centrality looses more influence when increasing own conformity.

We can also use Corollary 1 to compare opinion leadership in our model $v$, with opinion leadership in the classic DeGroot model $w$, (i.e. in the special case of our model where every player $i$ is an honest type, $\delta_i = 0$). For this purpose consider first a society where all agents are characterized by the same trait, i.e. $\delta_j = \bar{\delta}$ for all $j \in \mathcal{N}$. Then (13) yields $v = w$: opinion leadership is not affected by conformity if all agents are characterized by the same level of conformity. More generally, we have $v_i \geq w_i$ if and only if $\delta_i \leq \sum_{j \neq i} \frac{w_j}{\sum_{k \neq i} w_k} \delta_j$, i.e. an agent's social influence in our model compared to the classic DeGroot model is fostered if $\delta_i$ is below some average of the others' conformity parameters.

Finally, Theorem 1 supports again the interpretation that counter-conformity can be considered a persuasion device. Not only the next period's opinion of others is drawn into direction of own opinion as noted in Example 1, but also their long-run opinion is influenced towards the own initial opinion. It must be pointed out, however, that "too much" counter-conformity of multiple players can lead to divergence of opinions.

# 5 Wisdom

The discussion so far applies to any continuous opinion including those for which no true value can be determined. In some applications, however, agents' opinions are more or less accurate with respect to some objective truth. A statistical phenomenon in this context is the fact that aggregating independent individual opinions yields an arbitrarily accurate estimate when the group size becomes large. This effect is sometimes called the "wisdom of crowds." One approach to study this phenomenon in the framework of the DeGroot model is provided in Golub and Jackson (2010). They call a sequence of growing societies 'wise' if, asymptotically, information aggregation is done in a way such that all agents' long–run opinions converge to the objective truth. While Golub and Jackson (2010) focus solely on the asymptotic properties for large societies, we are less interested in the wisdom of a growing society, but address how the accuracy of information aggregation within a given society depends on the conformity of its members.

In the following, we will therefore assume that there is some true value $\mu \in \mathbb{R}$ and that all agents of the society receive independent unbiased signals about $\mu$ which constitute the agents' initial opinions. Formally, this means that, for all $i \in \mathcal{N}$, agent $i$'s initial opinion $x_i(0)$ is a random variable with expected value $\mu$ and some individual variance $\sigma_i^2$, and that all $x_i(0)$ are uncorrelated random variables. Assuming that opinion dynamics converge, a very natural question to ask is how close the different steady state opinions will be to the true, but to the agents unknown, value $\mu$. To measure this closeness between $\mu$ and an estimate $\hat{\mu}$, we use the mean squared error (MSE), which is defined as $E((\hat{\mu} - \mu)^2)$. The

MSE can be decomposed into the squared bias, $(E(\hat{\mu}-\mu))^2$, and the estimator's variance, $\text{Var}(\hat{\mu})$:

$$E((\hat{\mu}-\mu)^2) = (E(\hat{\mu}-\mu))^2 + \text{Var}(\hat{\mu}).$$

As $x(\infty) = M^\infty x(0)$ and $M^\infty \mathbb{1} = \mathbb{1}$, it is obvious that $E(x(\infty)) = \mu\mathbb{1}$, i.e. all agents' long–run opinions are unbiased estimates for $\mu$. Denoting by $\Sigma$ the covariance matrix of $x(0)$, the corresponding MSEs are therefore given by the entries on the diagonal of $M^\infty\Sigma(M^\infty)'$. To study the effects of conformity on wisdom, we begin with an illustrative example.

### 5.1 Wisdom: an Example

Let $n = 10$, $(\sigma_1^2, \ldots, \sigma_{10}^2)$=(6, 4, 8, 7, 6, 3, 10, 12, 14, 16), and

$$T = \begin{pmatrix} 0.9 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.8 & 0.2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.7 & 0.3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.3 & 0.7 & 0 & 0 & 0 & 0 \\ 0.1 & 0 & 0 & 0 & 0 & 0 & 0.9 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0.3 & 0 & 0 & 0 & 0.5 & 0 & 0 \\ 0.1 & 0 & 0 & 0.1 & 0 & 0 & 0 & 0 & 0.8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.2 & 0 & 0 & 0.2 & 0.6 \end{pmatrix}.$$

In this situation, we have $K = 3$ closed and strongly connected groups, $\mathcal{C}_1 = \{1, 2\}$, $\mathcal{C}_2 = \{3, 4\}$, and $\mathcal{C}_3 = \{5, 6\}$, while players 7 to 10 form the rest of the world. If all agents report their opinions truthfully ($\Delta = 0$), we find the MSEs equal to (4, 4, 4, 4, 2.25, 2.25, 4, 4, 2, 1.0625). There are several striking features of this result. First of all, due to the fact that their long–run opinions are equal, all agents within a closed and strongly connected group share the same level of wisdom. Comparing the first two groups, we find the surprising result that the MSEs of these two groups are 4 each, although the first group enjoys significantly better initial signals of variances 6 and 4, while the second group seems to combine their less precise signals of variances 8 and 7 much more effectively than the first group. It is also remarkable that player 2, by communicating with player 1, ends up with exactly the same MSE of 4 that she would reach if she used only her own signal. With respect to the rest of the world, notice that these agents typically have different MSEs. Furthermore, players 7 and 8 each end up with the same MSE as the first two groups while players 9 and 10 achieve MSEs better than all members of the closed and strongly connected groups.

Now suppose that players 2, 3, and 5 are conforming with $\delta_2 = 5/9$, $\delta_3 = 2/3$, and $\delta_5 = 1/2$ (and $\delta_i = 0$ for all other players). Then wisdom levels can be calculated to be (4.9, 4.9, 4, 4, 2, 2, 4.9, 4, 2.225, 1.05625). Thus, increasing conformism may lead to less wisdom (as the first group's MSE becomes larger), the same wisdom (as the second group's MSE does not change), or more wisdom (as the third group's MSE becomes smaller). We also find that the players in the rest of the world are affected by the changes in conformity of the players in the closed and strongly connected groups: the MSE of players 7 and 9 becomes larger, while player 10's MSE improves slightly. It still holds that player 7 and 8's MSEs equal that of the first and second group, respectively.

We will now proceed by systematically analyzing the principles underlying the distribution of wisdom within the society.

## 5.2 Wisdom of Groups

Due to (11), a group $\mathcal{C}_k$ will, given convergence, eventually end up reaching a consensus where all players' opinions are equal to $\hat{\mu}_k := v'_{|C_k} x(0)_{|C_k}$. Hence, we can directly derive group $\mathcal{C}_k$'s wisdom as given by the MSE of $\hat{\mu}_k$.

**Lemma 1.** *The MSE of $\hat{\mu}_k$ is given by*

$$\mathrm{MSE}_k := E((\hat{\mu}_k - \mu)^2) = \sum_{i \in \mathcal{C}_k} v_i^2 \sigma_i^2 = \sum_{i \in \mathcal{C}_k} \left( \frac{(1-\delta_i)w_i}{\sum\limits_{j \in \mathcal{C}_k} (1-\delta_j)w_j} \right)^2 \sigma_i^2$$

As $v_i^2 \le v_i$ due to $v_i \in (0,1]$ for all agents $i$, we easily find that

$$\mathrm{MSE}_k = \sum_{i \in \mathcal{C}_k} v_i^2 \sigma_i^2 \le \sum_{i \in \mathcal{C}_k} v_i \sigma_i^2 \le \min_{i \in \mathcal{C}_k} \sigma_i^2. \tag{16}$$

Thus, group $\mathcal{C}_k$'s long-run opinion is on average at least as close to the true value $\mu$ as that of the agent with the least precise signal. This worst case is given when both inequalities in (16) become equalities, which is the case for $v_i \in \{0,1\}$ for all $i \in \mathcal{C}_k$ (first inequality) and $v_i = 0$ for all $i$ with $\sigma_i^2 < \max\limits_{j \in \mathcal{C}_k} \sigma_j^2$ (second inequality). Therefore, information updating within group $\mathcal{C}_k$ is worst when importance is given to only one player whose signal is most imprecise. This case would be approached if all other players were close to full conformity, i.e. $\delta_i$ close to 1.

We now consider the comparative static effect of one agent's conformity on the wisdom of her group.

**Proposition 6.** *The wisdom of a closed and strongly connected group $C_k$ is increasing in the conformity level of a group member $i$ if and only if $i$'s product of signal variance and power is larger than the group's MSE, i.e.*

$$\frac{\partial \, \mathrm{MSE}_k}{\partial \delta_i} \le 0 \;\Leftrightarrow\; v_i \sigma_i^2 \ge \mathrm{MSE}_k \, .$$

To give an interpretation for Proposition 6, let us rewrite $v_i \sigma_i^2 = \frac{v_i}{1/\sigma_i^2}$ and $MSE_k = \sum\limits_{j \in \mathcal{C}_k} v_j \frac{v_j}{\frac{1}{\sigma_j^2}}$. This shows that it is not a person's expertise alone which is decisive for the question of how this person can increase the group's wisdom, rather, it is the ratio of power over signal precision, $\frac{v_i}{1/\sigma_i^2}$: if agents with a high ratio as compared to the group's average are more conforming, then this will reduce their power within the group, decrease the group's MSE, and thereby increase its wisdom. Vice versa, agents who are not powerful enough in relation to their signal precision will increase the group's wisdom if they are less conforming, because this will increase their network importance, decrease the group's MSE, and foster its wisdom.

The above discussion implies that in optimum, the ratio of network importance over signal precision must be constant within the group: $v_i \sigma_i^2 = v_j \sigma_j^2$ for all $i, j \in \mathcal{C}_k$. This is formalized in the following corollary:

**Corollary 2.** *For the wisdom of the group $\mathcal{C}_k$ as measured by* $\mathrm{MSE}_k$*, we have:*

$$\mathrm{MSE}_k \geq \frac{1}{\sum\limits_{j\in\mathcal{C}_k} \frac{1}{\sigma_j^2}} =: \mathrm{MSE}_k^*, \tag{17}$$

*with equality in* (17) *if and only if $v_i\sigma_i^2 = v_j\sigma_j^2$ for all $i, j \in \mathcal{C}_k$. The latter condition is equivalent to*

$$\delta_i = 1 - C\,\frac{1}{\sigma_i^2 w_i \sum\limits_{j\in\mathcal{C}_k} \frac{1}{\sigma_j^2}} \text{ for all } i \in \mathcal{C}_k \tag{18}$$

*for some constant $C \in (0, 2\sum\limits_{j\in\mathcal{C}_k} \frac{1}{\sigma_j^2} \min\limits_{j\in\mathcal{C}_k} w_j\sigma_j^2)$.*

Corollary 2 delivers the analogue to (16). While (16) describes the worst case with respect to wisdom, Corollary 2 considers the best scenario: all agents within the same closed and strongly connected group share the same ratio of network importance over signal precision, and this case can always be achieved if the players' conformity is distributed suitably. Notice that, in particular, choosing $C \in (0, \sum\limits_{j\in\mathcal{C}_k} \frac{1}{\sigma_j^2} \max\limits_{j\in\mathcal{C}_k} w_j\sigma_j^2)$ in (18) ensures $\delta_i > 0$ for all $i \in \mathcal{C}_k$ and therefore guarantees convergence of the opinions in $\mathcal{C}_k$ to the best possible consensus $\hat{\mu}_k$. Notice also that the optimal MSE is smaller than agent $i$'s signal's variance, $\sigma_i^2$, for all agents $i$ in group $\mathcal{C}_k$, as is easily seen from (17). Therefore, in the optimum, all agents benefit from the communication within $\mathcal{C}_k$.

Reconsidering the example discussed in subsection 5.1, we find $w_1 = 0.8$, $w_2 = 0.2$, $w_3 = 0.6$, $w_4 = 0.4$, $w_5 = 0.5$, and $w_6 = 0.4$. Therefore, in (18), the constant $C$ can be chosen in $(0, 2/3)$ (group 1) and $(0, 3/2)$ (groups 2 and 3). Choosing $C = 1/3$ (group 1) and $C = 3/4$ (groups 2 and 3) delivers $\delta_1 = 5/6$, $\delta_3 = 5/12$, and $\delta_5 = 1/2$ (and $\delta_i = 0$ for all other players). Thus, choosing the players' degree of conformity according to these values ensures the optimal wisdom within the respective groups given by (2.4, 2.4, $3.7\overline{3}$, $3.7\overline{3}$, 2, 2, 2.4, $3.7\overline{3}$, $1.5\overline{3}$, $0.88\overline{3}$). The same level could also be reached for other conformity levels, for instance, choosing $C = 1/4$ (first group), $C = 3/7$ (second group), and $C = 3/8$ (third group) in (18), we find that the conformism levels $\delta_{1:6} = (7/8, 1/4, 2/3, 3/7, 3/4, 1/2)$ also lead to the optimal wisdom. Notice that, as in Golub and Jackson (2010), wisdom thus is independent of the speed of convergence, as we have two examples with the same optimal wisdom but different speeds of convergence (the last-mentioned conformity levels lead to slightly slower convergence than the earlier mentioned ones).

## 5.3 Wisdom within the Rest of the World

Let us recall that agents in the rest of the world do not necessarily share a consensus opinion in the long–run, so that we will typically have individual wisdom levels. Due to (12), we have the following formula for the long–run opinions within the rest of the world: $x(\infty)_{|\mathcal{R}} = \Gamma\hat{\mu}$, with $\hat{\mu} := (\hat{\mu}_1, \ldots, \hat{\mu}_K)'$. Therefore, it is obvious that the wisdom levels in the rest of the world depend on the conformity levels of the agents in the closed and strongly connected groups as these affect the consensus opinions $\hat{\mu}_k$ of these groups. On the other hand, as neither the initial signals nor the conformity levels of the agents in the rest of the world play any role for their long–run opinions, these agents' wisdom is independent of their conformity levels as well as of their initial signals. In other words, if the rest of the world is non-empty, information processing in the society is necessarily inefficient as the information contained in these agents' initial signals is inevitably lost. Assuming convergence, let $\gamma_{i,k}$ denote the long-term weight of the group $\mathcal{C}_k$ on the opinion of agent $i \in \mathcal{R}$, i.e. $x_i(\infty) = \sum_{k=1}^{K} \gamma_{i,k}\hat{\mu}_k$ (cf. (12)). This immediately translates into the wisdom of an agent $i \in \mathcal{R}$ as follows:

$$E((x_i(\infty) - \mu)^2) = \sum_{k=1}^{K} \gamma_{i,k}^2 \,\mathrm{MSE}_k \leq \max_{k=1,\dots,K} \mathrm{MSE}_k \,. \tag{19}$$

The wisdom of an agent in the rest of the world depends on the wisdom within the closed and strongly connected groups. More precisely, an agent $i$'s wisdom only depends on the wisdom of groups $\mathcal{C}_k$ to which there is a directed path in the network $T$ because this corresponds to $\gamma_{i,k} > 0$. The worst case for an agent in the rest of the world is to be influenced only by agents of one closed and strongly connected group with maximal MSE. With regard to the example discussed in subsection 5.1 this is the case for Players 7 and 8 who have directed paths only into group 1, respectively group 2 such that they share their MSEs of 4. Player 9, however, who has directed paths into both groups with MSE of 4 reaches an MSE of 2 since the long-term weights $\gamma_{9,1}$ and $\gamma_{9,2}$ are squared in (19). Finally, Player 10 has a directed path into these groups via Player 9 and, moreover, has a directed path into group 3. Player 10 is able to combine MSEs of 4, 4, and 2.25 into an MSE as low as 1.0625. It is intuitive that for maximal wisdom of a player in the rest of the world, all groups' signals have to be accessed with some kind of balanced group weights. The following proposition confirms this intuition.

**Proposition 7.** *For agents $i \in \mathcal{R}$, we have:*

$$E((x_i(\infty) - \mu)^2) \geq \frac{1}{\sum_{k=1}^{K} \frac{1}{\mathrm{MSE}_k}} \,, \tag{20}$$

*with equality if and only if* $\gamma_{i,k} = \frac{\frac{1}{\mathrm{MSE}_k}}{\sum_{l=1}^{K} \frac{1}{\mathrm{MSE}_l}}$ *for all* $k = 1, \dots, K$.

Therefore, the highest wisdom is achieved if an agent in the rest of the world averages the different groups' opinions in such a way that the product of weight put on a group and its MSE is constant for all groups: the better a group's estimate, the more weight it should get. Nevertheless, as all the optimal weights are positive, this optimum can only be achieved if from agent $i$ there is a directed path into all the closed and strongly connected groups.Notice also that the optimal weights depend on the groups' MSEs, so an agent in the rest of the world who is initially characterized by optimal weights would no longer average the groups' opinions optimally if conformity levels within the groups were to change. It is remarkable that an agent in the rest of the world who is connected to multiple groups can reach a significantly lower MSE than the best informed agents ('experts') from those groups. Thus, the fact that agents in the rest of the world are absolutely powerless does not imply that they are not wise.

# 6 Concluding Remarks

In this paper we present a model of opinion formation that is based on the model by DeGroot (1974) but incorporates the fact that individuals may state an opinion that is different from their true opinion. Thereby the individuals interact strategically in every discussion round. The incentive to depart from the true opinion is given by each individual's preference for conformity. A highly conforming player will state an opinion that is close to her peer-group's opinion, while non-conforming players exaggerate their true opinion to countervail the opinion of others (Proposition 1). Hence, in addition to *informational social influence* modeled by naïve updating through the network $T$ as in the DeGroot model, we also model *normative social influence* by including conforming/counter–conforming behavior. In particular, players are heterogeneous with respect to their network position and their degree of conformity. If the degree of conformity of all players is zero, then the DeGroot model is obtained as a special case of our model.

We elaborate on the long–run implications and convergence to a steady state of repeated play of Nash equilibrium. This implies that the law of motion of opinion profiles is given by a time–homogeneous matrix $M$ which is a transformation of the interaction structure $T$, but preserves the group structure (Proposition 3). Steady states are then characterized by opinion profiles such that in each closed and strongly connected group all players have the same opinion, i.e. reach a consensus (Proposition 2). A sufficient condition for convergence to a steady state is given if all players have at least some self–confidence (positive diagonal of $T$) and the conformity parameter of all players is non–negative (Proposition 5).

As the main result of this paper, we show to which steady state the opinion dynamics under conformity converges (Theorem 1). The players in each closed and strongly connected group reach a consensus in the long–run which is characterized by a weighted average of the initial opinions such that the averaging weights depend on network centrality and degree of conformity. We interpret this result in terms of opinion leadership and wisdom within groups.

Opinion leaders are those whose initial opinion has high influence on consensus. We show that influence of a given player is increasing in network centrality, increasing in other players' conformity and decreasing in own degree of conformity (Corollary 1). Taking the network as given, we conclude that low conformity fosters opinion leadership while high conformity undermines opinion leadership. Therefore, counter–conformity can also be interpreted as a persuasion device since not only the connected players' opinions of next period are influenced towards own opinion but a higher impact on the consensus opinion is achieved.

We also elaborate on a phenomenon that is often called "wisdom of the crowds" which occurs when aggregation of individual opinions yields an accurate estimate if the group size is large. In our context, we particularly ask how information aggregation within a given group is affected by the individual degrees of conformity, keeping the group size fixed. We assume that each player's initial opinion is a noisy but unbiased signal about some true state of the world such that the players are heterogeneous with respect to signal precision (the inverse of the variance). Wisdom of a given group is then defined as the mean squared error of the consensus opinion. We show the comparative static effects of varying individuals' conformity parameter on wisdom (Proposition 6) and find that increasing conformity of players with high ratio of network centrality over signal precision as compared to the group's average decreases the group's MSE, and thereby increases its wisdom. In particular, optimal wisdom within a given closed and strongly connected group is achieved if distribution of conformity degrees is such that this ratio is balanced (Corollary 2). Since the conformity of players in the rest of the world does not affect consensus, we analyze how the network structure affects the wisdom within this group. We find that although players in the rest of the world are powerless in terms of influence, they can be quite wise if they are connected to many wise closed and strongly connected groups.

The model presented here contains some simplifying assumptions which may be relaxed in future research. First, we assumed that the social network is exogenous and stays fixed over time. In the literature we can find models where the network structure may vary over time such that only players with "close opinions" are trusted (Hegselmann and Krause, 2002), self confidence varies (DeMarzo et al., 2003), and general changes are possible (Lorenz, 2005). It would be interesting to see how changes in the interaction structure, either exogenously or endogenously, affect our results. Second, we assumed that the network which determines how the true opinions are influenced and the network which determines how the stated opinions are formed (i.e. the network which determines the players to which a given player conforms) coincide. If such an assumption is relaxed, the group structure may no longer be preserved. Moreover, interesting applications to lobbying (addressing a certain group) are imaginable. We leave these ideas and possible extensions to future research.

# 6 Appendix: Proofs

## 6.1 Nash Equilibrium

### Proof of Proposition 1

*Proof.* Given some opinion profile $x \in \mathbb{R}^n$, let $s_i^*$ denote the best reply of player $i$ to the strategy profile $s \in \mathbb{R}^n$. Note that the best reply is unique and satisfies the first order condition:

$$\frac{\partial u_i(s_i, s_{-i}|x_i)}{\partial s_i}\Big|_{s_i=s_i^*} = -2\alpha_i\left(s_i^* - x_i\right) - 2\beta_i\left(s_i^* - \sum_{j\neq i}\frac{t_{ij}}{(1-t_{ii})}s_j\right) = 0,$$

for all $i \in \mathcal{N}$. A strategy profile is a Nash equilibrium if and only if $s_i^*$ is a best reply to $s^* \in \mathbb{R}^n$. Thus using the notation introduced above and letting $\delta_i = \frac{\beta_i}{\alpha_i+\beta_i}$ we get that Nash equilibria $s^* \in \mathbb{R}^n$ satisfy:

$$(I-\Delta)(s^*-x) + \Delta\left(I-(I-D)^{-1}(T-D)\right)s^* = 0$$

Note that $I - D$ is invertible since $D$ is by assumption diagonal with entries $0 \leq t_{ii} < 1$. Further, since $(I-D)^{-1}(T-D)$ is row stochastic, and since $\Delta$ has all entries $|\delta_{ii}| < 1$ we also have that $(I - \Delta(I-D)^{-1}(T-D))$ invertible.[17] Thus, we get a unique solution to the first order condition given by,

$$s^* = \left(I - \Delta(I-D)^{-1}(T-D)\right)^{-1}(I-\Delta)x.$$

□

## 6.2 Rewriting I-M

**Lemma 2** (I-M)**.** *$I - M = \left(I-(T-D)\Delta(I-D)^{-1}\right)^{-1}(I-T)$.*

### Proof of Lemma 2 (I-M).

First, we can rewrite $M$, given by (3), to obtain

$$M = T - (T-D)(I-\Delta(I-D)^{-1}(T-D))^{-1}\Delta(I-(I-D)^{-1}(T-D)),$$

This can be verified by the following calculation.

$$\begin{aligned} M &= D + (T-D)(I-\Delta(I-D)^{-1}(T-D))^{-1}(I-\Delta) \\ &= D + (T-D)\big[I-\Delta(I-D)^{-1}(T-D)\big]^{-1}\big[I-\Delta(I-D)^{-1}(T-D) \\ &\quad + \Delta(I-D)^{-1}(T-D) - \Delta\big] \\ &= D + (T-D)(I + \big[I-\Delta(I-D)^{-1}(T-D)\big]^{-1}\big[\Delta(I-D)^{-1}(T-D)-\Delta)\big] \\ &= T - (T-D)\big[I-\Delta(I-D)^{-1}(T-D)\big]^{-1}\Delta\big[I-(I-D)^{-1}(T-D)\big]. \end{aligned}$$

[17] Since $Y := (I-D)^{-1}(T-D)$ is row stochastic we have $|\lambda_i(Y)| \leq 1$ for all eigenvalues $\lambda_i$ of $Y$. Multiplication with a diagonal matrix $\Delta$ with entries $|\delta_{ii}| < 1$ implies that $|\lambda_i(\Delta Y)| < 1$ for all eigenvalues $\lambda_i$ of $\Delta Y$, see Ostrowski (1959), Theorem 1. Thus, $I - \Delta Y$ is non-singular since there exists no 0-eigenvalue.

Thus,

$$
\begin{aligned}
I - M &= I - T + (T-D)\big[I - \Delta(I-D)^{-1}(T-D)\big]^{-1}\Delta(I-D)^{-1}(I-T) \\
&= \Big(I + (T-D)\left[I - \Delta(I-D)^{-1}(T-D)\right]^{-1}\Big)\,\Delta(I-D)^{-1}(I-T) \qquad (21)
\end{aligned}
$$

Now, note that for any $n \times m$-matrix $A$ and any $m \times n$-matrix $B$ we have that $I_n - AB$ is invertible if and only if $I_m - BA$ is invertible, and then $(I_n - AB)^{-1} = I_n + A(I_m - BA)^{-1}B$, since $(I_n + A(I_m - BA)^{-1}B)(I_n - AB) = I_n - AB + A(I_m - BA)^{-1}B - A(I_m - BA)^{-1}BAB = I_n - AB + A(I_m - BA)^{-1}(I_m - BA)B = I_n$. Here, $I_k$ denotes the $k$-dimensional identity matrix, $k \in \{n, m\}$. Taking $A = T - D$ and $B = \Delta(I-D)^{-1}$ in (21) then gives $I - M = \left(I - (T-D)\Delta(I-D)^{-1}\right)^{-1}(I - T)$. □

## 6.3 Steady states

**Proof of Proposition 2**

1. $x$ is a steady state of $T$, i.e. $Tx = x \;\Leftrightarrow\; (I-T)x = 0$ if and only if $\big[I - (T - D)\Delta(I-D)^{-1}\big]^{-1}(I-T)x = 0$, since by Lemma 2 $\big[I - (T-D)\Delta(I-D)^{-1}\big]$ is invertible. Thus by Lemma 2, $Tx = x$ if and only if $Mx = x$.

   It therefore suffices to show that $Mx = x \Rightarrow q = x \Rightarrow q = s \Rightarrow Mx = x$.

   (a) $x = Mx = Dx + (I-D)q$ implies $(I-D)x = (I-D)q$, thus $q = x$.

   (b) $q = x$ implies $s = (I-\Delta)x + \Delta q = (I-\Delta)q + \Delta q = q$.

   (c) $q = s$ implies $s = (I-\Delta)x + \Delta q = (I-\Delta)x + \Delta s$ and therefore $(I-\Delta)s = (I-\Delta)x$ and $s = q = x$, from which we find $Mx = Dx + (I-D)q = Dx + (I-D)x = x$.

2. Suppose $x = s$. Note that $s = \left(I - \Delta(I-D)^{-1}(T-D)\right)^{-1}(I-\Delta)x$ by Proposition 1. Thus

$$
\begin{aligned}
& x = s \\
\Leftrightarrow\quad & \left(I - \Delta(I-D)^{-1}(T-D)\right)x = (I-\Delta)x \\
\Leftrightarrow\quad & \Delta\left(I - (I-D)^{-1}(T-D)\right)x = 0 \\
\Leftrightarrow\quad & \Delta(I-D)^{-1}\left(I - D - (T-D)\right)x = 0 \\
\overset{(*)}{\Leftrightarrow}\quad & (I-D)^{-1}\Delta\left(I - D - (T-D)\right)x = 0 \\
\Leftrightarrow\quad & \Delta\left(I - T\right)x = 0.
\end{aligned}
$$

   where (*) holds, since $(I-D)^{-1}$ and $\Delta$ are diagonal. □

**Lemma 3.** *The following are equivalent:*

1. *True opinions $x(t)$ converge for $t \to \infty$.*

2. *Stated opinions $s(t)$ converge for $t \to \infty$.*

*3. Perceived opinions $q(t)$ converge for $t \to \infty$.*

*Moreover, if the true, stated and perceived opinions converge then the* $\lim_{t\to\infty} x(t) = \lim_{t\to\infty} s(t) = \lim_{t\to\infty} q(t)$.

**Proof of Lemma 3**

From Proposition 1 we get that $s(t) = \left(I - \Delta(I-D)^{-1}(T-D)\right)^{-1}(I-\Delta)x(t)$. Thus convergence of $x(t)$ implies convergence of $s(t)$. By definition we have that $q(t) = (I - D)^{-1}(T-D)s(t)$, and hence convergence of $s(t)$ implies convergence of $q(t)$. To see that convergence of $q(t)$ implies convergence of $x(t)$ consider $(I-D)q(t) = (T-D)s(t)$ and $x(t+1) = Dx(t) + (T-D)s(t) = Dx(t) + (I-D)q(t)$, by definition. For all $t \geq 0$, this implies $x(t) = D^t x(0) + \sum_{l=0}^{t-1} D^{t-1-l}(I-D)q(l)$, the first part of which converges to 0 because all elements of the diagonal matrix $D$ belong to $[0,1)$. The limit of $x(t)$ therefore equals

$$\lim_{t\to\infty}\sum_{l=0}^{t-1} D^{t-1-l}(I-D)q(l) = \lim_{t\to\infty}\sum_{l=0}^{t-1} D^{t-1-l}(I-D)\left(q(l)-q(\infty)\right) + \lim_{t\to\infty}\sum_{l=0}^{t-1} D^{t-1-l}(I-D)q(\infty).$$

First of all, notice that the second limit obviously is $q(\infty)$, because $\sum_{l=0}^{\infty} D^l = (I-D)^{-1}$. For the first limit, note that for any $\varepsilon > 0$, we can find an index $l_\varepsilon$ such that we have $||q(l) - q(\infty)|| < \varepsilon$ for all $l > l_\varepsilon$. Splitting the sum into small $l$ ($l \leq l_\varepsilon$) and large $l$ ($l > l_\varepsilon$), we then easily see that the first term converges to 0, so that all in all, $x(t)$ converges to $q(\infty)$. Since $s(t) = (I-\Delta)x(t) + \Delta q(t)$, also $s(t)$ shares the same limit. □

## 6.4 Block structure

**Proof of Proposition 3.**

Let $Z := [I - \Delta(I-D)^{-1}(T-D)]^{-1}(I-\Delta)$ to simplify $s^* = Zx$ and $M = D + (T-D)Z$. We now proceed in three steps: we first characterize $Z$, then $M$, and finally $M^t$. Let $T$ be given as in (8). Then simple but tedious block matrix algebra together with Lemma 2 yields:

1.
$$Z = \begin{pmatrix} Z_{11} & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & Z_{KK} & 0 \\ Z_{\mathcal{R}1} & \cdots & \cdots & Z_{\mathcal{R}K} & Z_{\mathcal{R}\mathcal{R}} \end{pmatrix}$$

with

$$Z_{kk} = (I - \Delta_{kk}(I-D_{kk})^{-1}(T_{kk}-D_{kk}))^{-1}(I-\Delta_{kk}),$$
$$Z_{\mathcal{R}k} = Z_{\mathcal{R}\mathcal{R}}(I-\Delta_{\mathcal{R}\mathcal{R}})^{-1}\Delta_{\mathcal{R}\mathcal{R}}(I-D_{\mathcal{R}\mathcal{R}})^{-1}T_{\mathcal{R}k}Z_{kk}$$

for all $k = 1, \ldots, K$, and

$$Z_{\mathcal{R}\mathcal{R}} = (I - \Delta_{\mathcal{R}\mathcal{R}}(I-D_{\mathcal{R}\mathcal{R}})^{-1}(T_{\mathcal{R}\mathcal{R}}-D_{\mathcal{R}\mathcal{R}}))^{-1}(I-\Delta_{\mathcal{R}\mathcal{R}}).$$

2. For $M = D + (T-D)Z = I - \left(I - (T-D)\Delta(I-D)^{-1}\right)^{-1}(I-T)$, we get

$$M = \begin{pmatrix} M_{11} & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & M_{KK} & 0 \\ M_{\mathcal{R}1} & \cdots & \cdots & M_{\mathcal{R}K} & M_{\mathcal{R}\mathcal{R}} \end{pmatrix}$$

with

$$\begin{aligned} M_{kk} &= D_{kk} + (T_{kk} - D_{kk})(I - \Delta_{kk}(I - D_{kk})^{-1}(T_{kk} - D_{kk}))^{-1}(I - \Delta_{kk}) \\ &= I - \left(I - (T_{kk} - D_{kk})\Delta_{kk}(I - D_{kk})^{-1}\right)^{-1}(I - T_{kk}), \end{aligned}$$

$$\begin{aligned} M_{\mathcal{R}k} &= T_{\mathcal{R}k}Z_{kk} + (T_{\mathcal{R}\mathcal{R}} - D_{\mathcal{R}\mathcal{R}})Z_{\mathcal{R}k} \\ &= (I - (T_{\mathcal{R}\mathcal{R}} - D_{\mathcal{R}\mathcal{R}})\Delta_{\mathcal{R}\mathcal{R}}(I - D_{\mathcal{R}\mathcal{R}})^{-1})^{-1}T_{\mathcal{R}k}Z_{kk} \end{aligned}$$

for all $k = 1, \dots, K$, and

$$\begin{aligned} M_{\mathcal{R}\mathcal{R}} &= D_{\mathcal{R}\mathcal{R}} + (T_{\mathcal{R}\mathcal{R}} - D_{\mathcal{R}\mathcal{R}})(I - \Delta_{\mathcal{R}\mathcal{R}}(I - D_{\mathcal{R}\mathcal{R}})^{-1}(T_{\mathcal{R}\mathcal{R}} - D_{\mathcal{R}\mathcal{R}}))^{-1}(I - \Delta_{\mathcal{R}\mathcal{R}}) \\ &= I - \left(I - (T_{\mathcal{R}\mathcal{R}} - D_{\mathcal{R}\mathcal{R}})\Delta_{\mathcal{R}\mathcal{R}}(I - D_{\mathcal{R}\mathcal{R}})^{-1}\right)^{-1}(I - T_{\mathcal{R}\mathcal{R}}). \end{aligned}$$

3. Finally, we claim that for every $t \in \mathbb{N}^*$,

$$M^t = \begin{pmatrix} M_{11}^t & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & M_{KK}^t & 0 \\ (M^t)_{\mathcal{R}1} & \cdots & \cdots & (M^t)_{\mathcal{R}K} & M_{\mathcal{R}\mathcal{R}}^t \end{pmatrix}$$

with $(M^t)_{\mathcal{R}k} = \sum_{l=0}^{t-1} M_{\mathcal{R}\mathcal{R}}^l M_{\mathcal{R}k} M_{kk}^{t-1-l}$ for all $k = 1, \dots, K$.

The assertion for the diagonal elements $M_{11}^t, \dots, M_{KK}^t$ and $M_{\mathcal{R}\mathcal{R}}^t$ is trivial. We prove the formula for $M_{\mathcal{R}k}^t$ by induction:

- For $t = 1$, the assertion is trivial.
- $t \mapsto t+1$: first, we have $M_{\mathcal{R}k}^{t+1} = (M^t M)_{\mathcal{R}k} = M_{\mathcal{R}k}^t M_{kk} + M_{\mathcal{R}\mathcal{R}}^t M_{\mathcal{R}k}$ by simple matrix multiplication. Inserting $(M^t)_{\mathcal{R}k} = \sum_{l=0}^{t-1} M_{\mathcal{R}\mathcal{R}}^l M_{\mathcal{R}k} M_{kk}^{t-1-l}$, we find

$$M_{\mathcal{R}k}^{t+1} = \left(\sum_{l=0}^{t-1} M_{\mathcal{R}\mathcal{R}}^l M_{\mathcal{R}k} M_{kk}^{t-1-l}\right) M_{kk} + M_{\mathcal{R}\mathcal{R}}^t M_{\mathcal{R}k} = \sum_{l=0}^{t+1-1} M_{\mathcal{R}\mathcal{R}}^l M_{\mathcal{R}k} M_{kk}^{t+1-1-l}, \tag{22}$$

which concludes the proof. □

## 6.5 Convergence

**Proof of Proposition 4**

1. 'Only if': this is proven in the first part of the proof of Proposition 5.

2. 'If': Now suppose each $M_{kk}$ converges and $M_{\mathcal{RR}}$ converges to $\mathbf{0}$. First, since $M_{kk}$ converges, its only eigenvalue with $|\lambda| \geq 1$ is $\lambda = 1$ with algebraic and geometric multiplicity equal to 1 for every $k = 1, \ldots, K$. On the other hand, $M^t_{\mathcal{RR}} \to \mathbf{0}$ implies that the eigenvalues of $M_{\mathcal{RR}}$ are all smaller than 1 in absolute value and, thus, $M_{\mathcal{RR}} - \lambda I$ is invertible for all complex numbers $\lambda$ with $|\lambda| \geq 1$.

   Now let the complex number $\tilde{\lambda}$ be either outside of the unit circle ($|\tilde{\lambda}| > 1$) or exactly on the unit circle ($|\tilde{\lambda}| = 1$), but different from 1. Denoting $x = (x_{11}, \ldots, x_{KK}, x_{\mathcal{RR}})$ and taking into account the block structure of $M$, we easily see that any solution of $(M - \tilde{\lambda}I)x = 0$ must satisfy $x_{11} = 0, \ldots, x_{KK} = 0$ and therefore also $x_{\mathcal{RR}} = 0$, so that we can conclude that $\lambda = 1$ is the only possible eigenvalue of $M$ with $|\lambda| \geq 1$.

   In order to show convergence of $M^t$, we therefore have to show that algebraic and geometric multiplicity of $\lambda = 1$ coincide. With regard to algebraic multiplicity, the block structure of $M$ implies $\det(M - \lambda I) = \prod_{k=1}^{K} \det(M_{kk} - \lambda I) \det(M_{\mathcal{RR}-\lambda I})$, such that the algebraic multiplicity of $\lambda = 1$ is the sum of the algebraic multiplicities of $M_{11}, \ldots, M_{KK}$ and $M_{\mathcal{RR}}$, which are given by 1 and 0, respectively, since $M_{kk}$ is by definition irreducible for all $k = 1, ..., K$. All in all, the algebraic multiplicity equals $K$. With regard to geometric multiplicity, the block structure of $M$ implies that every vector of the form $(c_1\mathbb{1}, \ldots, c_K\mathbb{1}, (I - M_{\mathcal{RR}})^{-1} \sum_{k=1}^{n} c_k M_{\mathcal{R}k}\mathbb{1})$ with constants $c_1, \ldots, c_K$ is an eigenvector to $M$ for $\lambda = 1$, implying that the geometric multiplicity is at least $K$, thereby concluding the proof. □

**Proof of Proposition 5**

Denote $Y := (I - D)^{-1}(T - D)$ which is row stochastic. Thus, if $|\delta_i| < 1$ for all $i \in \mathcal{N}$ we have that $[I - \Delta Y]$ is invertible and $[I - \Delta Y]^{-1} = \sum_{k=0}^{\infty} (\Delta Y)^k$. Moreover, if $\delta_i \geq 0$ for all $i \in \mathcal{N}$ the sum $\sum_{k=0}^{\infty} (\Delta Y)^k$ is a sum of non-negative matrices, implying that $[I - \Delta Y]^{-1}$ is non-negative. Hence $M = \Big[D + (T - D)[I - \Delta Y]^{-1}(I - \Delta)\Big]$ is non-negative since it is the product of non-negative matrices (since $0 < t_{ii} < 1$) added to $D$, which is a diagonal matrix with strictly positive entries ($0 < t_{ii}$). Finally, since $M\mathbb{1} = \mathbb{1}$ by Lemma 2 we get that $M$ is row stochastic. Since the diagonal of $D$ is strictly positive, we get that the diagonal of $M$ is strictly positive, $m_{ii} > 0$, implying aperiodicity of $M$. Thus $M^t$ converges. □

## 6.6 Long–run

To prove Theorem 1, the following Lemma is helpful.

**Lemma 4** (Convergence to Eigenvector)**.** *Let $A$ be an $n \times n$-matrix with $A\mathbb{1} = \mathbb{1}$ and* $\mathrm{rk}(I - A) = n - 1$. *If $A^t$ converges to $A^\infty$ for $t \to \infty$, then $A^\infty = \mathbb{1}w'$, with $w'$ the unique normalized left eigenvector of $A$ associated with the eigenvalue* 1.

**Proof of Lemma 4**
Obviously, $AA^\infty = A^\infty = A^\infty A$. This implies that

- the columns of $A^\infty$ must be multiples of $\mathbb{1}$,
- the rows of $A^\infty$ must be multiples of $w'$,

from which we find $A^\infty = c\,\mathbb{1}w'$ for some real number $c$ which is found to be equal to 1 as $\mathbb{1} = A^\infty\mathbb{1} = c\,\mathbb{1}w'\mathbb{1} = c\,\mathbb{1}$. $\square$

**Proof of Theorem 1**
We first derive the formula for $M^\infty_{kk}$. Then we turn to $M^\infty_{\mathcal{R}\mathcal{R}}$ and $M^\infty_{\mathcal{R}k}$.
Let $v'$ denote the unique normalized left eigenvector of $M$ associated with the eigenvalue 1, i.e. $v'M = v'$ such that $v'\mathbb{1} = 1$. Moreover, assume for the moment that $\text{rk}(I-T) = n-1$. Then as $v'(M-I) = 0$, we have due to Lemma 2

$$0 = v'(I-M) = v'\left(I-(T-D)\Delta(I-D)^{-1}\right)^{-1}(I-T),$$

implying

$$v'\left(I-(T-D)\Delta(I-D)^{-1}\right)^{-1} = r\,w'$$

for some real number $r$. Using $w'T = w'$, we then find

$$v' = r\,w'\left(I-(T-D)\Delta(I-D)^{-1}\right) = r\,w'\left(I-(I-D)\Delta(I-D)^{-1}\right) = r\,w'(I-\Delta).$$

The normalization of $v$ then entails $r = \frac{1}{w'(I-\Delta)\mathbb{1}}$, which shows that $v = \dfrac{(I-\Delta)w}{\mathbb{1}'(I-\Delta)w}$. Now, relaxing the assumption $\text{rk}(I-T) = n-1$, the formula for $M^\infty_{kk}$ follows.

Furthermore, $MM^\infty x = M^\infty x$ and therefore due to Proposition 2, $TM^\infty x = M^\infty x$ for all $n$-dimensional vectors $x$, delivering $TM^\infty = M^\infty$. This implies

- $M^\infty_{\mathcal{R}\mathcal{R}} = T_{\mathcal{R}\mathcal{R}}M^\infty_{\mathcal{R}\mathcal{R}}$ and therefore $(I-T_{\mathcal{R}\mathcal{R}})M^\infty_{\mathcal{R}\mathcal{R}} = 0$, entailing $M^\infty_{\mathcal{R}\mathcal{R}} = 0$ because $I - T_{\mathcal{R}\mathcal{R}}$ is invertible,
- $M^\infty_{\mathcal{R}k} = T_{\mathcal{R}k}M^\infty_{kk} + T_{\mathcal{R}\mathcal{R}}M^\infty_{\mathcal{R}k}$, and therefore $M^\infty_{\mathcal{R}k} = (I-T_{\mathcal{R}\mathcal{R}})^{-1}T_{\mathcal{R}k}M^\infty_{kk}$. $\square$

**Proof of Corollary 1** The expression follows straightforwardly from Theorem 1. For the comparative statics consider,

$$\begin{aligned}
\frac{\partial v_i}{\partial \delta_k} &= \frac{\partial \frac{w_i(1-\delta_i)}{\sum_{j=1}^n w_j(1-\delta_j)}}{\partial \delta_k} \\
&= \frac{w_k}{\sum_{j=1}^n w_j(1-\delta_j)}\left(\frac{w_i(1-\delta_i)}{\sum_{j=1}^n w_j(1-\delta_j)} - 1_{i=k}\right) \\
&= \frac{w_k}{\sum_{j=1}^n w_j(1-\delta_j)}\left(v_i - 1_{i=k}\right). \quad \square
\end{aligned}$$

## 6.7 Wisdom

**Proof of Lemma 1**

First of all, $\hat{\mu}_k$ is easily seen to be unbiased for $\mu$ because $\sum_{i\in\mathcal{C}_k} v_i = 1$. Therefore, its MSE equals its variance which is given by $\sum_{i\in\mathcal{C}_k} v_i^2\sigma_i^2$ as the $x_i(0)$ are uncorrelated. □

**Proof of Proposition 6**

$$\frac{\partial\,\mathrm{MSE}_k}{\partial\delta_i} = \frac{\partial \sum_{j\in\mathcal{C}_k} v_j^2\sigma_j^2}{\partial\delta_i} = \sum_{j\in\mathcal{C}_k} 2\sigma_j^2 v_j \frac{\partial v_j}{\partial\delta_i} \overset{(14)}{=} \frac{2w_i}{\sum_{j\in\mathcal{C}_k} w_j(1-\delta_j)} \sum_{j\in\mathcal{C}_k} \sigma_j^2 v_j\left(v_j - 1_{j=i}\right).$$

The assertion follows easily when noting that $\mathrm{MSE}_k = \sum_{j\in\mathcal{C}_k} v_j v_j \sigma_j^2$. □

**Proof of Proposition 7** First, notice that $E((x_i(\infty)-\mu)^2) = \sum_{k=1}^{K} \gamma_{i,k}^2\,\mathrm{MSE}_k$, with $\sum_{k=1}^{K}\gamma_{i,k} = 1$ for all $i \in \mathcal{R}$. As the assertion is trivial for $K=1$, we assume $K>1$ and replace $\gamma_{i,K}$ by $1-\sum_{k=1}^{K-1}\gamma_{i,K}$. We can then understand

$$\sum_{k=1}^{K}\gamma_{i,k}^2\,\mathrm{MSE}_K = \sum_{k=1}^{K-1}\gamma_{i,k}^2\,\mathrm{MSE}_k + \left(1-\sum_{k=1}^{K-1}\gamma_{i,k}\right)^2 \mathrm{MSE}_K$$

as a function $f$ of $(\gamma_{i,1},\ldots,\gamma_{i,K-1})$, defined on the convex and compact set of all $(\gamma_{i,1},\ldots,\gamma_{i,K-1})$ with $\gamma_{i,k}\geq 0$ for all $k=1,\ldots,K-1$ and $\sum_{k=1}^{K-1}\gamma_{i,k}\leq 1$. We then easily find

$$\frac{\partial f}{\partial\gamma_{i,l}} = 2\gamma_{i,l}\,\mathrm{MSE}_l - 2\,\mathrm{MSE}_K\left(1-\sum_{k=1}^{K-1}\gamma_{i,K}\right) \tag{23}$$

and $\frac{\partial^2 f}{\partial\gamma_{i,l}\partial\gamma_{i,m}} = 2\left(\mathrm{MSE}_l\,1_{l=m} + \mathrm{MSE}_K\right)$. From the last expression, we find the Hessian of $f$ to be constant and positive definite, implying that $f$ is a strictly convex function defined on a compact set. Therefore, $f$ is known to take its unique minimum at the solution of the FOC's (23) if such a solution exists. As $\gamma_{i,k} = \frac{1}{\mathrm{MSE}_k \sum_{l=1}^{K}\frac{1}{\mathrm{MSE}_l}}$ delivers a solution, the proof is complete. □

# References


Ariely, D. and Levav, J. (2000). Sequential choice in group settings: Taking the road less traveled and less enjoyed. *Journal of Consumer Research*, 27(3):279–90.

Asch, S. (1955). Opinions and social pressure. *Scientific American*, 193(5):31–35.

Bala, V. and Goyal, S. (1998). Learning from neighbours. *Review of Economic Studies*, 65(3):595–621.

Bala, V. and Goyal, S. (2001). Conformism and diversity under social learning. *Economic Theory*, 17(1):101–120.

Berger, R. L. (1981). A Necessary and Sufficient Condition for Reaching Consensus Using DeGroot's Method. *Journal of the American Statistical Association*, 76:415–418.

Bernheim, B. D. (1994). A theory of conformity. *Journal of Political Economy*, 102(5):841–77.

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

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *The Journal of Mathematical Sociology*, 2(1):113–120.

Buechel, B., Hellmann, T., and Pichler, M. M. (2011). The dynamics of continuous cultural traits in social networks. Working Papers 457, Bielefeld University, Institute of Mathematical Economics.

Chan, K. K. and Misra, S. (1990). Characteristics of the opinion leader: A new dimension. *Journal of Advertising*, 19(3):pp. 53–60.

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

DeMarzo, P. M., Vayanos, D., and Zwiebel, J. (2003). Persuasion bias, social influence, and unidimensional opinions. *Quart. J. Econ.*, 118(3):909–968.

Deutsch, M. and Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgement. *Journal of Abnormal Psychology*, 51(3):629–636.

Ellison, G. and Fudenberg, D. (1993). Rules of thumb for social learning. *Journal of Political Economy*, 101(4):612–43.

Ellison, G. and Fudenberg, D. (1995). Word-of-mouth communication and social learning. *The Quarterly Journal of Economics*, 110(1):93–125.

Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row, Peterson.

Flache, A. and Torenvlied, R. (2004). When Will They Ever Make up Their Minds? The Social Structure of Unstable Decision Making. *Journal of Mathematical Sociology*, 28:171–196.

French, J. (1956). A formal theory of social power. *Psychological Review*, 63:181–194.

Friedkin, N. (1991). Theoretical foundations for centrality measures. *The American Journal of Sociology*.

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

Galeotti, A. and Goyal, S. (2009). Influencing the influencers: a theory of strategic diffusion. *The RAND Journal of Economics*, 40(3):509–532.

Golub, B. and Jackson, M. O. (2010). Naïve learning in social networks and the wisdom of crowds. *Amer. Econ. J.: Microecon.*, 2(1):112–49.

Grabisch, M. and Rusinowska, A. (2010). A model of influence in a social network. *Theory and Decision*, 69(1):69–96.

Grabisch, M. and Rusinowska, A. (2011). Influence functions, followers and command games. *Games and Economic Behavior*, 72(1):123–138.

Harary, F. (1959). Status and contrastatus. *Sociometry*, 22:23–43.

Hayakawa, H. and Venieris, Y. (1977). Consumer interdependence via reference groups. *Journal of Political Economy*, 85(3):pp. 599–616.

Hegselmann and Krause (2002). Opinion dynamics and bounded confidence, models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3):1–33.

Hogg, M. and Abrams, D. (1988). *Social Identifications: A Social Psychology of Intergroup Relations and Group Processes*. University paperbacks. Routledge.

Hornsey, M. J., Majkut, L., Terry, D. J., and McKimmie, B. M. (2003). On being loud and proud: Non-conformity and counter-conformity to group norms. *British Journal of Social Psychology*, 42(3):319–335.

Jones, S. R. G. (1984). *The economics of conformism*. B. Blackwell, Oxford; New York, NY.

Katz, E. and Lazarsfeld, P. (2005). *Personal Influence: The Part Played by People in the Flow of Mass Communications*. Transaction Publishers.

Lorenz, J. (2005). A stabilization theorem for dynamics of continuous opinions. *Physica A: Statistical Mechanics and its Applications*, 355(1):217–223.

Maslach, C., Stapp, J., and Santee, R. T. (1985). Individuation: Conceptual analysis and assessment. *Journal of Personality and Social Psychology*, 49:729–738.

Mueller-Frank, M. (2011). The Robustness of Consensus in Network-Based Non-Bayesian Learning Models. Technical report.

Ostrowski, A. M. (1959). A quantitative formulation of Sylvester's law of inertia. *Proc Natl Acad Sci U S A*, 45(5):740–744.

Snyder, C. R. and Fromkin, H. L. (1980). *Uniqueness, the human pursuit of difference*. Plenum Press, New York.

Tian, K. T., Bearden, W. O., and Hunter, G. L. (2001). Consumer's need for uniqueness: Scale development and validation. *Journal of Consumer Research*, 28(1):50–66.