Essays on dynamic social influence

Manuel Förster

Université Paris 1 Panthéon - Sorbonne

Université catholique de Louvain
Faculté des sciences économiques, sociales, politiques et de communication
Ecole des sciences économiques

ESSAYS ON DYNAMIC SOCIAL INFLUENCE

Manuel Förster

Thèse présentée et soutenue

à Paris le 10 juin 2014

en vue de l'obtention du grade de Docteur

en mathématiques appliquées de l'Université Paris 1 Panthéon - Sorbonne

et en sciences économiques et de gestion de l'Université catholique de Louvain

La thèse a été préparée dans le cadre du programme doctoral européen EDE-EM (European Doctorate in Economics – Erasmus Mundus), en cotutelle entre l'Université Paris 1 Panthéon - Sorbonne et l'Université catholique de Louvain

Directeurs/Promoteurs de thèse :

Prof. Agnieszka Rusinowska, Université Paris 1 Panthéon - Sorbonne
Prof. Vincent Vannetelbosch, Université catholique de Louvain
Prof. Michel Grabisch, Université Paris 1 Panthéon - Sorbonne (Codirecteur)

Composition du jury :

Prof. Dunia López-Pintado, Universidad Pablo de Olavide (Rapporteur)
Prof. Jean-Jacques Herings, Maastricht University (Rapporteur)
Prof. Ana Mauleon, Université Saint-Louis – Bruxelles
Prof. Francis Bloch, Université Paris 1 Panthéon - Sorbonne (Président)

This thesis has been carried out jointly

at the Centre d'Economie de la Sorbonne, Maison des Sciences Economiques, 106-112 Boulevard de l'Hôpital, 75647 Paris Cedex 13, France

and at the Center for Operations Research and Econometrics, Voie du Roman Pays 34, L1.03.01, B-1348 Louvain-la-Neuve, Belgium.

Cette thèse a été préparée conjointement

au Centre d'Economie de la Sorbonne, Maison des Sciences Economiques, 106-112 Boulevard de l'Hôpital, 75647 Paris Cedex 13, France

et au Center for Operations Research and Econometrics, Voie du Roman Pays 34, L1.03.01, B-1348 Louvain-la-Neuve, Belgique.

Meinen Eltern

Acknowledgements

Now, at the end of my Ph.D., the time has come to thank everyone who has helped me get there and who made my stays in Paris and Louvain-la-Neuve such a great experience.

First of all, I owe my sincere gratitude to my supervisors, Agnieszka Rusinowska, Vincent Vannetelbosch, and Michel Grabisch. It was a pleasure for me to work with them, I could not have wished for better and friendlier supervisors.

I want to thank Agnieszka and Michel for giving me such a warm welcome to Paris and the Centre d'Economie de la Sorbonne. They always listened carefully to my ideas and gave me the freedom to do the research of my liking and at my own rhythm. I am grateful for all their guidance and support. And more particularly I am thankful for their understanding of my decision to spend the last year of my Ph.D. in Belgium.

I want to thank Vincent for welcoming me at CORE in the second year of my Ph.D. He and Ana, who more than deserves to be mentioned here, as well always listened carefully to me and gave me the freedom to pursue my own ideas. I am grateful for all the discussions and their support during these two years. They also gave me the opportunity to visit regularly the Université Saint-Louis in Brussels, a place that turned out to be fruitful for my research.

Agnieszka, Michel, Vincent, and Ana, thank you for all!

I am also grateful to Dunia López-Pintado and Jean-Jacques Herings for agreeing to be part of my jury and for having read this thesis. They have made very helpful comments and gave me good advice for future research. Likewise, I would like to thank Francis Bloch for agreeing to be the chairman of my jury, and Pierre Dehez for replacing him in my pre-defense.

Furthermore, I would like to acknowledge the support from the doctoral program EDE-EM (European Doctorate in Economics – Erasmus Mundus) of the European Commission.

I am thankful to Berno Büchel, Tim Hellmann, Vladyslav Nora, Jan-Peter Siedlarek, Tom Truyts and many other seminar and conference participants for comments

and interesting discussions.

Special thanks also go to all my colleagues at the Centre d'Economie de la Sorbonne and at CORE for the friendly work environment. In particular, I am grateful to Andrew and Baris, with whom I shared an office at CORE in the last two years, to Stéphane for helping me with French translations, and to Caroline and Catherine for lots of help and advice.

My time in Paris and Louvain-la-Neuve (or rather Brussels) would not have been as pleasant without the great people that I can now call my friends, I will never forget our trips and evenings spent together, thank you very much guys!

Finally, I would like to thank my family, in particular my mum, my dad, and my brother. This work would not have been possible without their support. Mama, Papa und Tim, vielen Dank für Eure Unterstützung!

Abstract

This Ph.D. dissertation develops theories of dynamic social influence. In a dynamic framework, individuals interact repeatedly with their social environment and exchange beliefs and opinions on various economic, political and social issues.

In Chapter 2, we study influence processes modeled by ordered weighted averaging operators. These operators are anonymous: they only depend on how many agents share a belief. We find a necessary and sufficient condition for convergence to consensus and characterize outcomes where the society ends up polarized. Furthermore, we apply our results to fuzzy linguistic quantifiers.

In Chapter 3, we introduce the possibility of manipulation into the model by DeGroot (1974). We show that manipulation can modify the trust structure and lead to a connected society. Manipulation fosters opinion leadership, but the manipulated agent may even gain influence on the long-run beliefs. Finally, we investigate the tension between information aggregation and spread of misinformation.

In Chapter 4, we introduce conflicting interests into a model of non-Bayesian belief dynamics. Agents meet with their neighbors in the social network and exchange information strategically. With conflicting interests, the belief dynamics typically fails to converge: each agent's belief converges to some interval and keeps fluctuating on it forever.

Keywords: Influence, social networks, anonymity, manipulation, conflict of interest, consensus, belief fluctuations.

Résumé

Titre : Essais sur l'influence sociale dynamique.

Cette dissertation de doctorat développe des théories de l'influence sociale dynamique. Dans un cadre dynamique, les individus interagissent à plusieurs reprises avec leur environnement social et échangent leurs croyances et opinions sur différentes questions économiques, politiques et sociales.

Dans le Chapitre 2, nous étudions les processus d'influence modélisés par les moyennes ordonnées pondérées. Ces dernières sont anonymes : elles ne dépendent que du nombre d'agents qui partagent la même croyance. Nous exhibons une condition nécessaire et suffisante pour la convergence au consensus et caractérisons les résultats où la société se retrouve polarisée. Enfin, nous appliquons nos résultats aux quantificateurs linguistiques flous.

Dans le Chapitre 3, nous introduisons la possibilité de manipulation dans le modèle de DeGroot (1974). Nous montrons que la manipulation peut modifier la structure de confiance et mène à une société connectée. La manipulation promeut le leadership d'opinion, mais même l'agent manipulé peut gagner de l'influence sur les croyances à long terme. Finalement, nous étudions la tension entre l'agrégation d'informations et le déploiement de désinformations.

Dans le Chapitre 4, nous introduisons des conflits d'intérêt dans un modèle de dynamique de croyance non-bayésienne. Les agents se rencontrent avec leurs voisins dans le réseau social et échangent des informations stratégiquement. Avec des conflits d'intérêt, la dynamique de croyance ne converge pas en général : la croyance de chaque agent converge vers un certain intervalle et continue à fluctuer sur celui-ci pour toujours.

Mots clés : Influence, réseaux sociaux, anonymat, manipulation, conflit d'intérêt, consensus, fluctuations de croyance.

Résumé prolongé

L'influence sociale s'intéresse à comment nos croyances, opinions et sentiments sont affectés par d'autres. Provenant de la psychologie sociale, ce sujet a été étudié dans différents domaines, y compris l'économie.

Cette dissertation de doctorat développe des théories de l'influence sociale dynamique. Dans un cadre dynamique, les individus interagissent à plusieurs reprises avec leur environnement social. Pendant ces interactions, ils échangent leurs croyances et opinions sur différentes questions économiques, politiques et sociales.

Notre objectif est de contribuer à la littérature sur l'influence sociale dynamique non-bayésienne. Nous étudions trois aspects de l'influence sociale qui n'ont pas reçu beaucoup d'attention dans la littérature : l'influence anonyme, la manipulation, et les conflits d'intérêt. Nous analysons comment ces aspects affectent les croyances et opinions à long terme dans la société.

Dans le Chapitre 2, nous étudions les processus stochastiques d'influence modélisés par les moyennes ordonnées pondérées. Les agents ont une inclination (croyance) à dire « oui » ou « non » sur une question commune, et les croyances peuvent changer, dû à l'influence mutuelle entre agents. Chaque agent agrège à plusieurs reprises les croyances des autres agents et de lui-même en utilisant sa moyenne ordonnée pondérée. Ces dernières sont anonymes : elles ne dépendent que du nombre d'agents qui partagent la même croyance. Ceci permet, par exemple, d'étudier des situations semblables au vote par majorité, qui ne sont pas couvertes par les approches classiques basées sur les moyennes pondérées. Nous exhibons une condition nécessaire et suffisante pour la convergence au consensus et caractérisons les résultats où la société se retrouve polarisée. Nos résultats peuvent aussi être utilisés pour comprendre des situations plus générales, où les moyennes ordonnées pondérées ne sont utilisées que partiellement. Nous analysons la vitesse de convergence et les probabilités des différents résultats du processus. Enfin, nous appliquons nos résultats aux quantificateurs linguistiques flous, c.-à-d., des expressions comme « la plus grande partie » ou « au moins quelques ».

Dans le Chapitre 3, nous introduisons la possibilité de manipulation dans le

modèle de DeGroot (1974). Chaque agent a une croyance initiale sur une question commune. Les agents communiquent à plusieurs reprises avec leurs voisins dans le réseau social, peuvent exercer un effort afin de manipuler la confiance des autres, et mettent à jour leurs croyances par des moyennes pondérées des croyances des voisins. La motivation de manipulation est donnée par les préférences des agents. Nous montrons que la manipulation peut modifier la structure de confiance et mène à une société connectée. La manipulation promeut le leadership d'opinion, mais même l'agent manipulé peut gagner de l'influence sur les croyances à long terme. Nous notons que, dans une société suffisamment homophile, la manipulation accélère (ralentit) la convergence si elle diminue (augmente) l'homophilie. Finalement, nous étudions la tension entre l'agrégation d'informations et le déploiement de désinformations. Si la manipulation est plutôt coûteuse et les agents qui (ne) vendent (pas) bien leur informations perdent (gagnent) de l'influence globale, alors la manipulation réduit la désinformation et les agents convergent conjointement vers des croyances plus précises sur un état vrai sous-jacent.

Dans le Chapitre 4, nous introduisons des conflits d'intérêt dans un modèle de dynamique de croyance non-bayésienne. Les agents se rencontrent deux par deux avec leurs voisins dans le réseau social et échangent des informations stratégiquement. Nous démêlons les termes croyance (ce qui est considéré être) et opinion (ce qui devrait être, dû à un biais) : l'expéditeur de l'information voudrait propager son opinion (croyance biaisée), alors que le destinataire voudrait découvrir la croyance exacte de l'expéditeur. A l'équilibre, l'expéditeur ne communique qu'un message imprécis contenant des informations sur sa croyance. Le destinataire interprète le message envoyé et met à jour sa croyance par la moyenne de l'interprétation et sa croyance précédente. Avec des conflits d'intérêt, la dynamique de croyance ne converge pas en général : la croyance de chaque agent converge vers un certain intervalle et continue à fluctuer sur celui-ci pour toujours. Ces intervalles se confirment mutuellement : ils sont les combinaisons convexes des interprétations utilisées en communiquant, étant donné que tous les agents ont des croyances dans les intervalles correspondants.

Contents

Chapter 1

Introduction

1.1 Social Influence

Our beliefs, opinions and feelings shape our social interactions and behaviors. They are formed through personal experiences, observing the actions and experiences of others as well as through communication with others. *Social influence* is concerned with how these beliefs, opinions and feelings are affected by society.

Originating from social psychology, the topic has been intensively studied by different fields, among them economics. This work develops theories of *dynamic social influence*. We study how different aspects of social influence shape *beliefs* and *opinions* in society.

In a dynamic framework, individuals interact repeatedly with their social environment – often referred to as their *social network*. This may include friends, family, neighbors and coworkers as well as political actors, celebrities and news sources. During these interactions, individuals exchange their beliefs and opinions on various economic, political and social issues.

Our objective is to shed light on three particular aspects of social influence: *anonymous influence*, *manipulation*, and *conflicting interests*. We analyze how these aspects affect long-run beliefs and opinions in society.

1.2 State of the Art

Providing a complete survey of the literature on dynamic social influence is beyond the scope of this work. We discuss the two main lines of research on the topic in economics, which differ in the mechanism of information processing: in *Bayesian models* rational agents update their beliefs using Bayes' rule, while in *non-Bayesian models*

agents use some rather naïve rule to update their beliefs.[1] Our work contributes to the second stream of literature.

In models where agents are rational and update their beliefs using Bayes' rule, the common objective is to form a belief about (or to learn) an underlying state by aggregating information that is initially dispersed in society.[2,3] Gale and Kariv (2003) were the first to study Bayesian learning in a dynamic framework. Each agent starts with a prior (initial belief) about some underlying state and updates this belief by repeatedly communicating with her neighbors in a social network. The paper ignores strategic considerations of the agents, i.e., it is assumed that agents communicate truthfully, and show that connected societies converge to a consensus, but that in general this consensus will not be optimal.[4]

Acemoglu et al. (2014) study a model of Bayesian learning where they allow for *non-truthful communication*. The agents' objective is to form beliefs (acquire information) about an irreversible decision that each agent has to make, eventually. Each agent starts with an initial signal about the optimal decision and acquires additional information by repeatedly communicating with her neighbors in a social network. Each period, agents can decide whether to take the irreversible decision or to wait, where waiting is costly in the sense that their payoff from taking the right decision is discounted. Notice that in this setting agents might want to misreport their information in order to delay the decisions of other agents. They show that it is an equilibrium to report truthfully whenever truthful communication leads to asymptotic learning, i.e., the fraction of agents taking the right decision converges to 1 (in probability) as the society grows. Furthermore, they find that in some situations, misreporting can lead to asymptotic learning while truthful communication

[1] We focus on dynamic (repeated) models. For a survey on recent developments in Bayesian and non-Bayesian learning, see Acemoglu and Ozdaglar (2011).

[2] In this literature, the objective is often phrased as to learn the action that maximizes the agent's payoff. When agents maximize their static payoffs, we can interpret observable actions as truthful communication of beliefs.

[3] In a static framework, such models have been studied first by Banerjee (1992) and Bikhchandani et al. (1992). Later, Acemoglu et al. (2011) introduced a network structure to this model.

[4] Bala and Goyal (1998, 2001) study models of social experimentation in a dynamic framework. Boundedly rational agents repeatedly observe the actions and payoffs of their neighbors and update their beliefs on the optimal action using these observations. Their model differs from social learning models since agents learn from observing the outcome of experiments instead of trying to infer their neighbors' private information.

would not.[5,6]

Next, we discuss the stream of literature on dynamic social influence our work contributes to. In non-Bayesian models agents usually use some kind of "rule of thumb" to update their beliefs or change their beliefs in a way similar to being "infected" by a disease. A classical model where agents use a rule of thumb to update their beliefs was introduced in DeGroot (1974), see also French (1956) and Harary (1959) for antecedents. Each agent holds an initial belief about some common issue of interest (which might not be further specified, but could as well be an underlying state the agents would like to learn) and updates this belief by repeatedly communicating with her neighbors in a social network. This network is weighted such that the weight some agent places on another agent reflects the trust of the former in the latter agent. Each period, agents communicate truthfully with their neighbors and each agent's updated belief is the weighted average of her neighbors' beliefs (and possibly her own belief) from the previous period. In this model, the conditions for convergence to consensus are fairly weak: the agents' beliefs converge to a consensus whenever the social network is connected and some weak regularity condition is fulfilled.

The DeGroot model has been extensively studied in the literature. DeMarzo et al. (2003) were the first to study social learning in this framework. Each agent starts with an initial belief that is correlated with some underlying state. Furthermore, agents assign weights to their neighbors in a social network proportional to the precision of their initial beliefs. While this leads to optimal updating of beliefs in the beginning, agents fail to account for the repetition of information they receive in later periods. They refer to this phenomenon as persuasion bias and show that it implies that the agents' social influence depends not only on the precision of their signals, but also on their network position. This explains why information is aggregated non-optimally with the DeGroot updating rule.

Golub and Jackson (2010) study asymptotic learning in this model. Agents receive a noisy signal about the underlying state and communicate repeatedly with their neighbors using the DeGroot updating rule. They show that all beliefs in a large society converge to the underlying state if and only if the influence of the

[5]In an extension of the model they study endogenous formation of the social network. Initially, agents are split up in several social cliques, which are groups of agents linked at zero cost. To connect these cliques, agents need to form costly links. They show that sufficiently large cliques kill incentives to connect to other cliques and thus prevent asymptotic learning.

[6]Closely related to Acemoglu et al. (2014) are Hagenbach and Koessler (2010) and Galeotti et al. (2013), who study cheap-talk games on a network, but maintain the one-shot nature of cheap-talk games à la Crawford and Sobel (1982).

most influential agent vanishes as the society grows. Büchel et al. (2012) also study learning and introduce non-truthful communication to the model. Agents act strategically in the sense that they misreport their beliefs depending on their preferences for conformity. The paper finds that lower conformity fosters opinion leadership. In addition, the society becomes wiser if agents who are well informed are less conform, while uninformed agents conform more with their neighbors.

A related model is studied in Acemoglu et al. (2010). They investigate the tension between information aggregation and spread of misinformation in society. Each agent starts with identical information about the underlying state. Agents meet pairwise with their neighbors according to a stochastic process and update their beliefs by adopting the average of both beliefs. They introduce forceful agents who influence the beliefs of the other agents they meet, but almost do not change their own beliefs. They show that all beliefs converge to a stochastic consensus. Furthermore, they quantify the extent of misinformation by providing bounds on the gap between the consensus value and the benchmark without forceful agents where there is efficient information aggregation.

Grabisch and Rusinowska (2013) develop a framework where agents start with yes-no beliefs about some issue and update their beliefs by repeatedly communicating with the other agents. Their approach is more general than the models discussed above with respect to the mechanism used to aggregate the other agents' beliefs. They allow for arbitrary aggregation functions as the updating rule, which, for instance, allow to account for the influence of groups of agents. They characterize convergence of long-run beliefs in terms of influential coalitions (and agents).

The models discussed so far share a common feature: roughly speaking, agents reach a mutual consensus whenever the society is connected.[7] While this feature is desirable when we are interested in social learning since it allows to compare the consensus with the underlying state, this might be less the case in other situations. For instance, when we want to explain voting behavior or the evolution of public opinions on certain political issues, see, e.g., Kramer (1971) who documents large swings in US voting behavior within short periods, and works in social and political psychology that study how political parties and other organizations influence political beliefs, e.g., Cohen (2003); Zaller (1992).

Several authors have proposed models to explain non-convergence of beliefs, usu-

[7] Things are a bit different in Grabisch and Rusinowska (2013) due to their complex updating mechanism. Nevertheless, the mechanism's monotonicity property imposes a tendency towards consensus. Whether or not consensus will be attained in this model depends on the connectedness of the associated hypergraph of influence capturing which coalitions are influential for which agents.

ally incorporating some kind of homophily that leads to segregated societies and polarized beliefs.[8] Axelrod (1997) proposed such a model in a discrete belief setting, and later on Hegselmann and Krause (2002) and Deffuant et al. (2000) studied the continuous case, see also Lorenz (2005); Blondel et al. (2009); Como and Fagnani (2011). In these works, agents have "bounded confidence" in the sense that they only listen to agents that hold beliefs that are similar to their own beliefs. In other words, they disregard agents that are too different. This behavior typically leads to segregated societies, where each cluster of agents reaches a different consensus. Another approach is developed in Golub and Jackson (2012), who study societies consisting of different types of agents. They show that the presence of homophily, i.e., agents of the same type are well connected, while there are not many connections between the types, can substantially slow down convergence and thus lead to a high persistence of disagreement.

Though explaining persistent disagreement in society, the models discussed above cannot capture the phenomenon of *belief fluctuations* like the large swings in US voting behavior documented in Kramer (1971). Acemoglu et al. (2013) study a model where agents meet pairwise with their neighbors according to a stochastic process and update their beliefs by adopting the average of both beliefs. They introduce stubborn agents that never change their beliefs, which leads to fluctuating beliefs when the other agents update regularly from different stubborn agents. We can see these stubborn agents as a more extreme version of the forceful agents introduced in Acemoglu et al. (2010). While the latter are able to mislead the society in the sense that information aggregation is less efficient, stubborn agents can completely prevent information aggregation.

Finally, we briefly discuss non-Bayesian models where agents change their beliefs in a way similar to being "infected" by a disease. These models study the diffusion of beliefs or behaviors in a society, i.e., how beliefs or observable behaviors spread from few individuals to the whole population. Morris (2000) studies a dynamic framework where each agent interacts strategically with a finite subset of an infinite population. Each period, agents take a binary action (behavior) that is a best response to the actions of their neighbors in the previous period, i.e., each agent chooses an action that was played by a sufficiently large fraction of her neighbors. The paper characterizes when diffusion from a finite set of agents to the whole population is possible.[9] López-Pintado (2008) studies how behaviors spread in a

[8] An exception being Friedkin and Johnsen (1990), who study a variation of the DeGroot model where agents can adhere to their initial beliefs to some degree. This leads as well to persistent disagreement among the agents.

[9] Ellison (1993) studies learning in a dynamic large population coordination game. He focusses

social network.[10] Agents decide whether or not to adopt a new behavior as a function of the decisions taken by their neighbors. She finds the threshold for the spreading rate above which diffusion takes place and the new behavior becomes persistent in the population. This diffusion threshold depends on the connectivity distribution of the social network and the diffusion rule.

Both approaches, Bayesian and non-Bayesian models, have clear advantages and disadvantages. Bayesian models assume that agents update their beliefs optimally (from a statistical point of view), which makes them a nice benchmark of what we can expect in an ideal situation. However, it also makes issues like spread of misinformation difficult, almost impossible, to study, see also Acemoglu and Ozdaglar (2011). Choi et al. (2012) report an experimental investigation of learning in three-person networks and use the Bayesian framework of Gale and Kariv (2003) to interpret the generated data. They adapt the Quantal Response Equilibrium model by McKelvey and Palfrey (1995) to test the theory.[11] The paper finds that the theory can account for the behavior observed in the laboratory in a variety of networks and informational settings. In particular, they observe that individuals fail to account for repeated information, see also Corazzini et al. (2012).

These results suggest that the rationality assumption of Bayesian models is indeed quite demanding for individuals, especially when interacting on complex networks. On the other hand, the updating mechanism of non-Bayesian models may be too simple, for instance, as Choi et al. (2012) showed, individuals act at least boundedly rational. In another experimental work, Chandrasekhar et al. (2012) run a unique lab experiment in the field across 19 villages in rural Karnataka, India, to discriminate between models using Bayes' rule and the DeGroot mechanism.[12] They find evidence that the DeGroot model better explains the data than Bayesian learning models.[13] Moreover, they emphasize that many individuals come across

on the rate of diffusion and shows that when agents only interact with a small set of agents, it is likely that evolution instead of historical factors determine the strategies of the agents, i.e., convergence is fast enough such that we can expect to see the limit behavior being played.

[10] See also Jackson and Yariv (2007) and López-Pintado (2012).

[11] Roughly speaking, the Quantal Response Equilibrium model allows for idiosyncratic preference shocks such that the probability of a certain mistake is a decreasing function of the associated payoff difference and agents take into account that others make mistakes.

[12] Notice that in order to compare the two concepts, they study DeGroot action models, i.e., agents take an action after aggregating the actions of their neighbors using the DeGroot updating rule.

[13] At the network level (i.e., when the observational unit is the sequence of actions), both models do a decent job with Bayesian learning explaining 62% of the actions and the degree weighting DeGroot model explaining 76% of the actions taken by individuals. At the individual level (i.e., when the observational unit is the action of an individual given a history), both the degree weighting

information sets the Bayesian model attaches zero probability to, which could be interpreted as a lack of fit of the model.

1.3 The Approach

The objective of this work is to contribute to the literature on non-Bayesian social influence models. We study three aspects of social influence that have not received much attention in the literature.

- Anonymous social influence: agents are influenced only by the number of agents sharing a belief. We study belief updating rules based on ordered weighted averages, i.e., different to the weighted averaging rules widely studied in the literature, weights are not attached to agents, but to the ranks of the agents in the vector of beliefs.
- Manipulation: agents can manipulate the social network by increasing the attention other agents pay to them. We introduce the possibility of manipulating the trust weights of other agents into the model by DeGroot (1974).
- Conflicting interests: agents with conflicting interests communicate strategically with their neighbors. We introduce conflicting interests à la Crawford and Sobel (1982) to a model of non-Bayesian belief dynamics.

These elements are key in understanding how beliefs and opinions evolve in our societies. *Anonymous social influence* means the phenomenon that individuals are influenced by groups of people whose identity is unknown to them. This kind of influence has gained significant importance with the emergence of the internet, where individuals often follow positive evaluations of products and advices of anonymous people.

Second, *manipulation* is an aspect of social influence that is of importance when individuals need the support of others to enforce their interests in society. In politics, majorities are needed to pass laws and in companies, decisions might be taken by a hierarchical superior. It can therefore be advantageous for individuals to increase their influence on others and to manipulate the way others form their beliefs. This behavior is often referred to as lobbying and widely observed in society, especially in politics.

Furthermore, individuals in our societies typically have *conflicting interests* and widely diverging views on many issues, as can be seen in daily political discussions

and the uniform DeGroot model largely outperform Bayesian learning models.

or in all kinds of bargaining situations. In election campaigns, politicians have incentives to argue solutions or proposals that differ from their beliefs. In budget allocation problems, the recipients of capital, e.g., ministries, local governments or departments of companies or universities, have incentives to overstate their capital requirement, while the other side is concerned with efficiency. Another example are court trials, where the accused has clearly incentives to misreport the events in question.

1.4 Overview of the Results

We study societies that consist of n agents. Each agent i starts with an initial belief (or opinion) $x_i(0)$ about some common issue of interest, which could be an underlying state the agents would like to learn. The agents update their beliefs by repeatedly meeting and communicating with the other agents. At time $t \geq 0$, each agent i holds a belief $x_i(t)$. Our work is concerned with how these beliefs evolve in the long-run, i.e., when time tends to infinity.

In Chapter 2 (joint with Michel Grabisch and Agnieszka Rusinowska), we study influence processes modeled by *ordered weighted averaging operators*, commonly called *OWA operators* and introduced in Yager (1988). Agents start with "yes" or "no" inclinations (beliefs) on some common issue, i.e., $x_i(0) \in \{0, 1\}$ (where "yes" is coded as 1), and beliefs may change due to mutual influence among the agents. Each agent repeatedly (and independently) aggregates the beliefs of the other agents and possibly herself at discrete time instants using her OWA operator. This aggregation determines the probability that "yes" is her updated belief after one step of influence (and otherwise it is "no"). The other agents (only) observe the updated beliefs of all agents, i.e., the social network is the complete network.

We show that OWA operators are the only aggregation functions that are *anonymous* in the sense that the aggregation does only depend on how many agents hold a belief instead of which agents do so.[14] Accordingly, we call a model *anonymous* if the transitions between states of the process do only depend on how many agents share a belief. We show that the concept is consistent: if all agents use anonymous aggregation functions, then the model is anonymous. In particular, anonymous models allow to study situations where the influence process is based on *majori-*

[14] An aggregation function is defined by the following two conditions: (i) unanimity of beliefs persists (boundary conditions), and (ii) influence is positive (nondecreasingness), see Grabisch and Rusinowska (2013).

ties, which means that agents say "yes" if some kind of majority holds this belief.[15] These situations are not covered by the classical approach of weighted averaging aggregation.

We discuss the different types of terminal classes and characterize terminal states, i.e., singleton terminal classes. The condition is simple: the OWA operators must be such that all beliefs persist after mutual influence. In our main result, we find a necessary and sufficient condition for convergence to consensus. The condition says that there must be a certain number of agents such that if at least this number of agents says "yes," it is possible that after mutual influence more agents say "yes" and if less than that number of agents says "yes," it is possible that after mutual influence more agents say "no." In other words, we have a cascade that leads either to the "yes"- or "no"-consensus. Additionally, we also present an alternative characterization based on *influential coalitions*. We call a coalition influential on an agent if the latter follows (adopts) the belief of this coalition – given all other agents hold the opposite belief – with some probability. Furthermore, we generalize the model based on OWA operators and allow agents to use a (convex) combination of OWA operators and general aggregation functions (*OWA-decomposable* aggregation functions). We show that the sufficiency part of our main result still holds.

Besides identifying all possible terminal classes of the influence process, it is also important to know how quickly opinions will reach their limit. In Grabisch and Rusinowska (2013) no analysis of the speed of convergence has been provided. In this paper, we study the *speed of convergence* to terminal classes as well as the *probabilities of convergence* to certain classes in the general aggregation model. Computing the distribution of the speed of convergence and the probabilities of convergence can be demanding if the number of agents is large. However, we find that for anonymous models, we can reduce this demand substantially.

As an application of our model we study *fuzzy linguistic quantifiers*, which were introduced in Zadeh (1983) and are also called *soft quantifiers*. Typical examples of such quantifiers are expressions like "almost all," "most," "many" or "at least a few," see Yager and Kacprzyk (1997). For instance, an agent could say "yes" if "most of the agents say 'yes.'"[16] Yager (1988) has shown that for each quantifier we can find a unique corresponding OWA operator.[17] We find that if the agents use quantifiers that are similar in some sense, then they reach a consensus. Moreover, this result holds even if some agents deviate to quantifiers that are not similar in that sense.

[15] Examples are simple majorities as well as unanimity of beliefs, among others.

[16] Note that the formalization of such quantifiers is clearly to some extent ambiguous.

[17] With the only restriction that, due to our model, the quantifier needs to represent positive influence.

Loosely speaking, quantifiers are similar if their literal meanings are "close," e.g., "most" and "almost all."

In Chapter 3 (joint with Ana Mauleon and Vincent Vannetelbosch), we introduce the possibility of *manipulation* into the model by DeGroot (1974). Each agent starts with an initial belief $x_i(0) \in \mathbb{R}$ about some common issue and repeatedly communicates with her neighbors in the social network. At each period, first one agent is selected randomly and can exert effort to manipulate the social trust of an agent of her choice. If she decides to provide some costly effort to manipulate another agent, then the manipulated agent weights relatively more the belief of the agent who manipulated her when updating her belief. Second, all agents communicate with their neighbors and update their beliefs using the DeGroot updating rule, i.e., using her (possibly manipulated) weights, an agent's updated belief is the weighted average of her neighbors' beliefs (and possibly her own belief) from the previous period.

We first show that manipulation can modify the *trust structure*. If the society is split up into several disconnected clusters of agents and there are also some agents outside these clusters, then the latter agents might connect different clusters by manipulating the agents therein. Such an agent, previously outside any of these clusters, would not only get influential on the agents therein, but also serve as a bridge and connect them. As we show by means of an example, this can lead to a connected society, and thus, make the society reaching a consensus.

Second, we analyze the long-run beliefs and show that manipulation fosters *opinion leadership* in the sense that the manipulating agent always increases her influence on the long-run beliefs. For the other agents, this is ambiguous and depends on the social network. Surprisingly, the manipulated agent may thus even gain influence on the long-run beliefs. As a consequence, the expected change of influence on the long-run beliefs is ambiguous and depends on the agents' preferences and the social network. We also show that a definitive trust structure evolves in the society and, if the satisfaction of agents only depends on the current and future beliefs and not directly on the trust, manipulation will come to an end and they reach a consensus (under some weak regularity condition). At some point, beliefs become too similar to be manipulated. Furthermore, we discuss the speed of convergence and note that manipulation can accelerate or slow down convergence. In particular, in sufficiently homophilic societies, i.e., societies where agents tend to trust those agents who are similar to them, and where costs of manipulation are rather high compared to its benefits, manipulation accelerates convergence if it decreases homophily and otherwise it slows down convergence.

Finally, we investigate the tension between *information aggregation* and *spread of misinformation*. We find that if manipulation is rather costly and the agents underselling their information gain and those overselling their information lose overall influence (i.e., influence in terms of their initial information), then manipulation reduces misinformation and agents converge jointly to more accurate beliefs about some underlying true state. In particular, this means that an agent for whom manipulation is cheap can severely harm information aggregation.

In Chapter 4, we introduce conflicting interests into a model of non-Bayesian belief dynamics. We disentangle the terms *belief* and *opinion* (or *biased belief*): the belief of an individual about some issue of common interest will be what she holds to be true given her information about the issue. On the other hand, her opinion (or biased belief) will be what is ought to be the answer to the issue given her bias.[18]

At time $t \geq 0$, each agent holds a belief $x_i(t) \in [0,1]$ about some common issue. Furthermore, each agent has a bias $b_i \in \mathbb{R}$ that is common knowledge and that determines her opinion (biased belief) $x_i(t) + b_i$. Each agent starts with an initial belief $x_i(0) \in [0,1]$ and repeatedly meets (communicates with) agents in her social neighborhood according to a Poisson process in continuous time that is independent of the other agents.

When an agent is selected by her associated Poisson process, she receives information from one of her neighbors (called the sender of information) according to a stochastic process that forms her social network. We assume that the sender wants to spread his opinion, while the receiver wants to infer his belief in order to update her own belief. In equilibrium, this conflict of interest leads to noisy communication à la Crawford and Sobel (1982): the sender sends one of finite messages that contains information about his belief, which is then interpreted by the receiver. In optimal equilibrium, communication is as informative as possible given the conflict of interest, i.e., the sender uses as many messages as possible and discriminates as finely as possible between different beliefs. Finally, the receiver updates her belief by taking the average of the interpretation of the sent message and her pre-meeting belief.

Our framework induces a belief dynamics process as well as an opinion dynamics process. As a first observation, we note that we can concentrate our analysis on the belief dynamics process since both processes have the same convergence properties. We say that an agent's belief *fluctuates* on an interval if her belief will never leave the interval and if this does not hold for any subinterval. In other words, the belief

[18] In this sense, her opinion is a personal judgement about the issue for strategic reasons or taste considerations.

"travels" the whole interval, but not beyond.

In our main result, we show that for any initial beliefs, the belief dynamics process converges to a set of intervals that is *minimal mutually confirming*. Given each agent's belief lies in her corresponding interval, these intervals are the convex combinations of the interpretations the agents use when communicating. Furthermore, we show that the belief of an agent eventually fluctuates on her corresponding interval whenever the interval is proper, i.e., whenever it contains infinitely many elements (beliefs). As a consequence, the belief dynamics has a steady state if and only if there exists a minimal mutually confirming set such that all its intervals are degenerate, i.e., contain only a single point. Furthermore, we notice that as long as conflicts are small and some agents communicate with several different agents, outcomes with a steady state are non-generic.

The introduction of conflict of interest leads not only to persistent disagreement among the agents, but also to fluctuating beliefs and opinions, a phenomenon that is frequently observed in social sciences, see, e.g., Kramer (1971) who documents large swings in US voting behavior within short periods, and works in social and political psychology that study how political parties and other organizations influence political beliefs, e.g., Cohen (2003); Zaller (1992). At the same time, our result is surprising in view of the literature on dynamic social influence: in most models, a strongly connected network leads to mutual consensus among the agents in the long-run.

Chapter 2

Anonymous Social Influence*

2.1 Introduction

In the present work we study an important and widespread phenomenon which affects many aspects of human life – the phenomenon of *influence*. Being undoubtedly present, e.g., in economic, social and political behaviors, influence frequently appears as a dynamic process. In particular, social influence plays a crucial role in the formation of opinions, beliefs and the diffusion of information and thus, it is not surprising that numerous scientific works investigate different dynamic models of influence.[19]

Grabisch and Rusinowska (2010, 2011) investigate a one-step deterministic model of influence, where agents have "yes" or "no" inclinations (beliefs) on some common issue and their opinions may change due to mutual influence among the agents. Grabisch and Rusinowska (2013) extend it to a dynamic stochastic model based on aggregation functions, which determine how the agents update their opinions depending on the current opinions in the society. Each agent repeatedly (and independently) aggregates the opinions of the other agents and possibly herself at discrete time instants. This aggregation determines the probability that "yes" is her updated opinion after one step of influence (and otherwise it is "no"). The other agents only observe this updated opinion. Since any aggregation function is allowed when updating the opinions, the framework covers numerous existing models of opinion formation. The only restrictions come from the definition of an aggregation function: unanimity of opinions persists (boundary conditions) and influence

*This chapter is a modified version of the article published as: Förster, Manuel, Michel Grabisch, and Agnieszka Rusinowska (2013). Anonymous social influence. Games and Economic Behavior 82, 621–35.

[19]For an overview of the vast literature on influence we refer, e.g., to Jackson (2008).

is positive (nondecreasingness). Grabisch and Rusinowska (2013) provide a general analysis of convergence in the aggregation model and find all terminal classes, which are sets of states the process will not leave once they have been reached. Such a class could only consist of one single state, e.g., the states where we have unanimity of opinions ("yes"- and "no"-consensus) or a state where the society is polarized, i.e., some group of agents finally says "yes" and the rest says "no."

Due to the generality of the model of influence based on arbitrary aggregation functions introduced in Grabisch and Rusinowska (2013), it would be difficult to obtain a deeper insight into some particular phenomena of influence by using this model. This is why the analysis of particular classes of aggregation functions and the exhaustive study of their properties are necessary for explaining many social and economic interactions. One of them concerns *anonymous social influence*, which is particularly present in real-life situations. Internet, accompanying us in everyday life, intensifies enormously anonymous influence: when we need to decide which washing machine to buy, which hotel to reserve for our eagerly awaited holiday, we will certainly follow all anonymous customers and tourists that have expressed their positive opinion on the object of our interest. In the present paper we examine a particular way of aggregating the opinions and investigate influence processes modeled by *ordered weighted averaging operators (ordered weighted averages)*, commonly called *OWA operators* and introduced in Yager (1988), because they appear to be a very appropriate tool for modeling and analyzing anonymous social influence. Roughly speaking, OWA operators are similar to the ordinary weighted averages (weighted arithmetic means), with the essential difference that weights are not attached to agents, but to the *ranks* of the agents in the input vector. As a consequence, OWA operators are in general nonlinear, and include as particular cases the median, the minimum and the maximum, as well as the (unweighted) arithmetic mean.

We show that OWA operators are the only aggregation functions that are *anonymous* in the sense that the aggregation does only depend on how many agents hold an opinion instead of which agents do so. Accordingly, we call a model *anonymous* if the transitions between states of the process do only depend on how many agents share an opinion. We show that the concept is consistent: if all agents use anonymous aggregation functions, then the model is anonymous. However, as we show by example, a model can be anonymous although agents do not use anonymous functions. In particular, anonymous models allow to study situations where the influence process is based on *majorities*, which means that agents say "yes" if

some kind of majority holds this opinion.[20] These situations are not covered by the classical (commonly used) approach of weighted averaging aggregation.

In the main part, we study the convergence of models based on OWA operators.[21] We discuss the different types of terminal classes and characterize terminal states, i.e., singleton terminal classes. The condition is simple: the OWA operators must be such that all opinions persist after mutual influence. In our main result, we find a necessary and sufficient condition for convergence to consensus. The condition says that there must be a certain number of agents such that if at least this number of agents says "yes," it is possible that after mutual influence more agents say "yes" and if less than that number of agents says "yes," it is possible that after mutual influence more agents say "no." In other words, we have a cascade that leads either to the "yes"- or "no"-consensus. Additionally, we also present an alternative characterization based on *influential coalitions*. We call a coalition influential on an agent if the latter follows (adopts) the opinion of this coalition – given all other agents hold the opposite opinion – with some probability.[22] Furthermore, we generalize the model based on OWA operators and allow agents to use a (convex) combination of OWA operators and general aggregation functions (*OWA-decomposable* aggregation functions). In particular, this allows us to combine OWA operators and ordinary weighted averaging operators. As a special case of this, we study models of *mass psychology* (also called herding behavior) in an example. We find that this model is equivalent to a convex combination of the majority influence model and a completely self-centered agent. We also study an example on *important agents* where agents trust some agents directly that are important for them and otherwise follow a majority model. Furthermore, we show that the sufficiency part of our main result still holds.[23]

Besides identifying all possible terminal classes of the influence process, it is also important to know how quickly opinions will reach their limit. In Grabisch and Rusinowska (2013) no analysis of the speed of convergence has been provided. In this paper, we study the *speed of convergence* to terminal classes as well as the *probabilities of convergence* to certain classes in the general aggregation model. Computing the distribution of the speed of convergence and the probabilities of

[20] Examples are simple majorities as well as unanimity of opinions, among others.

[21] Note that (implicitly) the social network is the complete network since agents observe all opinions.

[22] Note that although Grabisch and Rusinowska (2013) have already studied conditions for convergence to consensus and other terminal classes in the general model, our results are inherently different due to our restriction to anonymous aggregation functions.

[23] When applying the condition to the OWA operators in the convex combinations.

convergence can be demanding if the number of agents is large. However, we find that for anonymous models, we can reduce this demand substantially.[24]

As an application of our model we study *fuzzy linguistic quantifiers*, which were introduced in Zadeh (1983) and are also called *soft quantifiers*. Typical examples of such quantifiers are expressions like "almost all," "most," "many" or "at least a few," see Yager and Kacprzyk (1997). For instance, an agent could say "yes" if "most of the agents say 'yes.'"[25] Yager (1988) has shown that for each quantifier we can find a unique corresponding OWA operator.[26] We find that if the agents use quantifiers that are similar in some sense, then they reach a consensus. Moreover, this result holds even if some agents deviate to quantifiers that are not similar in that sense. Loosely speaking, quantifiers are similar if their literal meanings are "close," e.g., "most" and "almost all." We also give examples to provide some intuition.

We terminate this section with a brief overview of the related literature. One of the main differences between our work and the existing models on opinion formation lies in the way agents are assumed to aggregate the opinions. Except, e.g., Grabisch and Rusinowska (2013) many related works assume a convex combination as the way of aggregating opinions. Additionally, while we consider "yes"/"no" opinions, in some models of influence, like in the seminal model of opinion and consensus formation due to DeGroot (1974), the opinion of an agent is a number in $[0, 1]$. Moreover, in DeGroot (1974) every agent aggregates the opinions (beliefs) of other agents through an ordinary weighted average. The interaction among agents is captured by the social influence matrix. Several scholars have analyzed the DeGroot framework and proposed different variations of it, in which the updating of opinions can vary in time and along circumstances. However, most of the influence models usually assume a convex combination as the way of aggregating opinions. Golub and Jackson (2010) examine convergence of the social influence matrix and reaching a consensus, and the speed of convergence of beliefs, among other things. DeMarzo et al. (2003) consider a model where an agent may place more or less weight on her own belief over time. Another framework related to the DeGroot model is presented in Asavathiratham (2000) and López-Pintado and Watts (2008). Büchel et al. (2011) introduce a generalization of the DeGroot model by studying the transmission of cultural traits from one generation to the next one. Büchel et al. (2012) analyze an influence model in which agents may misrepresent their opinion in a conforming or counter-conforming

[24] We have to compute powers and inverses of matrices whose dimensions grow exponentially in the number of agents. In anonymous models this reduces to linear growth.

[25] Note that the formalization of such quantifiers is clearly to some extent ambiguous.

[26] With the only restriction that, due to our model, the quantifier needs to represent positive influence.

way. Calvó-Armengol and Jackson (2009) study an overlapping-generations model in which agents, that represent some dynasties forming a community, take yes-no actions.

López-Pintado (2008, 2012) studies the spreading of behavior in society and investigate the role of social influence therein. While these papers focus on the social network and use simple diffusion rules that are the same for all agents, we do not impose a network structure and allow for heterogeneous agents. Van den Brink and his co-authors study power measures in weighted directed networks, see, e.g., van den Brink and Gilles (2000); Borm et al. (2002). A different approach to influence, i.e., a method based on simulations, is presented in Mäs (2010). Morris (2000) analyzes the phenomenon of contagion which occurs if an action can spread from a finite set of individuals to the whole population.

Another stream of related literature concerns models of Bayesian and observational learning, where agents observe choices over time and update their beliefs accordingly, see, e.g., Banerjee (1992), Ellison (1993), Bala and Goyal (1998, 2001), Gale and Kariv (2003) and Banerjee and Fudenberg (2004). This literature differs from the influence models mentioned above as in the latter the choices depend on the influence of others. Mueller-Frank (2010) considers continuous aggregation functions with a special property called "constricting" and studies convergence applied to non-Bayesian learning in social networks. Galeotti and Goyal (2009) model networks in terms of degree distributions and study influence strategies in the presence of local interaction.

The literature on OWA operators comprises, in particular, applications to multi-criteria decision-making. Jiang and Eastman (2000), for instance, apply OWA operators to geographical multi-criteria evaluation, and Malczewski and Rinner (2005) present a fuzzy linguistic quantifier extension of OWA in geographical multi-criteria evaluation. Using ordered weighted averages in (social) networks is quite new, although some scholars have already initiated such an application, see Cornelis et al. (2010), who apply OWA operators to trust networks. To the best of our knowledge, ordered weighted averages have not been used to model social influence yet.

The remainder of the paper is organized as follows. In Section 2.2 we present the model and basic definitions. Section 2.3 introduces the notion of anonymity. Section 2.4 concerns the convergence analysis in the aggregation model with OWA operators. In Section 2.5 the speed of convergence and the probabilities of different outcomes are studied. In Section 2.6 we apply our results on ordered weighted averages to fuzzy linguistic quantifiers. Section 2.7 contains some concluding remarks. The longer proofs of some of our results are presented in Appendix 2.A.

2.2 Model and Notation

Let $N = \{1, 2, \ldots, n\}$, $n \geq 2$, be the set of agents that have to make a "yes" or "no" decision on some issue. Each agent $i \in N$ has an *initial opinion* $x_i(0) \in \{0, 1\}$ (called *inclination*) on the issue, where "yes" is coded as 1. Let us denote by 1_S the characteristic vector of $S \subseteq N$, i.e., $(1_S)_j = 1$ if $j \in S$ and $(1_S)_j = 0$ otherwise. We can represent the vector of initial opinions $x(0) = (x_1(0), x_2(0), \ldots, x_n(0))'$ by such a characteristic vector.[27] We say that S is the initial *state* or *coalition* if $x(0) = 1_S$ is the vector of initial opinions. In other words, the initial state consists of the agents that have the inclination "yes." We sometimes denote a state $S = \{i, j, k\}$ simply by ijk and its cardinality or *size* by s. During the influence process, agents may change their opinion due to mutual influence among the agents. They update their opinion simultaneously at discrete time instants.

Definition 1 (Aggregation function)**.** An *n-place aggregation function* is any mapping $A : \{0, 1\}^n \to [0, 1]$ satisfying

(i) $A(0, \ldots, 0) = 0, A(1, \ldots, 1) = 1$ (boundary conditions) and

(ii) if $x \leq x'$ then $A(x) \leq A(x')$ (nondecreasingness).

To each agent i we assign an aggregation function A_i that determines the way she reacts to the opinions of the other agents and herself.[28] Note that by using these functions we model positive influence only. Our aggregation model $\mathbf{A} = (A_1, A_2, \ldots, A_n)'$ is stochastic, the A_i, $i = 1, 2, \ldots, n$, are mutually independent and the output $A_i(1_S) \in [0, 1]$ of agent i's aggregation function is her probability to say "yes" after one step of influence when the current opinions are $x(t) = 1_S$ (at time t), i.e., $x_i(t+1) = 1$ with probability $A_i(1_S)$, and otherwise $x_i(t+1) = 0$. The other agents do not know these probabilities, but they observe the realization $x(t + 1)$ of the updated opinions. Note that we do not explicitly model the realization of the updated opinions, which is for agent i a (biased) coin toss with probability $A_i(1_S)$ of "yes" and probability $1 - A_i(1_S)$ of "no." Therefore, we can represent the realized and observed opinions (after one step of influence) again by a state $S' \subseteq N$ such that $i \in S'$ with probability $A_i(1_S)$.

The aggregation functions our paper is mainly concerned with are *ordered weighted averaging operators* or simply *ordered weighted averages*. This class of aggregation functions was first introduced by Yager (1988).

[27] We denote the transpose of a vector x by x'.

[28] Note that we use a modified version of aggregation functions by restricting the opinions to be from $\{0, 1\}$ instead of $[0, 1]$. We discuss this issue later on in Example 1.

Definition 2 (Ordered weighted average)**.** We say that an n-place aggregation function A is an *ordered weighted average* $A = \mathsf{OWA}_w$ with weight vector w, i.e., $0 \leq w_i \leq 1$ for $i = 1, 2, \ldots, n$ and $\sum_{i=1}^{n} w_i = 1$, if $A(x) = \sum_{i=1}^{n} w_i x_{(i)}$ for all $x \in \{0,1\}^n$, where $x_{(1)} \geq x_{(2)} \geq \ldots \geq x_{(n)}$ are the ordered components of x.

The definition of an aggregation function ensures that the two consensus states – the "yes"-consensus $\{N\}$ where all agents say "yes" and the "no"-consensus $\{\emptyset\}$ where all agents say "no" – are fixed points of the aggregation model $\mathbf{A} = (A_1, A_2, \ldots, A_n)'$. We call them *trivial terminal classes*. Before we go on, let us give an example of an ordered weighted average already presented in Grabisch and Rusinowska (2013), the *majority influence model*. Furthermore, we also use this example to argue why we do restrict opinions to be either "yes" or "no."

Example 1 (Majority)**.** A straightforward way of making a decision is based on majority voting. If the majority of the agents says "yes," then all agents agree to say "yes" after mutual influence and otherwise, they agree to say "no." We can model simple majorities as well as situations where more than half of the agents are needed to reach the "yes"-consensus. Let $m \in \{\lfloor \frac{n}{2} \rfloor + 1, \lfloor \frac{n}{2} \rfloor + 2, \ldots, n\}$. Then, the *majority* aggregation model is given by

$$\mathsf{Maj}_i^{[m]}(x) = x_{(m)} \text{ for all } i \in N.$$

All agents use an ordered weighted average where $w_m = 1$. Obviously, the convergence to consensus is immediate.

The restriction of opinions to $\{0,1\}$ is crucial in order to study situations that depend on how many agents share an opinion. It allows us to specify the probability to say "yes" after mutual influence for any possible number of agents having the current opinion "yes" (with the restrictions given by the definition of an aggregation function).[29]

Furthermore, let us look at some examples apart from the majority model.

Example 2 (Some ordered weighted averages)**.** Consider some agent $i \in N = \{1, 2, 3, 4, 5\}$ who uses an ordered weighted average, $A_i = \mathsf{OWA}_w$.

(i) If $w = \left(0, 0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)'$, then this agent will say "no" for sure if there is not even a simple majority in favor of the issue. Otherwise, she will say "yes" with a positive probability, which increases by $\frac{1}{3}$ with each additional agent being in favor of the issue.

[29] Allowing for opinions in $[0,1]$ would lead to a deterministic model where each agent's updated opinion is a weighted average of the ordered opinions in society. In particular, such a model would require a different analytical approach.

(ii) If $w = \left(\frac{1}{3}, \frac{2}{3}, 0, 0, 0\right)'$, then this agent will already say "yes" if only one agent does so and she will be in favor for sure whenever at least two agents say "yes." This could represent a situation where it is perfectly fine for the agent if only a few of the others are in favor of the issue.

(iii) If $w = \left(\frac{1}{2}, 0, 0, 0, \frac{1}{2}\right)'$, then this agent will say "yes" with probability $\frac{1}{2}$ if neither all agents say "no" nor all agents say "yes." This could be interpreted as an agent who is indifferent and so decides randomly.

We have already seen that there always exist the two trivial terminal classes. In general, a terminal class is defined as follows:

Definition 3 (Terminal class)**.** A *terminal class* is a collection of states $\mathcal{C} \subseteq 2^N$ that forms a strongly connected and closed component, i.e., for all $S, T \in \mathcal{C}$, there exists a path[30] from S to T and there is no path from S to T if $S \in \mathcal{C}, T \notin \mathcal{C}$.

We can decompose the state space into disjoint terminal classes – also called absorbing classes – $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_l \subseteq 2^N$, for some $l \geq 2$, and a set of transient states $\mathcal{T} = 2^N \backslash (\bigcup_{k=1}^{l} \mathcal{C}_k)$. Let us now define the notion of an *influential agent* (Grabisch and Rusinowska, 2013).

Definition 4 (Influential agent)**.** (i) An agent $j \in N$ is *"yes"-influential* on $i \in N$ if $A_i(1_{\{j\}}) > 0$.

(ii) An agent $j \in N$ is *"no"-influential* on $i \in N$ if $A_i(1_{N\backslash\{j\}}) < 1$.

The idea is that j is "yes"-(or "no"-)influential on i if j's opinion to say "yes" (or "no") matters for i in the sense that there is a positive probability that i follows the opinion that is solely held by j. Analogously to influential agents, we can define *influential coalitions* (Grabisch and Rusinowska, 2013).

Definition 5 (Influential coalition)**.** (i) A nonempty coalition $S \subseteq N$ is *"yes"-influential* on $i \in N$ if $A_i(1_S) > 0$.

(ii) A nonempty coalition $S \subseteq N$ is *"no"-influential* on $i \in N$ if $A_i(1_{N\backslash S}) < 1$.

Making the assumption that the probabilities of saying "yes" are independent among agents[31] and only depend on the current state, we can represent our aggregation model by a time-homogeneous Markov chain with transition matrix $\mathbf{B} =$

[30] We say that there is a path from S to T if there is $K \in \mathbb{N}$ and states $S = S_1, S_2, \ldots, S_{K-1}, S_K = T$ such that $A_i(S_k) > 0$ for all $i \in S_{k+1}$ and $A_i(S_k) < 1$ otherwise, for all $k = 1, 2, \ldots, K-1$.

[31] This assumption is not limitative, and correlated opinions may be considered as well. In the latter case, only the next equation giving $b_{S,S'}$ will differ.

$(b_{S,S'})_{S,S'\subseteq N}$, where

$$b_{S,S'} = \Pi_{i\in S'} A_i(1_S)\Pi_{i\notin S'}(1 - A_i(1_S)).$$

Hence, the states of this Markov chain are the states or coalitions of the agents that currently say "yes" in the influence process. Thus, $b_{S,S'}$ denotes the probability, given the current state $S \subseteq N$, that the process is in state $S' \subseteq N$ after one step of influence. Note that for each state S, the transition probabilities to states S' are represented by a certain row of $\mathbf{B}$. Notice also that this Markov chain is neither irreducible nor recurrent since it has at least two terminal classes.[32] The m-th power of a matrix, e.g., $\mathbf{B} = (b_{S,S'})_{S,S'\subseteq N}$, is denoted by $\mathbf{B}^m = (b_{S,S'}(m))_{S,S'\subseteq N}$.

2.3 Anonymity

We establish the notions of *anonymous* aggregation functions and models. In what follows, we show that the notions of anonymity are consistent and that anonymous functions are characterized by OWA operators.

Definition 6 (Anonymity)**.** (i) We say that an n-place aggregation function A is *anonymous* if for all $x \in \{0,1\}^n$ and any permutation $\sigma : N \to N$, $A(x_1, x_2, \ldots, x_n) = A(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)})$.

(ii) Suppose $\mathbf{B}$ is obtained from an aggregation model with aggregation functions $A_1, A_2, \ldots, A_n$. We say that the model is *anonymous* if for all $s, u \in \{0, 1, \ldots, n\}$,

$$\sum_{\substack{U\subseteq N:\\ |U|=u}} b_{S,U} = \sum_{\substack{U\subseteq N:\\ |U|=u}} b_{S',U} \text{ for all } S, S' \subseteq N \text{ of size } s.$$

For an agent using an anonymous aggregation function, only the size of the current coalition matters. Similarly, in models that satisfy anonymity, only the size of the current coalition matters for the further influence process. In other words, it matters how many agents share an opinion, but not which agents do so. Let us now confirm that our notions of anonymity are consistent in the sense that models where agents use anonymous functions are anonymous. Moreover, we characterize anonymous aggregation functions by ordered weighted averages.

Proposition 1. *(i) An aggregation model with anonymous aggregation functions $A_1, A_2, \ldots, A_n$ is anonymous.*

[32] In the language of Markov chains, terminal classes are also called communication classes.

(ii) An aggregation function A is anonymous if and only if it is an ordered weighted average.

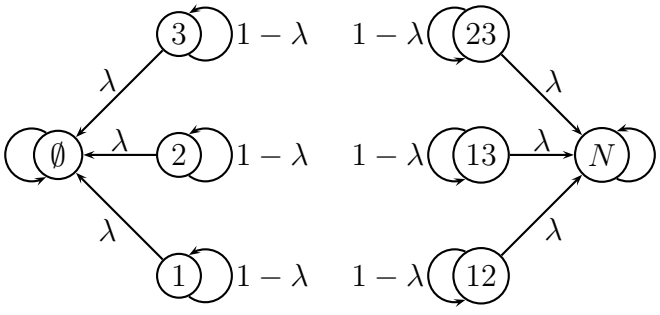
Proof. We omit the proof of (i) as well as the necessity part of (ii). For the sufficiency part, suppose that A is an anonymous aggregation function, i.e., for all $x \in \{0,1\}^n$ and any permutation $\sigma : N \to N$, $A(x_1, x_2, \ldots, x_n) = A(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)})$. This is equivalent to $A(1_S) = A(1_{S'})$ for all $S, S' \subseteq N$ such that $|S| = |S'|$. Hence, there exists $w \in \mathbb{R}^n$ such that $A(1_S) = \sum_{i \in N} w_i (1_S)_{(i)}$ for all $S \subseteq N$. It follows by the definition of aggregation functions that $w_i \geq 0$ for all $i \in N$ (nondecreasingness) and $\sum_{i=1}^n w_i = 1$ (boundary condition), which finishes the proof. □

Note that the converse of the first part does not hold, a model can be anonymous although not all agents use anonymous aggregation functions as we now show by example. We study the phenomenon of *mass psychology*, also called herding behavior, considered in Grabisch and Rusinowska (2013).

Example 3 (Mass psychology)**.** Mass psychology or herding behavior means that if at least a certain number $m \in \{\lfloor \frac{n}{2} \rfloor + 1, \lfloor \frac{n}{2} \rfloor + 2, \ldots, n\}$ of agents share the same opinion, then these agents attract others, who had a different opinion before. We assume that an agent changes her opinion in this case with probability $\lambda \in (0,1)$. In particular, we consider $n = 3$ agents and a threshold of $m = 2$. This means whenever only two agents are of the same opinion, the third one might change her opinion. This corresponds to the following *mass psychology* aggregation model:

$$\mathsf{Mass}_i^{[2]}(x) = \lambda x_{(2)} + (1-\lambda) x_i \text{ for all } i \in N.$$

Agents are "yes"- and "no"-influential on themselves and coalitions of size two or more are "yes"- and "no"-influential on all agents. The model gives the following digraph of the Markov chain:



The aggregation functions are not anonymous since agents consider their own opinion with weight $1 - \lambda > 0$. However, the model turns out to be anonymous, there is no differentiation between different coalitions of the same size, as can be seen from the digraph.

An immediate consequence of Proposition 1 is that models where agents use OWA operators are anonymous.

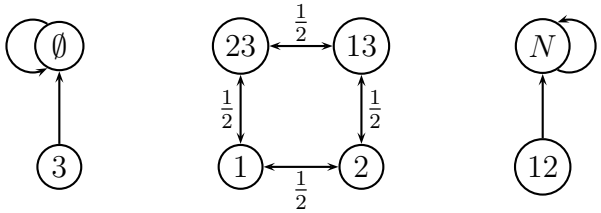
Corollary 1. *Aggregation models with aggregation functions* $A_i = \mathsf{OWA}_{w^i}, i \in N$, *are anonymous.*

2.4 Convergence Analysis

In this section, we study the convergence of aggregation models where the influence process is determined by OWA operators, i.e., by anonymous aggregation functions. In Grabisch and Rusinowska (2013, Theorem 2), the authors show that there are three different types of terminal classes in the general model. To terminal classes of the first type, singletons $\{S\}$, $S \subseteq N$, we usually refer to as *terminal states*. They represent the two consensus states, $\{N\}$ and $\{\emptyset\}$, as well as situations where the society is eventually polarized: agents within the class say "yes," while the others say "no." Classes of the second type are called *cyclic terminal classes*, their states form a cycle of nonempty sets $\{S_1, S_2, \ldots, S_k\}$ of any length $2 \leq k \leq \binom{n}{\lfloor n/2 \rfloor}$ (and therefore they are periodic of period k) with the condition that all sets are pairwise incomparable (by inclusion).[33] In other words, given the process has reached a state within such a class, the transition to the next state is deterministic. And the period of the class determines after how many steps a state is reached again.

Terminal classes of the third type are called *regular terminal classes*. They are collections $\mathcal{R}$ of nonempty sets with the property that $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \cdots \cup \mathcal{R}_p$, where each sub-collection $\mathcal{R}_j$ is an interval $\{S \in 2^N \mid S_j \subseteq S \subseteq S_j \cup K_j\}$, with $S_j \neq \emptyset, S_j \cup K_j \neq N$, and at least one K_j is nonempty.

Example 4 (Regular terminal class)**.** Consider an aggregation model with three agents and aggregation functions $A_1(x) = x_2, A_2(x) = x_1$ and $A_3(x) = (x_1 + x_2)/2$. Then, $\{\{1\}, \{1,3\}\} \cup \{\{2\}, \{2,3\}\}$ is a regular terminal class. The model gives the following digraph of the Markov chain:



[33] Sets $S_1, S_2, \ldots, S_k \subseteq N$ are called pairwise incomparable (by inclusion) if for any distinct $S_i, S_j, i, j \in \{1, 2, \ldots, k\}$, both $S_i \not\subseteq S_j$ and $S_i \not\supseteq S_j$.

If such a class only consists of a single interval $\mathcal{R}_1 = \{S \in 2^N \mid S_1 \subseteq S \subseteq S_1 \cup K_1\}$, where $S_1, K_1 \neq \emptyset$ and $S_1 \cup K_1 \neq N$, then we can interpret this terminal class as a situation where agents in S_1 finally decided to say "yes" and agents outside $S_1 \cup K_1$ finally decided to say "no," while the agents in K_1 change their opinion non-deterministically forever. With more than one interval, the interpretation is more complex and depends on the transitions between the intervals. Reaching an interval $\mathcal{R}_j$ means that the process attains one of its states, i.e., the agents in S_j say "yes" for sure and with some probability, also some agents in K_j do so.

Our aim is to investigate conditions for these outcomes under anonymous influence. We also relax our setup and study the case where agents use ordered weighted averages only to some extent. Our results turn out to be – due to the restriction to anonymous aggregation functions – inherently different from those in the general model, see Grabisch and Rusinowska (2013). We first consider influential coalitions and discuss (non-trivial) terminal classes. In the following, we derive a characterization of convergence to consensus and finally provide a generalization of our setting.

Due to anonymity, it is not surprising that the influence of a coalition indeed solely depends on the number of individuals involved.

Proposition 2. *Consider an aggregation model with aggregation functions* $A_i = \mathsf{OWA}_{w^i}$, $i \in N$.

(i) A coalition of size s, *where* $0 < s \leq n$, *is "yes"-influential on* $i \in N$ *if and only if* $\min\{k \in N \mid w^i_k > 0\} \leq s$.

(ii) A coalition of size s, *where* $0 < s \leq n$, *is "no"-influential on* $i \in N$ *if and only if* $\max\{k \in N \mid w^i_k > 0\} \geq n + 1 - s$.

Proof. Let $S \subseteq N$ have size $0 < s \leq n$ and be "yes"-influential on $i \in N$, i.e.,

$$A_i(1_S) = \sum_{k=1}^{s} w^i_k > 0 \Leftrightarrow \min\{k \in N \mid w^i_k > 0\} \leq s.$$

The second part is analogous. □

The result on influential agents follows immediately.

Corollary 2. *Consider an aggregation model with aggregation functions* $A_i = \mathsf{OWA}_{w^i}$, $i \in N$. *Then, all agents* $j \in N$ *are "yes"-("no"-)influential on* $i \in N$ *if and only if* $w^i_1 > 0$ *(*$w^i_n > 0$*).*

Note that this means that either all agents are "yes"-(or "no"-)influential on some agent $i \in N$ or none. Next, we study non-trivial terminal classes. We characterize terminal states, i.e., states where the society is polarized (except for the trivial terminal states), and show that – due to anonymity – there cannot be a cycle.

Proposition 3. *Consider an aggregation model with aggregation functions* $A_i = \mathsf{OWA}_{w^i}$, $i \in N$.

(i) A state $S \subseteq N$ *of size* s *is a terminal state if and only if* $\sum_{k=1}^{s} w_k^i = 1$ *for all* $i \in S$ *and* $\sum_{k=1}^{s} w_k^i = 0$ *otherwise.*

(ii) There does not exist any cycle.

Proof. The first part is obvious. For the second part, assume that there is a cycle $\{S_1, S_2, \ldots, S_k\}$ of length $2 \leq k \leq \binom{n}{\lfloor n/2 \rfloor}$. This implies that there exists $l \in \{1, 2, \ldots, k\}$ such that $s_l \leq s_{l+1}$, where $S_{k+1} \equiv S_1$. Thus,

$$\sum_{j=1}^{s_l} w_j^i = 1 \text{ for all } i \in S_{l+1}$$

and hence $S_{l+1} \subseteq S_{l+2}$, which is a contradiction to pairwise incomparability by inclusion, see Grabisch and Rusinowska (2013, Theorem 2). □

For regular terminal classes, note that an agent $i \in N$ such that $w_1^i = 1$ blocks a "no"-consensus and an agent $j \in N$ such that $w_n^j = 1$ blocks a "yes"-consensus – given that the process has not yet arrived at a consensus. Therefore, since there cannot be any cycle, these two conditions, while ensuring that there is no other terminal state, give us a regular terminal class with anonymous aggregation functions.

Example 5 (Anonymous regular terminal class)**.** Consider an aggregation model with aggregation functions $A_i = \mathsf{OWA}_{w^i}, i \in N = \{1, 2, 3\}$. Let agent 1 block a "no"-consensus and agent 3 block a "yes"-consensus, i.e., $w_1^1 = w_3^3 = 1$. Furthermore, choose $w_1^2 = w_3^2 = \frac{1}{2}$. Then, $\{\{1\}, \{1, 2\}\}$ is a regular terminal class. We have $\mathbf{A}(1_{\{1\}}) = \mathbf{A}(1_{\{1,2\}}) = (1 \ \frac{1}{2} \ 0)'$.

It is left to find conditions that avoid both non-trivial terminal states and regular terminal classes and hence ensure that the society ends up in a consensus. The following result characterizes the non-existence of non-trivial terminal classes. The idea is that – due to anonymity – for reaching a consensus, there must be some threshold such that whenever the size of the coalition is at least equal to this threshold, there is some probability that after mutual influence, more agents will say "yes." And whenever the size is below this threshold, there is some probability that after mutual influence, more agents will say "no."

Theorem 1. *Consider an aggregation model with aggregation functions* $A_i = \mathsf{OWA}_{w^i}$, $i \in N$. *Then, there are no other terminal classes than the trivial terminal classes if and only if there exists* $\bar{k} \in \{1, 2, \ldots, n\}$ *such that both:*

(i) For all $k = \bar{k}, \bar{k}+1, \ldots, n-1$, *there are distinct agents* $i_1, i_2, \ldots, i_{k+1} \in N$ *such that*

$$\sum_{j=1}^{k} w_j^{i_l} > 0 \text{ for all } l = 1, 2, \ldots, k+1.$$

(ii) For all $k = 1, 2, \ldots, \bar{k}-1$, *there are distinct agents* $i_1, i_2, \ldots, i_{n-k+1} \in N$ *such that*

$$\sum_{j=1}^{k} w_j^{i_l} < 1 \text{ for all } l = 1, 2, \ldots, n-k+1.$$

The proof is in Appendix 2.A. Note that Theorem 1 implies a straightforward – but very strict – sufficient condition:

Remark 1. Consider an aggregation model with aggregation functions $A_i = \mathsf{OWA}_{w^i}$, $i \in N$. Then, there are no other terminal classes than the trivial terminal classes if $w_1^i > 0$ for all $i \in N$ ($\bar{k} = 1$), or $w_n^i > 0$ for all $i \in N$ ($\bar{k} = n$).

We get a more intuitive formulation of Theorem 1 by using influential coalitions.

Corollary 3. *Consider an aggregation model with aggregation functions* $A_i = \mathsf{OWA}_{w^i}$, $i \in N$. *Then, there are no other terminal classes than the trivial terminal classes if and only if there exists* $\bar{k} \in \{1, 2, \ldots, n\}$ *such that both:*

(i) For all $k = \bar{k}, \bar{k}+1, \ldots, n-1$, *there are* $k+1$ *distinct agents such that coalitions of size* k *are "yes"-influential on each of them.*

(ii) For all $k = 1, 2, \ldots, \bar{k}-1$, *there are* $n-k+1$ *distinct agents such that coalitions of size* $n-k$ *are "no"-influential on each of them.*

In more general situations, the agents' behavior might only partially be determined by ordered weighted averages. We consider agents who use aggregation functions that are *decomposable* in the sense that they are (convex) combinations of ordered weighted averages and general aggregation functions.

Definition 7 (OWA-decomposable aggregation function)**.** We say that an n-place aggregation function A is OWA_w*-decomposable*, if there exists $\lambda \in (0, 1]$ and an n-place aggregation function A' such that $A = \lambda \mathsf{OWA}_w + (1-\lambda)A'$.

Such aggregation functions do exist since convex combinations of aggregation functions are again aggregation functions. Note that these functions are, in general, not anonymous any more, though. However, the mass psychology influence model presented in Section 2.3 – to which we will come back later on – is an example of an anonymous model that uses these decomposable aggregation functions. To provide some intuition for why these functions are useful, let us consider the class where ordered weighted averages are combined with weighted averages.[34]

Example 6 (OWA-/WA-decomposable aggregation functions)**.** Consider a convex combination of an ordered weighted average and a weighted average,

$$A = \lambda \mathsf{OWA}_w + (1-\lambda)\mathsf{WA}_{w'},$$

where $\lambda \in (0,1)$ and w, w' are any weight vectors. This allows us to somehow combine our model with the classical model by DeGroot.[35] We can interpret this as follows: to some extent λ, an agent updates her opinion anonymously to account, e.g., for majorities within her social group. But she might as well value her own opinion somehow – like in the mass psychology model – or some agents might be really important for her such that she wants to put also some weight directly on them, as we will show in Example 8.

As it turns out, the sufficiency part of Theorem 1 also holds if agents use such decomposable aggregation functions. If the ordered weighted average components of the decomposable functions fulfill the two conditions of Theorem 1, then the agents reach a consensus.[36]

Corollary 4. *Consider an aggregation model with* OWA_{w^i}*-decomposable aggregation functions* A_i*,* $i \in N$*. Then, there are no other terminal classes than the trivial terminal classes if there exists* $\bar{k} \in \{1, 2, \ldots, n\}$ *such that both:*

(i) For all $k = \bar{k}, \bar{k}+1, \ldots, n-1$*, there are distinct agents* $i_1, i_2, \ldots, i_{k+1} \in N$ *such that*

$$\sum_{j=1}^{k} w_j^{i_l} > 0 \text{ for all } l = 1, 2, \ldots, k+1.$$

[34] We say that an n-place aggregation function A is a *weighted average* $A = \mathsf{WA}_w$ with weight vector w, i.e., $0 \leq w_i \leq 1$ for $i = 1, 2, \ldots, n$ and $\sum_{i=1}^{n} w_i = 1$, if $A(x) = \sum_{i=1}^{n} w_i x_i$ for all $x \in \{0,1\}^n$.

[35] With the restriction that, differently to the DeGroot model, opinions are in {0,1}.

[36] It is clear that, in general, the necessity part does not hold since convergence to consensus may as well be (partly) ensured by the other component.

(ii) For all $k = 1, 2, \ldots, \bar{k} - 1$*, there are distinct agents* $i_1, i_2, \ldots, i_{n-k+1} \in N$ *such that*

$$\sum_{j=1}^{k} w_j^{i_l} < 1 \text{ for all } l = 1, 2, \ldots, n - k + 1.$$

Let us finally apply the concept of decomposable aggregation functions to more specific examples. As it turns out, the example on mass psychology combines the majority influence model and a completely self-centered agent.

Example 7 (Mass psychology, cont'd)**.** We have seen in Example 3 that for parameters $n = 3$, $m = 2$ and $\lambda \in (0, 1)$, we get the following mass psychology aggregation model:

$$\mathsf{Mass}_i^{[2]}(x) = \lambda x_{(2)} + (1 - \lambda)x_i \text{ for all } i \in N.$$

This aggregation function is OWA_w-decomposable with $w_2 = 1$ and by Corollary 4, taking $\bar{k} = 2$, we see that the group eventually reaches a consensus. This example is a particular case of Example 6 and furthermore, it is equivalent to a convex combination of the majority influence model and a completely self-centered agent:

$$\mathsf{Mass}_i^{[2]}(x) = \lambda \mathsf{Maj}_i^{[2]}(x) + (1 - \lambda)x_i \text{ for all } i \in N.$$

Hence, λ could be interpreted as a measure for how "democratically" – or, to put it the other way, "egoistically" – an agent behaves.

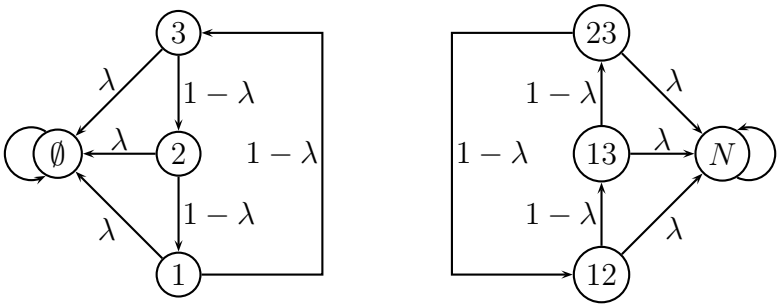
Finally, we study an example where agents use the majority influence model, but also put some weight directly on agents that are important for them. We study a case that turns out to be as well anonymous and furthermore, it is in some sense equivalent to the example on mass psychology.

Example 8 (Important agents)**.** Although agents might follow somehow a majority influence model, there might still be some important agents, e.g., very good friends or agents with an excellent reputation, whom they would like to trust directly as well. In particular, we consider $n = 3$ agents and that each agent follows to some extent $\lambda \in (0, 1)$ the simple majority model. Moreover, for each agent, the agent with the next higher index has a relative importance of $1 - \lambda$ for her.[37] This corresponds to the following *important agents* aggregation model:

$$\mathsf{Imp}_i^{[2;i+1]}(x) = \lambda \mathsf{Maj}_i^{[2]}(x) + (1 - \lambda)x_{i+1} \text{ for all } i \in N.$$

Agent $i + 1$ is "yes"- and "no"-influential on agent i for all $i \in N$ and coalitions of size two or more are "yes"- and "no"-influential on all agents. The model gives the following digraph of the Markov chain:

[37] We consider $4 \equiv 1$.



From the digraph, we can see that the model is anonymous although the aggregation functions are not.[38] Furthermore, when abstracting from the identity of the agents, i.e., considering only the size of a state, this digraph is identical to the one of the mass psychology example. Therefore, we can say that the two models are *anonymously equivalent*: starting in a state of size one or two, both models stay within the set of states of the same size with probability $1 - \lambda$ and converge to the "no"- or "yes"-consensus, respectively, with probability λ.

2.5 Speed of Convergence and Absorption

We first study the *speed of convergence* of the influence process to terminal classes.[39] Secondly, we investigate the probabilities of convergence to each of the consensus states and possibly other terminal classes. We call these probabilities *absorption probabilities*. Since this analysis has not been done in Grabisch and Rusinowska (2013), we provide it for the general aggregation model and also for anonymous models, which cover particularly the case where all agents use OWA operators. This section relies on results from Markov chain theory. We find that anonymity leads to a substantial gain in computational tractability.

Suppose that $\mathbf{B}$ is obtained from an aggregation model with aggregation functions $A_1, A_2, \ldots, A_n$ and that there is at least one transient state, i.e., $\mathcal{T} \neq \emptyset$. We assume that the process starts from a transient state $S \in \mathcal{T}$, i.e., $x(0) = 1_S$. Note that since the set of transient states is finite, we have convergence to the terminal classes almost surely. We say that the influence process $\mathbf{B}$ converges to the terminal classes at time t if $x(t-1) = 1_S$ such that $S \in \mathcal{T}$ and $x(t) = 1_{S'}$ such that $S' \notin \mathcal{T}$. Thus, the speed of convergence is the time it takes for the process to leave the set

[38] Note that this is a consequence of our choice of important agents. For most choices, the model would not be anonymous, e.g., if two agents would be important for each other and one of them would as well be important for the third one.

[39] In the language of Markov chains, the speed of convergence is also called time before absorption.

of transient states.[40]

We define, given $x(0) = 1_S$, $S \in \mathcal{T}$, the *speed of convergence* as

$$\tau_S := \inf\{t \in \mathbb{N} \mid x(t) = 1_{S'} \text{ such that } S' \notin \mathcal{T}\},$$

which is an almost surely finite stopping time on the probability space induced by the aggregation model $\mathbf{A} = (A_1, A_2, \ldots, A_n)'$. Our aim is to determine the distribution of the speed of convergence. It turns out that the latter is solely determined by the transition probabilities within the set of transient states. We denote the restriction of $\mathbf{B}$ to these states by $Q = (q_{S,S'})_{S,S' \in \mathcal{T}} := \mathbf{B}|_{\mathcal{T}} = (b_{S,S'})_{S,S' \in \mathcal{T}}$.

Proposition 4. *Suppose* $\mathbf{B}$ *is obtained from an aggregation model with aggregation functions* $A_1, A_2, \ldots, A_n$*. If* $x(0) = 1_S$*,* $S \in \mathcal{T}$*, then*

$$\mathbb{P}(\tau_S > t) = \sum_{S' \in \mathcal{T}} q_{S,S'}(t) \text{ and } \mathbb{E}[\tau_S] = \sum_{m=0}^{\infty} \sum_{S' \in \mathcal{T}} q_{S,S'}(m) < +\infty.$$

Proof. The first part follows from Brémaud (1999, p. 154, Theorem 5.2). For the expected value of τ_S, first note that it only takes nonnegative integer values. The first equality of the following computation follows from this fact, whereas the third equality and the inequality follow since $\mathcal{T}$ is finite and Q is strictly sub-stochastic, i.e., $\sum_{m=0}^{\infty} Q^m < +\infty$.[41]

$$\mathbb{E}[\tau_S] = \sum_{m=0}^{\infty} \mathbb{P}(\tau_S > m) = \sum_{m=0}^{\infty} \sum_{S' \in \mathcal{T}} q_{S,S'}(m) = \sum_{S' \in \mathcal{T}} \sum_{m=0}^{\infty} q_{S,S'}(m) < +\infty.$$

□

Next, suppose that $\mathbf{B}$ is obtained from an anonymous aggregation model $\mathbf{A} = (A_1, A_2, \ldots, A_n)'$. Then, $\mathbf{B}$ can be reduced from a $2^n \times 2^n$ transition matrix to an $(n+1) \times (n+1)$ matrix $\mathbf{B}^a = (b^a_{s,s'})_{s,s' \in \{0,1,\ldots,n\}}$, where

$$b^a_{s,s'} = \sum_{\substack{S' \subseteq N: \\ |S'| = s'}} b_{S,S'}, \text{ for any } S \subseteq N \text{ of size } s,$$

are the transition probabilities from coalitions of size s to coalitions of size s'. However, note that the gain in tractability (the dimensions of the transition matrix grow only linearly instead of exponentially in the number of agents) comes at the cost of

[40] Note that we do not consider the speed of convergence to certain terminal classes since its expected value will be infinite if there is a positive probability that this may not happen. Instead, we consider later on the absorption probabilities of certain terminal classes.

[41] cf. Brémaud (1999, p. 155, Theorem 6.1). It is understood that the right member is a matrix whose entries are all $+\infty$.

losing track of the transition probabilities to certain states. For a given terminal class $\mathcal{C}$ and the set of transient states $\mathcal{T}$, we define the corresponding *anonymous terminal class* and the *anonymous set of transient states* by $\mathcal{C}^a = \{s \in \{0, 1, \ldots, n\} \mid \exists\, S \in \mathcal{C} \text{ such that } |S| = s\}$ and $\mathcal{T}^a = \{s \in \{0, 1, \ldots, n\} \mid S \in \mathcal{T} \text{ if } |S| = s\}$, respectively.

Note that anonymous terminal classes are extended by states of the same size as states within the original class. This implies that the speed of convergence will be distorted in case it is possible that the process arrives at a state which is part of an anonymous terminal class, but not of the corresponding original one. We call such a model *distorted*. In this case, we need to use the original model to compute the speed of convergence. Models that only have singleton terminal classes are not distorted, though.

The speed of convergence, given $x(0) = 1_S$, $S \in \mathcal{T}$ such that $|S| = s$, is denoted by τ_s and the restriction of $\mathbf{B}^a$ to transient states by $Q^a = \mathbf{B}^a|_{\mathcal{T}}$. We find that anonymity leads to a substantial gain in computational tractability since it suffices to use Q^a instead of Q to compute the distribution of the speed of convergence.

Corollary 5. *Suppose $\mathbf{B}^a$ is obtained from an anonymous aggregation model with aggregation functions $A_1, A_2, \ldots, A_n$ that is not distorted. If $x(0) = 1_S$, $S \in \mathcal{T}$ such that $|S| = s$, then*

$$\mathbb{P}(\tau_s > t) = \sum_{s' \in \mathcal{T}^a} q^a_{s,s'}(t) \text{ and } \mathbb{E}[\tau_s] = \sum_{m=0}^{\infty} \sum_{s' \in \mathcal{T}^a} q^a_{s,s'}(m) < +\infty.$$

The next step is to look at the absorption probabilities of certain terminal classes. Define by

$$D_k = (d_{S,S'})_{S \in \mathcal{T}, S' \in \mathcal{C}_k} := (b_{S,S'})_{S \in \mathcal{T}, S' \in \mathcal{C}_k}$$

the matrix of transition probabilities from transient states to states within the terminal class $\mathcal{C}_k$. For our analysis it does not matter at which state the influence process enters a terminal class and hence we can reduce the matrices D_k to a vector by considering a terminal class $\mathcal{C}_k$ simply as a terminal state $\widetilde{\mathcal{C}_k}$. The transition probabilities from transient states to a terminal class $\mathcal{C}_k$ are then given by the vector

$$\widetilde{D_k} := \left(\sum_{S' \in \mathcal{C}_k} d_{S,S'} \right)_{S \in \mathcal{T}}.$$

Let us denote the matrix of transition probabilities from transient states to the terminal classes by $\widetilde{D} := (\widetilde{D_1} : \widetilde{D_2} : \cdots : \widetilde{D_l})$ and define $F := (\mathbf{I} - Q)^{-1}$.[42] Furthermore, we define, given $x(0) = 1_S$, $S \in \mathcal{T}$, the time of absorption by the terminal

[42] Note that for absorbing Markov chains the matrix F always exists since $Q^m \to 0$ for $m \to \infty$.

class $\mathcal{C}_k$ as

$$\tau_S^k := \inf\{t \in \mathbb{N} \mid x(t) = 1_{S'} \text{ such that } S' \in \mathcal{C}_k\}.$$

Note that this stopping time is not almost surely finite in general. We have $\tau_S^k = +\infty$ if the influence process was absorbed by a terminal class other than $\mathcal{C}_k$. The following result immediately follows from Brémaud (1999, p. 157, Theorem 6.2).

Proposition 5. *Suppose* $\mathbf{B}$ *is obtained from an aggregation model with aggregation functions* $A_1, A_2, \ldots, A_n$*. If* $x(0) = 1_S$*,* $S \in \mathcal{T}$*, then we get for the absorption probabilities:*

$$\mathbb{P}(\tau_S^k < \infty) = g_{S,k}, \text{ for } k = 1, 2, \ldots, l,$$

where $(g_{S,k})_{S\in\mathcal{T},k\in\{1,2,\ldots,l\}} := F\widetilde{D}$*.*

The corresponding result for anonymous models is straightforward. The reason is that if, in a distorted model, the influence process has reached a state that is part of an anonymous terminal class, but not of the corresponding original one, then it will converge to that original class immediately due to anonymity. This also provides a justification for not considering such states as possible initial states.

Let $D_k^a = (d_{s,s'})_{s\in\mathcal{T}^a,s'\in\mathcal{C}_k^a} := (b_{s,s'})_{s\in\mathcal{T}^a,s'\in\mathcal{C}_k^a}$ denote the matrix of transition probabilities from transient states to states within the anonymous terminal class $\mathcal{C}_k^a$. Furthermore, let $\widetilde{D_k^a} := \left(\sum_{s'\in\mathcal{C}_k^a} d_{s,s'}\right)_{s\in\mathcal{T}^a}$ denote the reduced matrices, $\widetilde{D^a} := (\widetilde{D_1^a} : \widetilde{D_2^a} : \cdots : \widetilde{D_l^a})$ their collection, and define $F^a := (\mathbf{I} - Q^a)^{-1}$. The time of absorption by the anonymous terminal class $\mathcal{C}_k^a$, given $x(0) = 1_S$, $S \in \mathcal{T}$ such that $|S| = s$, is denoted by τ_s^k.

Corollary 6. *Suppose* $\mathbf{B}^a$ *is obtained from an anonymous aggregation model with aggregation functions* $A_1, A_2, \ldots, A_n$*. If* $x(0) = 1_S$*,* $S \in \mathcal{T}$ *such that* $|S| = s$*, then we get for the absorption probabilities:*

$$\mathbb{P}(\tau_s^k < \infty) = g_{s,k}^a, \text{ for } k = 1, 2, \ldots, l,$$

where $\left(g_{s,k}^a\right)_{s\in\mathcal{T}^a,k\in\{1,2,\ldots,l\}} := F^a\widetilde{D^a}$*.*

This finishes our analysis of the speed of convergence and absorption probabilities.[43] To illustrate the results, we come back the example on mass psychology.

[43] We could also discuss the convergence after the process has entered a terminal class. This is obvious at least for singleton and cyclic terminal classes, though. For the latter, there is clearly no convergence to a stationary distribution. Furthermore, it holds that regular classes are convergent if and only if their corresponding transition matrix is aperiodic.

Example 9 (Mass psychology, cont'd)**.** We have seen in Example 3 that for parameters $n = 3$, $m = 2$ and $\lambda \in (0, 1)$, we get the following mass psychology aggregation model:

$$\mathsf{Mass}_i^{[2]}(x) = \lambda x_{(2)} + (1 - \lambda)x_i \text{ for all } i \in N.$$

Due to anonymity, we get for any initial opinions $x(0) = 1_S$, $1 \leq |S| = s \leq 2$:

$$\mathbb{P}(\tau_s > t) = (1 - \lambda)^t \text{ and } \mathbb{E}[\tau_s] = \frac{1}{\lambda}.$$

So, the speed of convergence hinges on λ, the probability that an agent follows the herd. If it is small, the process can take a long time. If initially two agents said "yes," the process terminates (with probability one) in the "yes"-consensus and otherwise, it terminates in the "no"-consensus.

Recall that Example 8 on important agents is anonymously equivalent to this example. Therefore, the speed of convergence is the same in both examples.

2.6 Applications to Fuzzy Linguistic Quantifiers

Instead of being sharp edged, e.g., as in the majority model, the threshold of an agent initially saying "no" for changing her opinion might be rather "soft." For instance, she could change her opinion if "*most* of the agents say 'yes.'" This is called a *soft majority* and phrases like "most" or "many" are so-called *fuzzy linguistic quantifiers*. Furthermore, *soft minorities* are also possible, e.g., "*at least a few* of the agents say 'yes.'" Our aim is to apply our findings on ordered weighted averages to fuzzy linguistic quantifiers. Mathematically, we define the latter by a function which maps the agents' proportion that says "yes" to the degree to which the quantifier is satisfied, see Zadeh (1983).

Definition 8 (Fuzzy linguistic quantifier)**.** A *fuzzy linguistic quantifier* $\mathcal{Q}$ is defined by a nondecreasing function

$$\mu_{\mathcal{Q}} : [0, 1] \to [0, 1] \text{ such that } \mu_{\mathcal{Q}}(0) = 0 \text{ and } \mu_{\mathcal{Q}}(1) = 1.$$

Furthermore, we say that the quantifier is *regular* if the function is strictly increasing on some interval $(\underline{c}, \bar{c}) \subseteq [0, 1]$ and otherwise constant.

Notice that fuzzy linguistic quantifiers are more general than ordered weighted averages since they assign a probability to say "yes" after mutual influence to any proportion of agents currently saying "yes" and therefore do not depend on the number of agents n in society.

Fuzzy linguistic quantifiers like "most" are ambiguous in the sense that it is not clear how to define them exactly mathematically. For example, one could well discuss which proportion of the agents should say "yes" for the quantifier "most" to be fully satisfied. Nevertheless, let us give some typical examples, see Yager and Kacprzyk (1997).

Example 10 (Typical quantifiers)**.** We define

(i) $\mathcal{Q}_{\text{aa}} =$ "almost all" by

$$\mu_{\mathcal{Q}_{\text{aa}}}(x) := \begin{cases} 1, & \text{if } x \geq \frac{9}{10} \\ \frac{5}{2}x - \frac{5}{4}, & \text{if } \frac{1}{2} < x < \frac{9}{10} \\ 0, & \text{otherwise} \end{cases},$$

(ii) $\mathcal{Q}_{\text{mo}} =$ "most" by

$$\mu_{\mathcal{Q}_{\text{mo}}}(x) := \begin{cases} 1, & \text{if } x \geq \frac{4}{5} \\ \frac{5}{2}x - 1, & \text{if } \frac{2}{5} < x < \frac{4}{5} \\ 0, & \text{otherwise} \end{cases},$$

(iii) $\mathcal{Q}_{\text{ma}} =$ "many" by

$$\mu_{\mathcal{Q}_{\text{ma}}}(x) := \begin{cases} 1, & \text{if } x \geq \frac{3}{5} \\ \frac{5}{2}x - \frac{1}{2}, & \text{if } \frac{1}{5} < x < \frac{3}{5} \\ 0, & \text{otherwise} \end{cases},$$

(iv) $\mathcal{Q}_{\text{af}} =$ "at least a few" by

$$\mu_{\mathcal{Q}_{\text{af}}}(x) := \begin{cases} 1, & \text{if } x \geq \frac{3}{10} \\ \frac{10}{3}x, & \text{otherwise} \end{cases}.$$

Note that these quantifiers are regular. For every quantifier, given the number of agents n in society, there exists a corresponding ordered weighted average in the sense that the latter represents the quantifier.[44] We can find its weights as follows.

Lemma 1 (Yager, 1988)**.** *Let $\mathcal{Q}$ be a fuzzy linguistic quantifier defined by $\mu_{\mathcal{Q}}$. Then, the weights of its corresponding ordered weighted average $\mathsf{OWA}_{\mathcal{Q}}$ are given by*

$$w_k = \mu_{\mathcal{Q}}\left(\frac{k}{n}\right) - \mu_{\mathcal{Q}}\left(\frac{k-1}{n}\right), \text{ for } k = 1, 2, \ldots, n.$$

[44] Note that this is due to our definition. The conditions in Definition 8 ensure that there exists such an ordered weighted average. In general, one can define quantifiers also by other functions, cf. Zadeh (1983).

In other words, the weights w_k of the corresponding ordered weighted average are equal to the increase of $\mu_{\mathcal{Q}}$ between $(k-1)/n$ and k/n, i.e., since $\mu_{\mathcal{Q}}$ is nondecreasing, all weights are nonnegative and by the boundary conditions, it is ensured that they sum up to one. We are now in the position to apply our results to regular quantifiers. We find that if all agents use such a quantifier, then under some similarity condition, the group will finally reach a consensus. This condition says that there must be a common point where all the fuzzy quantifiers are strictly increasing. This implies that there is a common non-zero weight of the corresponding OWA operators, which turns out to be sufficient to satisfy the condition of Theorem 1. Moreover, we show that the result still holds if some agents deviate to a quantifier that is not similar in that sense. In the following, we denote the quantifier of an agent i by $\mathcal{Q}^i$.

Proposition 6. *Consider an aggregation model with aggregation functions* $A_i = \mathsf{OWA}_{\mathcal{Q}^i}, i \in N$.

(i) If $\mathcal{Q}^i$ *is regular for all* $i \in N$ *and* $\cap_{i\in N}(\underline{c}_i, \bar{c}_i) \neq \emptyset$, *then there are no other terminal classes than the trivial terminal classes.*

(ii) Suppose $\min_{i\in N} \underline{c}_i > 0$, *then the result in (i) still holds if less than* $\lceil \bar{c}_d n \rceil$ *agents deviate to a regular quantifier* $\mathcal{Q}_d$ *such that* $\bar{c}_d < \min_{i\in N} \underline{c}_i$.

(iii) Suppose $\max_{i\in N} \bar{c}_i < 1$, *then the result in (i) still holds if less than* $\lceil (1 - \underline{c}_d)n \rceil$ *agents deviate to a regular quantifier* $\mathcal{Q}_d$ *such that* $\max_{i\in N} \bar{c}_i < \underline{c}_d$.

The proof is in Appendix 2.A. Note that this result can be generalized such that the deviating agents might also use different quantifiers. We can also characterize terminal states in a model where agents use regular quantifiers. We find that S is a terminal state if and only if the quantifiers of the agents in S are already fully satisfied at s/n, while the quantifiers of the other agents are not satisfied at all at this point.

Proposition 7. *Consider an aggregation model with aggregation functions* $A_i = \mathsf{OWA}_{\mathcal{Q}^i}, i \in N$. *If* $\mathcal{Q}^i$ *is regular for all* $i \in N$, *then a state* $S \subseteq N$ *of size* s *is a terminal state if and only if*

$$\max_{i\in S} \bar{c}_i \leq \frac{s}{n} \leq \min_{i\in N\setminus S} \underline{c}_i.$$

Proof. Suppose $S \subseteq N$ of size s is a terminal state. By Proposition 3, we know that

this is equivalent to

$$\begin{aligned}
&\sum_{k=1}^{s} w_k^i = 1 \text{ for all } i \in S \text{ and } \sum_{k=1}^{s} w_k^i = 0 \text{ otherwise} \\
\Leftrightarrow &\mu_{\mathcal{Q}^i}(s/n) = 1 \text{ for all } i \in S \text{ and } \mu_{\mathcal{Q}^i}(s/n) = 0 \text{ otherwise} \\
\Leftrightarrow &\max_{i \in S} \bar{c}_i \leq \frac{s}{n} \leq \min_{i \in N \setminus S} \underline{c}_i.
\end{aligned}$$

□

To provide some intuition, let us come back to Example 10 and look at the implications our findings have on the quantifiers defined therein.

Example 11 (Typical quantifiers, cont'd)**.** Consider an aggregation model with aggregation functions $A_i = \mathsf{OWA}_{\mathcal{Q}^i}, i \in N$.

(i) If $\mathcal{Q}^i \in \{\mathcal{Q}_{\text{aa}}, \mathcal{Q}_{\text{mo}}, \mathcal{Q}_{\text{ma}}\}$ for all $i \in N$, then there are no other terminal classes than the trivial terminal classes. The result still holds if less than $\lceil \frac{3}{10} n \rceil$ agents deviate to $\mathcal{Q}_{\text{af}}$.

(ii) If $\mathcal{Q}^i \in \{\mathcal{Q}_{\text{ma}}, \mathcal{Q}_{\text{af}}\}$ for all $i \in N$, then there are no other terminal classes than the trivial terminal classes. The result still holds if less than $\lceil \frac{1}{2} n \rceil$ agents deviate, each of them either to $\mathcal{Q}_{\text{aa}}$ or $\mathcal{Q}_{\text{mo}}$.

(iii) A state $S \subseteq N$ of size s is a terminal state if $\mathcal{Q}^i = \mathcal{Q}_{\text{af}}$ for all $i \in S$, $\mathcal{Q}^i = \mathcal{Q}_{\text{aa}}$ ($\mathcal{Q}^i \in \{\mathcal{Q}_{\text{aa}}, \mathcal{Q}_{\text{mo}}\}$) otherwise and $\frac{3}{10} \leq \frac{s}{n} \leq \frac{1}{2}$ ($\leq \frac{2}{5}$).

2.7 Conclusion

We study a stochastic model of influence where agents aggregate opinions using *OWA operators*, which are the only *anonymous* aggregation functions. As one would expect, an aggregation model is *anonymous* if all agents use these functions. However, our example on mass psychology shows that a model can be anonymous although agents do not use anonymous functions.

In the main part of the paper, we characterize influential coalitions, show that cyclic terminal classes cannot exist due to anonymity and characterize terminal states. Our main result provides a necessary and sufficient condition for *convergence to consensus*. It turns out that we can express this condition in terms of influential coalitions. Due to our restriction to anonymous functions, these results are inherently different to those obtained in the general case by Grabisch and Rusinowska (2013). We also extend our model to *decomposable* aggregation functions. In particular, this allows to combine OWA operators with the classical approach of

ordinary weighted averages. This class of decomposed functions comprises our example on mass psychology: it is equivalent to a convex combination of the majority influence model and a completely self-centered agent. We also study an example on important agents and show that in some cases, this model is anonymous as well and, additionally, anonymously equivalent to the example on mass psychology. Moreover, it turns out that our previous condition on convergence to consensus is still sufficient in this generalized setting.

We analyze the *speed of convergence* to terminal classes as well as *probabilities of absorption* by different terminal classes in the general model studied by Grabisch and Rusinowska (2013) and in our case of anonymous models. For the latter, and in particular models based on OWA operators, we can reduce the computational demand substantially compared to the general case. Furthermore, we apply our results to *fuzzy linguistic quantifiers* and show that if agents use in some sense similar quantifiers and not too many agents deviate from these quantifiers, the society will eventually reach a consensus.

These results rely on the fact that for each quantifier, we can find a unique corresponding ordered weighted average (Lemma 1), which allows to apply our results on OWA operators. Note that these corresponding ordered weighted averages clearly depend on the number of agents in the society. Therefore, we can see a quantifier as well as a more general definition of an OWA operator (usually called an *extended OWA operator*; see Grabisch et al., 2009), which does not anymore require a fixed number of agents. In other words, assigning to each agent such an extended OWA operator allows to vary the number of agents n in the society.

2.A Appendix

Proof of Theorem 1

First, suppose that there exists $\bar{k} \in \{1, 2, \ldots, n\}$ such that (i) and (ii) hold. Let us take any coalition $S \subsetneq N$ of size $s \geq \bar{k}$ and show that it is possible to reach the "yes"-consensus, which implies that S is not part of a terminal class. By choice of S, it is sufficient to show that there is a positive probability that after mutual influence, the size of the coalition has strictly increased. That is, it is sufficient to show that there exists a coalition $S' \subseteq N$ of size $s' > s$, such that $A_i(1_S) > 0$ for all agents $i \in S'$. Set $k := s$, then by condition (i), there are distinct agents $i_1, i_2, \ldots, i_{k+1} \in N$

such that

$$A_{i_l}(1_S) = \sum_{j=1}^{k} w_j^{i_l} > 0 \text{ for all } l = 1, 2, \ldots, k+1,$$

i.e., setting $S' := \{i_1, i_2, \ldots, i_{k+1}\}$ finishes this part. Analogously, we can show by condition (ii) that for any nonempty $S \subseteq N$ of size $s < \bar{k}$ it is possible to reach the "no"-consensus. Hence, there are only the trivial terminal classes.

Now, suppose to the contrary that for all $\bar{k} \in \{1, 2, \ldots, n\}$ either (i) or (ii) does not hold. Note that in order to establish that there exists a non-trivial terminal class, it is sufficient to show that there are $k_*, k^* \in \{1, 2, \ldots, n-1\}, k_* \leq k^*$, such that for all $S \subseteq N$ of size $s = k_*$,

$$A_i(1_S) < 1 \text{ for at most } n - k_* \text{ distinct agents } i \in N \qquad (C_*[k_*])$$

and for all $S \subseteq N$ of size $s = k^*$,

$$A_i(1_S) > 0 \text{ for at most } k^* \text{ distinct agents } i \in N. \qquad (C^*[k^*])$$

Indeed, condition $(C_*[k_*])$ says that it is not possible to reach a coalition with less than k_* agents starting from a coalition with at least k_* agents. Similarly, condition $(C^*[k^*])$ says that it is not possible to reach a coalition with more than k^* agents starting from a coalition with at most k^* agents.[45] Therefore, it is not possible to reach the trivial terminal states from any coalition S of size $k_* \leq s \leq k^*$, which proves the existence of a non-trivial terminal class.

Let now $\bar{k} = 1$. Then, clearly condition (ii) is satisfied and thus condition (i) cannot be satisfied by assumption. Hence, there exists $k^* \in \{1, 2, \ldots, n-1\}$ such that there are at most k^* distinct agents $i_1, i_2, \ldots, i_{k^*}$ such that

$$\sum_{j=1}^{k^*} w_j^{i_l} > 0 \text{ for } l = 1, 2, \ldots, k^*.$$

This implies that condition (i) is not satisfied for $\bar{k} = 1, 2, \ldots, k^*$. If $k^* \geq 2$ and additionally condition (ii) was not satisfied for some $\bar{k} \in \{2, 3, \ldots, k^*\}$, we were done since then there would exist $k_* \in \{1, 2, \ldots, k^* - 1\}$ such that there are at most $n - k_*$ distinct agents $i_1, i_2, \ldots, i_{n-k_*}$ such that

$$\sum_{j=1}^{k_*} w_j^{i_l} < 1 \text{ for } l = 1, 2, \ldots, n - k_*,$$

[45] Note that monotonicity of the aggregation function implies that $(C_*[k_*])$ also holds if we replace S by a coalition $S' \subseteq N$ of size $s' > k_*$. Analogously for $(C^*[k^*])$.

i.e., $(C_*[k_*])$ and $(C^*[k^*])$ were satisfied for $k_* \leq k^*$. Therefore, suppose without loss of generality that condition (ii) is satisfied for all $\bar{k} = 1, 2, \ldots, k^*$. (1)

For $\bar{k} = n$, clearly condition (i) is satisfied and thus condition (ii) cannot be satisfied. Hence, using (1), there exists $k_* \in \{k^*, k^* + 1, \ldots, n-1\}$ such that there are at most $n - k_*$ distinct agents $i_1, i_2, \ldots, i_{n-k_*}$ such that

$$\sum_{j=1}^{k_*} w_j^{i_l} < 1 \text{ for } l = 1, 2, \ldots, n - k_*,$$

i.e., $(C_*[k_*])$ and $(C^*[k^*])$ are satisfied. We now proceed by case distinction:

(1) If $k_* = k^*$, then we are done.

(2) If $k_* > k^*$, then let $\bar{k} = k_*$. By assumption, either (i) or (ii) does not hold.

(2.1) If (i) does not hold, then there exists $k^{**} \in \{k_*, k_* + 1, \ldots, n-1\}$ such that there are at most k^{**} distinct agents $i_1, i_2, \ldots, i_{k^{**}}$ such that

$$\sum_{j=1}^{k^{**}} w_j^{i_l} > 0 \text{ for } l = 1, 2, \ldots, k^{**},$$

i.e., $(C_*[k_*])$ and $(C^*[k^{**}])$ are satisfied for $k_* \leq k^{**}$ and hence we are done.

(2.2) If (ii) does not hold, then, using (1), there exists $k_{**} \in \{k^*, k^*+1, \ldots, k_*-1\}$ such that there are at most $n - k_{**}$ distinct agents $i_1, i_2, \ldots, i_{n-k_{**}}$ such that

$$\sum_{j=1}^{k_{**}} w_j^{i_l} < 1 \text{ for } l = 1, 2, \ldots, n - k_{**},$$

i.e., $(C_*[k_{**}])$ is satisfied. If $k_{**} = k^*$, then we are done, otherwise we can repeat this procedure using k_{**} instead of k_*.

Since $k_{**} \lneq k_*$, we find $k_{**} = k^*$ after a finite number of repetitions, which finishes the proof.

Proof of Proposition 6

(i) By assumption, there exists $c \in \cap_{i \in N}(\underline{c}_i, \bar{c}_i)$. Let us define $\bar{k} := \min\{k \in \mathbb{N} \mid \frac{k}{n} > c\}$, then clearly $\frac{\bar{k}-1}{n} \leq c$. We show that conditions (i) and (ii) of Theorem 1 are satisfied for $\bar{k}$. Since for all $i \in N$, $\mu_{\mathcal{Q}^i}$ is nondecreasing and, in particular, strictly increasing on the open ball $B_\epsilon(c)$ around c for some $\epsilon > 0$, we get by Lemma 1 that

$$w_{\bar{k}}^i = \mu_{\mathcal{Q}}\left(\frac{\bar{k}}{n}\right) - \mu_{\mathcal{Q}}\left(\frac{\bar{k}-1}{n}\right) \geq \mu_{\mathcal{Q}}\left(\frac{\bar{k}}{n}\right) - \mu_{\mathcal{Q}}(c) > 0 \text{ for all } i \in N.$$

This implies that for all $k = \bar{k}, \bar{k}+1, \ldots, n-1$,

$$\sum_{j=1}^{k} w_j^i \geq w_{\bar{k}}^i > 0 \text{ for all } i \in N$$

and for all $k = 1, 2, \ldots, \bar{k}-1$,

$$\sum_{j=1}^{k} w_j^i \leq \sum_{j \neq \bar{k}} w_j^i = 1 - w_{\bar{k}}^i < 1 \text{ for all } i \in N,$$

i.e., (i) and (ii) of Theorem 1 are satisfied for $\bar{k}$, which finishes the first part.

(ii) Suppose $\min_{i \in N} \underline{c}_i > 0$ and denote by $D \subseteq N$ the set of agents that deviate to the quantifier $\mathcal{Q}_d$. Similar to the first part, there exists $c \in \cap_{i \in N \setminus D}(\underline{c}_i, \bar{c}_i)$ and we can define $\bar{k} := \min\{k \in \mathbb{N} \mid \frac{k}{n} > c\}$. This implies that for all $k = \bar{k}, \bar{k}+1, \ldots, n-1$,

$$\sum_{j=1}^{k} w_j^i > 0 \text{ for all } i \in N \setminus D \tag{2}$$

and for all $k = 1, 2, \ldots, \bar{k}-1$,

$$\sum_{j=1}^{k} w_j^i < 1 \text{ for all } i \in N \setminus D. \tag{3}$$

Furthermore, we have by assumption $\mu_{\mathcal{Q}_d}(\bar{k}/n) = 1$, which implies $w_j^i = 0$ for all $j = \bar{k}+1, \bar{k}+2, \ldots, n$ and $i \in D$. Thus, for all $k = \bar{k}, \bar{k}+1, \ldots, n-1$

$$\sum_{j=1}^{k} w_j^i = \sum_{j=1}^{\bar{k}} w_j^i = 1 > 0 \text{ for all } i \in D,$$

i.e., in combination with (2), condition (i) of Theorem 1 is satisfied for $\bar{k}$. It is left to check condition (ii). Define for $i \in D$,

$$\tilde{k} := \max\{k \in \mathbb{N} \mid w_k^i > 0\} = \min\{k \in \mathbb{N} \mid k/n \geq \bar{c}_d\} \leq \bar{k}.$$

Hence, for $k = 1, 2, \ldots, \tilde{k}-1$,

$$\sum_{j=1}^{k} w_j^i < 1 \text{ for all } i \in D.$$

If $\tilde{k} = \bar{k}$, condition (ii) is – in combination with (3) – satisfied for $\bar{k}$ and any $D \subseteq N$. Otherwise, we have $\tilde{k} < \bar{k}$ and then, for $k = \tilde{k}, \tilde{k}+1, \ldots, \bar{k}-1$,

$$\sum_{j=1}^{k} w_j^i = 1 \text{ for all } i \in D.$$

This implies in combination with (3) that condition (ii) is only satisfied if $\max_{k=\tilde{k},\tilde{k}+1,\ldots,\bar{k}-1}(n-k+1) = n-\tilde{k}+1$ agents do not deviate, i.e.,

$$|D| \leq n-(n-\tilde{k}+1) = \tilde{k}-1 \;\Leftrightarrow\; |D| \lneq \tilde{k} \;\Leftrightarrow\; |D| \lneq \lceil \bar{c}_d n \rceil.$$

Thus, (i) and (ii) of Theorem 1 are satisfied for $\bar{k}$ if $|D| \lneq \lceil \bar{c}_d n \rceil$, which finishes the proof.

(iii) Analogous to the second part.

Chapter 3

Trust and Manipulation in Social Networks

3.1 Introduction

Individuals often rely on social connections (friends, neighbors and coworkers as well as political actors and news sources) to form beliefs or opinions on various economic, political or social issues. Every day individuals make decisions on the basis of these beliefs. For instance, when an individual goes to the polls, her choice to vote for one of the candidates is influenced by her friends and peers, her distant and close family members, and some leaders that she listens to and respects. At the same time, the support of others is crucial to enforce interests in society. In politics, majorities are needed to pass laws and in companies, decisions might be taken by a hierarchical superior. It is therefore advantageous for individuals to increase their influence on others and to *manipulate* the way others form their beliefs. This behavior is often referred to as lobbying and widely observed in society, especially in politics.[46] Hence, it is important to understand how beliefs and behaviors evolve over time when individuals can manipulate the trust of others. Can manipulation enable a segregated society to reach a consensus about some issue of broad interest? How long does it take for beliefs to reach consensus when agents can manipulate others? Can manipulation lead a society of agents who communicate and update naïvely to more efficient information aggregation?

We consider a model of opinion formation where agents repeatedly communicate with their neighbors in the social network, can exert some effort to manipulate the

[46]See Gullberg (2008) for lobbying on climate policy in the European Union, and Austen-Smith and Wright (1994) for lobbying on US Supreme Court nominations.

trust of others, and update their opinions taking weighted averages of neighbors' opinions. At each period, first one agent is selected randomly and can exert effort to manipulate the social trust of an agent of her choice. If she decides to provide some costly effort to manipulate another agent, then the manipulated agent weights relatively more the belief of the agent who manipulated her when updating her beliefs. Second, all agents communicate with their neighbors and update their beliefs using the DeGroot updating rule, see DeGroot (1974). This updating process is simple: using her (possibly manipulated) weights, an agent's new belief is the weighted average of her neighbors' beliefs (and possibly her own belief) from the previous period. When agents have no incentives to manipulate each other, the model coincides with the classical DeGroot model of opinion formation.

The DeGroot updating rule assumes that agents are boundedly rational, failing to adjust correctly for repetitions and dependencies in information that they hear multiple times. Since social networks are often fairly complex, it seems reasonable to use an approach where agents fail to update beliefs correctly.[47] Chandrasekhar et al. (2012) provide evidence from a framed field experiment that DeGroot "rule of thumb" models best describe features of empirical social learning. They run a unique lab experiment in the field across 19 villages in rural Karnataka, India, to discriminate between the two leading classes of social learning models – Bayesian learning models versus DeGroot models.[48] They find evidence that the DeGroot model better explains the data than the Bayesian learning model at the network level.[49] At the individual level, they find that the DeGroot model performs much better than Bayesian learning in explaining the actions of an individual given a history of play.[50]

Manipulation is modeled as a communicative or interactional practice, where the manipulating agent exercises some control over the manipulated agent against her

[47] Choi et al. (2012) report an experimental investigation of learning in three-person networks and find that already in simple three-person networks people fail to account for repeated information. They argue that the Quantal Response Equilibrium (QRE) model can account for the behavior observed in the laboratory in a variety of networks and informational settings.

[48] Notice that in order to compare the two concepts, they study DeGroot action models, i.e., agents take an action after aggregating the actions of their neighbors using the DeGroot updating rule.

[49] At the network level (i.e., when the observational unit is the sequence of actions), the Bayesian learning model explains 62% of the actions taken by individuals while the degree weighting DeGroot model explains 76% of the actions taken by individuals.

[50] At the individual level (i.e., when the observational unit is the action of an individual given a history), both the degree weighting and the uniform DeGroot model largely outperform Bayesian learning models.

will. In this sense, manipulation is illegitimate, see Van Dijk (2006). Agents only engage in manipulation if it is worth the effort. They face a trade-off between their increase in satisfaction with the opinions (and possibly the trust itself) of the other agents and the cost of manipulation. In examples, we will frequently use a utility model where agents prefer each other agent's opinion one step ahead to be as close as possible to their current opinion. This reflects the idea that the support of others is necessary to enforce interests. Agents will only engage in manipulation when it brings the opinion of the manipulated agent sufficiently closer to their current opinion compared to the cost of doing so. In our view, this constitutes a natural way to model lobbying incentives.

We first show that manipulation can modify the trust structure. If the society is split up into several disconnected clusters of agents and there are also some agents outside these clusters, then the latter agents might connect different clusters by manipulating the agents therein. Such an agent, previously outside any of these clusters, would not only get influential on the agents therein, but also serve as a bridge and connect them. As we show by means of an example, this can lead to a connected society, and thus, make the society reaching a consensus.

Second, we analyze the long-run beliefs and show that manipulation fosters opinion leadership in the sense that the manipulating agent always increases her influence on the long-run beliefs. For the other agents, this is ambiguous and depends on the social network. Surprisingly, the manipulated agent may thus even gain influence on the long-run opinions. As a consequence, the expected change of influence on the long-run beliefs is ambiguous and depends on the agents' preferences and the social network. We also show that a definitive trust structure evolves in the society and, if the satisfaction of agents only depends on the current and future opinions and not directly on the trust, manipulation will come to an end and they reach a consensus (under some weak regularity condition). At some point, opinions become too similar to be manipulated. Furthermore, we discuss the speed of convergence and note that manipulation can accelerate or slow down convergence. In particular, in sufficiently homophilic societies, i.e., societies where agents tend to trust those agents who are similar to them, and where costs of manipulation are rather high compared to its benefits, manipulation accelerates convergence if it decreases homophily and otherwise it slows down convergence.

Finally, we investigate the tension between information aggregation and spread of misinformation. We find that if manipulation is rather costly and the agents underselling their information gain and those overselling their information lose overall influence (i.e., influence in terms of their initial information), then manipulation re-

duces misinformation and agents converge jointly to more accurate opinions about some underlying true state. In particular, this means that an agent for whom manipulation is cheap can severely harm information aggregation.

There is a large and growing literature on learning in social networks. Models of social learning either use a Bayesian perspective or exploit some plausible rule of thumb behavior.[51] We consider a model of non-Bayesian learning over a social network closely related to DeGroot (1974), DeMarzo et al. (2003), Golub and Jackson (2010) and Acemoglu et al. (2010). DeMarzo et al. (2003) consider a DeGroot rule of thumb model of opinion formation and they show that persuasion bias affects the long-run process of social opinion formation because agents fail to account for the repetition of information propagating through the network. Golub and Jackson (2010) study learning in an environment where agents receive independent noisy signals about the true state and then repeatedly communicate with each other. They find that all opinions in a large society converge to the truth if and only if the influence of the most influential agent vanishes as the society grows.[52] Acemoglu et al. (2010) investigate the tension between information aggregation and spread of misinformation. They characterize how the presence of forceful agents affects information aggregation. Forceful agents influence the beliefs of the other agents they meet, but do not change their own opinions. Under the assumption that even forceful agents obtain some information from others, they show that all beliefs converge to a stochastic consensus. They quantify the extent of misinformation by providing bounds on the gap between the consensus value and the benchmark without forceful agents where there is efficient information aggregation.[53] Friedkin (1991) studies measures to identify opinion leaders in a model related to DeGroot. Recently, Büchel et al. (2012) develop a model of opinion formation where agents may state an opinion that differs from their true opinion because agents have preferences for conformity. They find that lower conformity fosters opinion leadership. In addition, the society becomes wiser if agents who are well informed are less conform, while uninformed agents conform more with their neighbors.

[51] Acemoglu et al. (2011) develop a model of Bayesian learning over general social networks, and Acemoglu and Ozdaglar (2011) provide an overview of recent research on opinion dynamics and learning in social networks.

[52] Golub and Jackson (2012) examine how the speed of learning and best-response processes depend on homophily. They find that convergence to a consensus is slowed down by the presence of homophily but is not influenced by network density.

[53] In contrast to the averaging model, Acemoglu et al. (2010) have a model of pairwise interactions. Without forceful agents, if a pair meets two periods in a row, then in the second meeting there is no information to exchange and no change in beliefs takes place.

We go further by allowing agents to manipulate the trust of others and we find that the implications of manipulation are non-negligible for opinion leadership, reaching a consensus, and aggregating dispersed information.

The paper is organized as follows. In Section 3.2 we introduce the model of opinion formation. In Section 3.3 we show how manipulation can change the trust structure of society. Section 3.4 looks at the long-run effects of manipulation. In Section 3.5 we investigate how manipulation affects the extent of misinformation in society. Section 3.6 concludes. The proofs are presented in Appendix 3.A.

3.2 Model and Notation

Let $\mathcal{N} = \{1, 2, \ldots, n\}$ be the set of agents who have to take a decision on some issue and repeatedly communicate with their neighbors in the social network. Each agent $i \in \mathcal{N}$ has an initial *opinion* or *belief* $x_i(0) \in \mathbb{R}$ about the issue and an initial vector of *social trust* $m_i(0) = (m_{i1}(0), m_{i2}(0), \ldots, m_{in}(0))$, with $0 \leq m_{ij}(0) \leq 1$ for all $j \in \mathcal{N}$ and $\sum_{j\in\mathcal{N}} m_{ij}(0) = 1$, that captures how much attention agent i pays (initially) to each of the other agents. More precisely, $m_{ij}(0)$ is the initial weight or trust that agent i places on the current belief of agent j in forming her updated belief. For $i = j$, $m_{ii}(0)$ can be interpreted as how much agent i is confident in her own initial opinion.

At period $t \in \mathbb{N}$, the agents' beliefs are represented by the vector $x(t) = (x_1(t), x_2(t), \ldots, x_n(t))' \in \mathbb{R}^n$ and their social trust by the matrix $M(t) = (m_{ij}(t))_{i,j\in\mathcal{N}}$.[54] First, one agent is chosen (probability $1/n$ for each agent) to meet and to have the opportunity to manipulate an agent of her choice. If agent $i \in \mathcal{N}$ is chosen at t, she can decide which agent j to meet and furthermore how much effort $\alpha \geq 0$ she would like to exert on j. We write $E(t) = (i; j, \alpha)$ when agent i is chosen to manipulate at t and decides to exert effort α on j. The decision of agent i leads to the following updated trust weights of agent j:

$$m_{jk}(t+1) = \begin{cases} m_{jk}(t)/\left(1+\alpha\right) & \text{if } k \neq i \\ \left(m_{jk}(t)+\alpha\right)/\left(1+\alpha\right) & \text{if } k = i \end{cases}.$$

The trust of j in i increases with the effort i invests and all trust weights of j are normalized. Notice that we assume for simplicity that the trust of j in an agent other than i decreases by the factor $1/(1+\alpha)$, i.e., the absolute decrease in trust is proportional to its level. If i decides not to invest any effort, the trust matrix does not change. We denote the resulting updated trust matrix by $M(t+1) = [M(t)](i; j, \alpha)$.

[54] We denote the transpose of a vector (matrix) x by x'.

Agent i decides on which agent to meet and on how much effort to exert according to her utility function

$$u_i\big(M(t), x(t); j, \alpha\big) = v_i\big([M(t)](i; j, \alpha), x(t)\big) - c_i(j, \alpha),$$

where $v_i\big([M(t)](i; j, \alpha), x(t)\big)$ represents her *satisfaction* with the other agents' opinions and trust resulting from her decision (j, α) and $c_i(j, \alpha)$ represents its *cost*. We assume that v_i is continuous in all arguments and that for all $j \neq i$, $c_i(j, \alpha)$ is strictly increasing in $\alpha \geq 0$, continuous and strictly convex in $\alpha > 0$, and that $c_i(j, 0) = 0$. Note that these conditions ensure that there is always an optimal level of effort $\alpha^*(j)$ given agent i decided to manipulate j.[55] Agent i's optimal choice is then $(j^*, \alpha^*(j^*))$ such that $j^* \in \operatorname{argmax}_{j \neq i} u_i\big(M(t), x(t); j, \alpha^*(j)\big)$.

Secondly, all agents communicate with their neighbors and update their beliefs using the updated trust weights:

$$x(t+1) = [x(t)](i; j, \alpha) = M(t+1)x(t) = [M(t)](i; j, \alpha)x(t).$$

In the sequel, we will often simply write $x(t+1)$ and omit the dependence on the agent selected to manipulate and her choice (j, α). We can rewrite this equation as $x(t+1) = \overline{M}(t+1)x(0)$, where $\overline{M}(t+1) = M(t+1)M(t)\cdots M(1)$ (and $\overline{M}(t) = I_n$ for $t < 1$, where I_n is the $n \times n$ identity matrix) denotes the *overall trust matrix*.

Now, let us give some examples of satisfaction functions that fulfill our assumptions.

Example 12 (Satisfaction functions)**.**

(i) Let $\gamma \in \mathbb{N}$ and

$$v_i\big([M(t)](i; j, \alpha), x(t)\big) = -\frac{1}{n-1} \sum_{k \neq i} \Big(x_i(t) - \big(M(t+1)^\gamma \, x(t)\big)_k\Big)^2,$$

where $M(t+1) = [M(t)](i; j, \alpha)$. That is, agent i's objective is that each other agent's opinion γ periods ahead is as close as possible to her current opinion, disregarding possible manipulations in future periods.

(ii)

$$v_i\big([M(t)](i; j, \alpha), x(t)\big) = -\left(x_i(t) - \frac{1}{n-1} \sum_{k \neq i} x_k(t+1)\right)^2,$$

[55] Note that for all j, $v_i(M(i; j, \alpha), x)$ is continuous in α and bounded from above since $v_i(\cdot, x)$ is bounded from above on the compact set $[0, 1]^{n \times n}$ for all $x \in \mathbb{R}^n$. In total, the utility is continuous in $\alpha > 0$ and since the costs are strictly increasing and strictly convex in $\alpha > 0$, there always exists an optimal level of effort, which might not be unique, though.

where $x_k(t+1) = \big([M(t)](i;j,\alpha)x(t)\big)_k$. That is, agent i wants to be close to the average opinion in society one period ahead, but disregards differences on the individual level.

We will frequently choose in examples the first satisfaction function with parameter $\gamma = 1$, together with a cost function that combines fixed costs and quadratic costs of effort.

Remark 2. If we choose satisfaction functions $v_i \equiv v$ for some constant v and all $i \in \mathcal{N}$, then agents do not have any incentive to exert effort and our model reverts to the classical model of DeGroot (1974).

We now introduce the notion of *consensus*. Whether or not a consensus is reached in the limit depends generally on the initial opinions.

Definition 9 (Consensus)**.** We say that a group of agents $G \subseteq \mathcal{N}$ reaches a *consensus* given initial opinions $(x_i(0))_{i\in\mathcal{N}}$, if there exists $x(\infty) \in \mathbb{R}$ such that

$$\lim_{t\to\infty} x_i(t) = x(\infty) \text{ for all } i \in G.$$

3.3 The Trust Structure

We investigate how manipulation can modify the structure of interaction or trust in society. We first shortly recall some graph-theoretic terminology.[56] We call a group of agents $C \subseteq \mathcal{N}$ *minimal closed* at period t if these agents only trust agents inside the group, i.e., $\sum_{j\in C} m_{ij}(t) = 1$ for all $i \in C$, and if this property does not hold for a proper subset $C' \subsetneq C$. The set of minimal closed groups at period t is denoted $\mathcal{C}(t)$ and is called the *trust structure*. A walk at period t of length K is a sequence of agents $i_1, i_2, \ldots, i_{K+1}$ such that $m_{i_k,i_{k+1}}(t) > 0$ for all $k = 1, 2, \ldots, K$. A walk is a *path* if all agents are distinct. A *cycle* is a walk that starts and ends in the same agent. A cycle is *simple* if only the starting agent appears twice in the cycle. We say that a minimal closed group of agents $C \in \mathcal{C}(t)$ is *aperiodic* if the greatest common divisor[57] of the lengths of simple cycles involving agents from C is 1.[58] Note that this is fulfilled if $m_{ii}(t) > 0$ for some $i \in C$.

[56] See Golub and Jackson (2010).

[57] For a set of integers $S \subseteq \mathbb{N}$, $\gcd(S) = \max\{k \in \mathbb{N} \mid m/k \in \mathbb{N} \text{ for all } m \in S\}$ denotes the greatest common divisor.

[58] Note that if one agent in a simple cycle is from a minimal closed group, then so are all.

At each period t, we can decompose the set of agents $\mathcal{N}$ into minimal closed groups and agents outside these groups, the *rest of the world*, $R(t)$:

$$\mathcal{N} = \bigcup_{C \in \mathcal{C}(t)} C \cup R(t).$$

Within minimal closed groups, all agents interact indirectly with each other, i.e., there is a path between any two agents. We say that the agents are *strongly connected*. For this reason, minimal closed groups are also called strongly connected and closed groups, see Golub and Jackson (2010). Moreover, agent $i \in \mathcal{N}$ is part of the rest of the world $R(t)$ if and only if there is a path at period t from her to some agent in a minimal closed group $C \not\ni i$.

We say that a manipulation at period t does not change the trust structure if $\mathcal{C}(t+1) = \mathcal{C}(t)$. It also implies that $R(t+1) = R(t)$. We find that manipulation changes the trust structure when the manipulated agent belongs to a minimal closed group and additionally the manipulating agent does not belong to this group, but may well belong to another minimal closed group. In the latter case, the group of the manipulated agent is disbanded since it is not anymore closed and its agents join the rest of the world. However, if the manipulating agent does not belong to a minimal closed group, the effect on the group of the manipulated agent depends on the trust structure. Apart from being disbanded, it can also be the case that the manipulating agent and possibly others from the rest of the world join the group of the manipulated agent.

Proposition 8. *Suppose that $E(t) = (i; j, \alpha)$, $\alpha > 0$, at period t.*

(i) Let $i \in \mathcal{N}, j \in R(t)$ or $i, j \in C \in \mathcal{C}(t)$. Then, the trust structure does not change.

(ii) Let $i \in C \in \mathcal{C}(t)$ and $j \in C' \in \mathcal{C}(t) \backslash \{C\}$. Then, C' is disbanded, i.e., $\mathcal{C}(t+1) = \mathcal{C}(t) \backslash \{C'\}$.

(iii) Let $i \in R(t)$ and $j \in C \in \mathcal{C}(t)$.

(a) Suppose that there exists no path from i to k for any $k \in \cup_{C' \in \mathcal{C}(t) \backslash \{C\}} C'$. Then, $R' \cup \{i\}$ joins C, i.e.,

$$\mathcal{C}(t+1) = \mathcal{C}(t) \backslash \{C\} \cup \{C \cup R' \cup \{i\}\},$$

where $R' = \{l \in R(t) \backslash \{i\} \mid \text{there is a path from } i \text{ to } l\}$.

(b) Suppose that there exists $C' \in \mathcal{C}(t) \backslash \{C\}$ such that there exists a path from i to some $k \in C'$. Then, C is disbanded.

All proofs can be found in Appendix 3.A. The following example shows that manipulation can enable a society to reach a consensus due to changes in the trust structure.

Example 13 (Consensus due to manipulation)**.** Take $\mathcal{N} = \{1, 2, 3\}$ and assume that

$$u_i\big(M(t), x(t); j, \alpha\big) = -\frac{1}{2}\sum_{k\neq i}\big(x_i(t) - x_k(t+1)\big)^2 - \big(\alpha^2 + 1/10 \cdot \mathbf{1}_{\{\alpha>0\}}(\alpha)\big)$$

for all $i \in \mathcal{N}$. Notice that the first part of the utility is the satisfaction function in Example 12 part (i) with parameter $\gamma = 1$, while the second part, the costs of effort, combines fixed costs, here $1/10$, and quadratic costs of effort. Let $x(0) = (10, 5, -5)'$ be the vector of initial opinions and

$$M(0) = \begin{pmatrix} .8 & .2 & 0 \\ .4 & .6 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

be the initial trust matrix. Hence, $\mathcal{C}(0) = \{\{1, 2\}, \{3\}\}$. Suppose that first agent 1 and then agent 3 are drawn to meet another agent. Then, at period 0, agent 1's optimal decision is to exert $\alpha = 2.54$[59] effort on agent 3. The trust of the latter is updated to

$$m_3(1) = (.72, 0, .28)\,,$$

while the others' trust does not change, i.e., $m_i(1) = m_i(0)$ for $i = 1, 2$, and the updated opinions become

$$x(1) = M(1)x(0) = (9, 7, 5.76)'\,.$$

Notice that the group of agent 3 is disbanded (see part (ii) of Proposition 8). In the next period, agent 3's optimal decision is to exert $\alpha = .75$ effort on agent 1. This results in the following updated trust matrix:

$$M(2) = \begin{pmatrix} .46 & .11 & .43 \\ .4 & .6 & 0 \\ .72 & 0 & .28 \end{pmatrix}.$$

Notice that agent 3 joins group $\{1, 2\}$ (see part (iii,a) of Proposition 8) and therefore, $\mathcal{N}$ is minimal closed, which implies that the group will reach a consensus, as we will see later on.

[59] Stated values are rounded to two decimals for clarity reasons.

However, notice that if instead of agent 3 another agent is drawn in period 1, then agent 3 never manipulates since when finally she would have the opportunity, her opinion is already close to the others' opinions and therefore, she stays disconnected from them. Nevertheless, the agents would still reach a consensus in this case due to the manipulation at period 0. Since agent 3 trusts agent 1, she follows the consensus that is reached by the first two agents.

3.4 The Long-Run Dynamics

We now look at the long-run effects of manipulation. First, we study the consequences of a single manipulation on the long-run opinions of minimal closed groups. In this context, we are interested in the role of manipulation in opinion leadership. Secondly, we investigate the outcome of the influence process. Finally, we discuss how manipulation affects the speed of convergence of minimal closed groups and illustrate our results by means of an example.

3.4.1 Opinion Leadership

Typically, an agent is called *opinion leader* if she has substantial influence on the long-run beliefs of a group. That is, if she is among the most influential agents in the group. Intuitively, manipulating others should increase her influence on the long-run beliefs and thus foster opinion leadership.

To investigate this issue, we need a measure for how remotely agents are located from each other in the network, i.e., how directly agents trust other agents. For this purpose, we can make use of results from Markov chain theory. Let $(X_s^{(t)})_{s=0}^{\infty}$ denote the homogeneous Markov chain induced by the transition matrix $M(t)$. The agents are then interpreted as states of the Markov chain and the trust of i in j, $m_{ij}(t)$, is interpreted as the transition probability from state i to state j. Then, the *mean first passage time* from state i to state j is defined as $\mathbb{E}[\inf\{s \geq 0 \mid X_s^{(t)} = j\} \mid X_0^{(t)} = i]$. Given the current state of the Markov chain is i, the mean first passage time to j is the expected time it takes for the chain to reach state j.

In other words, the mean first passage time from i to j corresponds to the average (expected) length of a random walk on the weighted network $M(t)$ from i to j that takes each link with probability equal to the assigned weight.[60] This average length

[60] More precisely, it is a random walk on the state space $\mathcal{N}$ that, if currently in state k, travels to state l with probability $m_{kl}(t)$. The length of this random walk to j is the time it takes for it to reach state j.

is small if the weights along short paths from i to j are high, i.e., if agent i trusts agent j rather directly. We therefore call this measure *weighted remoteness* of j from i.

Definition 10 (Weighted remoteness)**.** Take $i, j \in \mathcal{N}$, $i \neq j$. The *weighted remoteness* at period t of agent j from agent i is given by

$$r_{ij}(t) = \mathbb{E}[\inf\{s \geq 0 \mid X_s^{(t)} = j\} \mid X_0^{(t)} = i],$$

where $(X_s^{(t)})_{s=0}^{\infty}$ is the homogeneous Markov chain induced by $M(t)$.

The following remark shows that the weighted remoteness attains its minimum when i trusts solely j.

Remark 3. Take $i, j \in \mathcal{N}$, $i \neq j$.

(i) $r_{ij}(t) \geq 1$,

(ii) $r_{ij}(t) < +\infty$ if and only if there is a path from i to j, and, in particular, if $i, j \in C \in \mathcal{C}(t)$,

(iii) $r_{ij}(t) = 1$ if and only if $m_{ij}(t) = 1$.

To provide some more intuition, let us look at an alternative (implicit) formula for the weighted remoteness. Suppose that $i, j \in C \in \mathcal{C}(t)$ are two distinct agents in a minimal closed group. By part (ii) of Remark 3, the weighted remoteness is finite for all pairs of agents in that group. The unique walk from i to j with (average) length 1 is assigned weight (or has probability, when interpreted as a random walk) $m_{ij}(t)$. And the average length of walks to j that first pass through $k \in C\backslash\{j\}$ is $r_{kj}(t) + 1$, i.e., walks from i to j with average length $r_{kj}(t) + 1$ are assigned weight (have probability) $m_{ik}(t)$. Thus,

$$r_{ij}(t) = m_{ij}(t) + \sum_{k \in C\backslash\{j\}} m_{ik}(t)(r_{kj}(t) + 1) \ .$$

Finally, applying $\sum_{k \in C} m_{ik}(t) = 1$ leads to the following remark.

Remark 4. Take $i, j \in C \in \mathcal{C}(t)$, $i \neq j$. Then,

$$r_{ij}(t) = 1 + \sum_{k \in C\backslash\{j\}} m_{ik}(t) r_{kj}(t).$$

Note that computing the weighted remoteness using this formula amounts to solving a linear system of $|C|(|C|-1)$ equations, which has a unique solution.

We denote by $\pi(C;t)$ the probability vector of the agents' influence on the final consensus of their group $C \in \mathcal{C}(t)$ at period t, given that the group is aperiodic and the trust matrix does not change any more.[61] In this case, the group converges to

$$x(\infty) = \pi(C;t)'\, x(t)|_C = \sum_{i \in C} \pi_i(C;t) x_i(t),$$

where $x(t)|_C = (x_i(t))_{i \in C}$ is the restriction of $x(t)$ to agents in C. In other words, $\pi_i(C;t)$, $i \in C$, is the influence weight of agent i's opinion at period t, $x_i(t)$, on the consensus of C. Notice that the influence vector $\pi(C;t)$ depends on the trust matrix $M(t)$ and therefore it changes with manipulation. A higher value of $\pi_i(C;t)$ corresponds to more influence of agent i on the consensus. Each agent in a minimal closed group has at least some influence on the consensus: $\pi_i(C;t) > 0$ for all $i \in C$.[62]

We now turn back to the long-run consequences of manipulation and thus, opinion leaders. We restrict our analysis to the case where both the manipulating and the manipulated agent are in the same minimal closed group. Since in this case the trust structure is preserved we can compare the influence on the long-run consensus of the group before and after manipulation.

Proposition 9. *Suppose that at period t, group $C \in \mathcal{C}(t)$ is aperiodic and $E(t) = (i;j,\alpha)$, $i,j \in C$. Then, aperiodicity is preserved and the influence of agent $k \in C$ on the final consensus of her group changes as follows,*

$$\pi_k(C;t+1) - \pi_k(C;t) =$$
$$\begin{cases} \alpha/(1+\alpha)\pi_i(C;t)\pi_j(C;t+1)\sum_{l \in C\setminus\{i\}} m_{jl}(t) r_{li}(t) & \text{if } k = i \\ \alpha/(1+\alpha)\pi_k(C;t)\pi_j(C;t+1)\left(\sum_{l \in C\setminus\{k\}} m_{jl}(t) r_{lk}(t) - r_{ik}(t)\right) & \text{if } k \neq i \end{cases}.$$

Corollary 7. *Suppose that at period t, group $C \in \mathcal{C}(t)$ is aperiodic and $E(t) = (i;j,\alpha)$, $i,j \in C$. If $\alpha > 0$, then*

(i) *agent i strictly increases her long-run influence, $\pi_i(C;t+1) > \pi_i(C;t)$,*

[61] In the language of Markov chains, $\pi(C;t)$ is known as the unique stationary distribution of the aperiodic communication class C. Without aperiodicity, the class might fail to converge to consensus.

[62] See Golub and Jackson (2010).

(*ii*) *any other agent $k \neq i$ of the group can either gain or lose influence, depending on the trust matrix. She gains if and only if*

$$\sum_{l \in C \setminus \{k,i\}} m_{jl}(t)\big(r_{lk}(t) - r_{ik}(t)\big) > m_{jk}(t) r_{ik}(t),$$

(*iii*) *agent $k \neq i, j$ loses influence for sure if j trusts solely her, i.e., $m_{jk}(t) = 1$.*

Proposition 9 tells us that the change in long-run influence for any agent k depends on the effort agent i exerts to manipulate agent j, agent k's current long-run influence and the future long-run influence of the manipulated agent j. In particular, the magnitude of the change increases with i's effort, and it is zero if agent i does not exert any effort. Furthermore, notice that dividing both sides by agent k's current long-run influence, $\pi_k(C;t)$, yields the relative change in her long-run influence.

When agent $k = i$, we find that this change is strictly positive whenever she exerts some effort. In this sense, manipulation fosters opinion leadership. It is large if the weighted remoteness of i from agents (other than i) that are significantly trusted by j is large. To understand this better, notice that the long-run influence of an agent depends on how much she is trusted by agents that are trusted. Or, in other words, an agent is influential if she is influential on other influential agents. Thus, there is a direct gain of influence due to an increase of trust from j and an indirect loss of influence (that is always dominated by the direct gain) due to a decrease of trust from j faced by agents that (indirectly) trust i. This explains why it is better for i if agents facing a large decrease of trust from j (those trusted much by j) do not (indirectly) trust i much, i.e., i has a large weighted remoteness from them.

For any other agent $k \neq i$, it turns out that the change can be positive or negative. It is positive if, broadly speaking, j does not trust k a lot, the weighted remoteness of k from i is small and furthermore the weighted remoteness of k from agents (other than i) that are significantly trusted by j is larger than that from i. In other words, it is positive if the manipulating agent, who gains influence for sure, (indirectly) trusts agent k significantly (small weighted remoteness of k from i), k does not face a large decrease of trust from j and those agents facing a large decrease from j (those trusted much by j) (indirectly) trust k less than i does.

Notice that for any agent $k \neq i, j$, this is a trade-off between an indirect gain of trust due to the increase of trust that i obtains from j, on the one hand, and an indirect loss of influence due to a decrease of trust from j faced by agents that (indirectly) trust k as well as the direct loss of influence due to a decrease of trust

from j, on the other hand. In the extreme case where j only trusts k, the direct loss of influence dominates the indirect gain of influence for sure.

In particular, it means that even the manipulated agent j can gain influence. In a sense, such an agent would like to be manipulated because she trusts the "wrong" agents. For agent j, being manipulated is positive if her weighted remoteness from agents she trusts significantly is large and furthermore, her weighted remoteness from i is small. Hence, it is positive if the manipulating agent (indirectly) trusts her significantly (small weighted remoteness from i) and agents facing a large decrease of trust from her (those she trusts) do not (indirectly) trust her much. Here, the trade-off is between the indirect gain of trust due to the increase of trust that i obtains from her and the indirect loss of influence due to a decrease of trust from her faced by agents that (indirectly) trust her. Note that the gain of influence is particularly large if the manipulating agent trusts j significantly.

The next example shows that indeed in some situations an agent can gain from being manipulated in the sense that her influence on the long-run beliefs increases.

Example 14 (Being manipulated can increase influence)**.** Take $\mathcal{N} = \{1, 2, 3\}$ and assume that

$$M(0) = \begin{pmatrix} .25 & .25 & .5 \\ .5 & .5 & 0 \\ .4 & .5 & .1 \end{pmatrix}$$

is the initial trust matrix. Notice that $\mathcal{N}$ is minimal closed. Suppose that agent 1 has the opportunity to meet another agent and decides to exert effort $\alpha > 0$ on agent 3. Then, from Proposition 9, we get

$$\begin{aligned} \pi_3(\mathcal{N}; 1) - \pi_3(\mathcal{N}; 0) &= \frac{\alpha}{1+\alpha} \pi_3(\mathcal{N}; 0)\pi_3(\mathcal{N}; 1) \sum_{l=1,2} m_{3l}(0) r_{l3}(0) - r_{13}(0) \\ &= \frac{\alpha}{1+\alpha} \pi_3(\mathcal{N}; 0)\pi_3(\mathcal{N}; 1) \frac{7}{10} > 0, \end{aligned}$$

since $\pi_3(\mathcal{N}; 0), \pi_3(\mathcal{N}; 1) > 0$. Hence, being manipulated by agent 1 increases agent 3's influence on the long-run beliefs. The reason is that, initially, she trusts too much agent 2 – an agent that does not trust her at all. She gains influence from agent 1's increase of influence on the long-run beliefs since this agent trusts her. In other words, after being manipulated she is trusted by an agent that is trusted more.

Furthermore, we can use Proposition 9 to compare the expected influence on the long-run consensus of society before and after manipulation when all agents are in

the same minimal closed group.[63] For this result we need to slightly change our notation. We denote the decision of agent $i \in \mathcal{N}$ when she is selected to meet another agent by $\big(j(i), \alpha(i; j(i))\big)$, i.e., agent i decides to exert effort $\alpha(i; j(i))$ on agent $j(i)$.

Corollary 8. *Suppose that at period t, $\mathcal{C}(t) = \{\mathcal{N}\}$ and that $\mathcal{N}$ is aperiodic. Then, aperiodicity is preserved and, in expectation, the influence of agent $k \in \mathcal{N}$ on the final consensus of the society changes as follows from period t to $t+1$,*

$$\mathbb{E}[\pi_k(\mathcal{N}; t+1) - \pi_k(\mathcal{N}; t) \mid M(t), x(t)] =$$
$$\frac{\pi_k(\mathcal{N}; t)}{n}\left[\sum_{i\in\mathcal{N}}\left(\frac{\alpha(i; j(i))}{1+\alpha(i; j(i))}\pi_{j(i)}(\mathcal{N}; t+1)\sum_{l\neq k} m_{j(i)l}(t) r_{lk}(t)\right) -\right.$$
$$\left.\sum_{i\neq k}\frac{\alpha(i; j(i))}{1+\alpha(i; j(i))}\pi_{j(i)}(\mathcal{N}; t+1) r_{ik}(t)\right].$$

Notice that an agent gains long-run influence in expectation if and only if the term in the square brackets is positive. For this to hold, it is necessary that $\alpha(i; j(i)) > 0$ for some $i \in \mathcal{N}$ at period t. Moreover, it follows from Corollary 7 part (i) that $\alpha(k; j(k)) > 0$ and $\alpha(i; j(i)) = 0$ for all $i \neq k$ at period t (i.e., only agent k would manipulate if she was selected at t) is a sufficient condition for that she gains influence in expectation. The reason is that agent k gains influence for sure when she manipulates herself, and since no other agent manipulates when selected, she gains in expectation. Notice that by dividing both sides by agent k's current long-run influence, $\pi_k(C; t)$, we get the expected relative change in her long-run influence.

3.4.2 Convergence

We now determine where the process finally converges to. First, we look at the case where all agents are in the same minimal closed group. Given the group is aperiodic, we show that if the satisfaction level only depends on the opinions (before and after manipulation), i.e., a change in trust that does not affect opinions does not change the satisfaction of an agent, and if there is a fixed cost for exerting effort, then manipulation comes to an end, eventually. At some point, opinions in the society become too similar to be manipulated. Second, we determine the final consensus the society converges to.

[63] Notice that if not all agents are in the same minimal closed group, then the group in question could be disbanded with some probability and hence would not anymore reach a consensus.

Lemma 2. *Suppose that $\mathcal{C}(0) = \{\mathcal{N}\}$ and that $\mathcal{N}$ is aperiodic. If for all $i, j \in \mathcal{N}$ and $\alpha > 0$,*

(i) $v_i\big(M(i;j,\alpha),x\big) - v_i\big(M(i;j,0),x\big) \to 0$ *if* $\|x(i;j,\alpha) - x(i;j,0)\| \to 0$, *and*

(ii) $c_i(j,\alpha) \geq \underline{c} > 0$,

then, there exists an almost surely finite stopping time τ such that from period $t = \tau$ on there is no more manipulation, where $\|\cdot\|$ is any norm on $\mathbb{R}^n$.[64] *The society converges to the random variable*

$$x(\infty) = \pi(\mathcal{N};\tau)'\overline{M}(\tau-1)x(0).$$

Now, we turn to the general case of any trust structure. We show that after a finite number of periods, the trust structure settles down. Then, it follows from the above result that, under the beforementioned conditions, manipulation within the minimal closed groups that have finally been formed comes to an end. We also determine the final consensus opinion of each aperiodic minimal closed group.

Proposition 10. *(i) There exists an almost surely finite stopping time τ such that for all $t \geq \tau$, $\mathcal{C}(t) = \mathcal{C}(\tau)$.*

(ii) If $C \in \mathcal{C}(\tau)$ is aperiodic and for all $i, j \in C$, $\alpha > 0$,

(1) $v_i\big(M(i;j,\alpha),x\big) - v_i\big(M(i;j,0),x\big) \to 0$ *if* $\|x(i;j,\alpha) - x(i;j,0)\| \to 0$, *and*

(2) $c_i(j,\alpha) \geq \underline{c} > 0$,

then, there exists an almost surely finite stopping time $\widehat{\tau} \geq \tau$ such that at all periods $t \geq \widehat{\tau}$, agents in C are not manipulated. Moreover, they converge to the random variable

$$x(\infty) = \pi(C;\widehat{\tau})'\, M(\widehat{\tau}-1)|_C\, M(\widehat{\tau}-2)|_C \cdots M(1)|_C\, x(0)|_C\,.$$

In what follows we use τ and $\widehat{\tau}$ in the above sense. We denote by $\overline{\pi}_i(C;t)$ the *overall influence* of agent i's initial opinion on the consensus of group C at period t given no more manipulation affecting C takes place. The overall influence is implicitly given by Proposition 10.

[64] In our context, this means that τ is a random variable such that the event $\tau = t$ only depends on which agents were selected to meet another agent at periods $1, 2, \ldots, t$, and furthermore τ is almost surely finite, i.e., the event $\tau < +\infty$ has probability 1.

Corollary 9. *The overall influence of the initial opinion of agent $i \in \mathcal{N}$ on the consensus of an aperiodic group $C \in \mathcal{C}(\tau)$ is given by*

$$\overline{\pi}_i(C;\widehat{\tau}) = \begin{cases} \left(\pi(C;\widehat{\tau})' \, M(\widehat{\tau}-1)|_C \, M(\widehat{\tau}-2)|_C \cdots M(1)|_C\right)_i & \text{if } i \in C \\ 0 & \text{if } i \notin C \end{cases}.$$

It turns out that an agent outside a minimal closed group that has finally formed can never have any influence on its consensus opinion.

3.4.3 Speed of Convergence

We have seen that within an aperiodic minimal closed group $C \in \mathcal{C}(t)$ agents reach a consensus given that the trust structure does not change anymore. This means that their opinions converge to a common opinion. By *speed of convergence* we mean the time that this convergence takes. That is, it is the time it takes for the expression

$$|x_i(t) - x_i(\infty)|$$

to become small. It is well known that this depends crucially on the second largest eigenvalue $\lambda_2(C;t)$ of the trust matrix $M(t)|_C$, where $M(t)|_C = (m_{ij}(t))_{i,j\in C}$ denotes the restriction of $M(t)$ to agents in C. Notice that $M(t)|_C$ is a stochastic matrix since C is minimal closed. The smaller the eigenvalue in absolute value, the faster the convergence to consensus (see Jackson, 2008).

Thus, the change in the second largest eigenvalue due to manipulation tells us whether the speed of convergence has increased or decreased. In this context, the concept of *homophily* is important, that is, the tendency of people to interact relatively more with those people who are similar to them.[65]

Definition 11 (Homophily)**.** The *homophily* of a group of agents $G \subseteq \mathcal{N}$ at period t is defined as

$$\text{Hom}(G;t) = \frac{1}{|G|}\left(\sum_{i,j\in G} m_{ij}(t) - \sum_{i\in G, j\notin G} m_{ij}(t)\right).$$

The homophily of a group of agents is the normalized difference of their trust in agents inside and outside the group. Notice that a minimal closed group $C \in \mathcal{C}(t)$ attains the maximum homophily, $\text{Hom}(C;t) = 1$. Consider a *cut of society* $(S, \mathcal{N}\backslash S)$, $S \subseteq \mathcal{N}, S \neq \emptyset$, into two groups of agents S and $\mathcal{N}\backslash S$.[66] The next lemma establishes

[65] Notice that we do not model explicitly the characteristics that lead to homophily.

[66] There exist many different notions of homophily in the literature. Our measure is similar to the one used in Golub and Jackson (2012). We can consider the average homophily $(\text{Hom}(S;t) + \text{Hom}(\mathcal{N}\backslash S;t))/2$ with respect to the cut $(S, \mathcal{N}\backslash S)$ as a generalization of degree-weighted homophily to general weighted averages.

that manipulation across the cut decreases homophily, while manipulation within a group increases it.

Lemma 3. *Take a cut of society $(S, \mathcal{N}\backslash S)$. If $i \in \mathcal{N}$ manipulates $j \in S$ at period t, then*

(i) the homophily of S (strictly) increases if $i \in S$ (and $\sum_{k\in S} m_{jk}(t) < 1$), and

(ii) the homophily of S (strictly) decreases if $i \notin S$ (and $\sum_{k\in S} m_{jk}(t) > 0$).

Now, we come back to the speed of convergence. Given the complexity of the problem for $n \geq 3$, we consider an example of a two-agent society that suggests that homophily helps to explain the change in speed of convergence.

Example 15 (Speed of convergence with two agents)**.** Take $\mathcal{N} = \{1, 2\}$ and suppose that at period t, $\mathcal{N}$ is minimal closed and aperiodic. Then, we have that $\lambda_2(\mathcal{N}; t) = m_{11}(t) - m_{21}(t) = m_{22}(t) - m_{12}(t)$. Therefore, we can characterize the change in the second largest eigenvalue as follows:

$$\begin{aligned} |\lambda_2(\mathcal{N}; t+1)| \leq |\lambda_2(\mathcal{N}; t)| &\Leftrightarrow |m_{11}(t+1) - m_{21}(t+1)| \leq |m_{11}(t) - m_{21}(t)| \\ &\Leftrightarrow |m_{22}(t+1) - m_{12}(t+1)| \leq |m_{22}(t) - m_{12}(t)|. \end{aligned}$$

It means that convergence is faster after manipulation if afterwards agents behave more similar, i.e., the trust both agents put on agent 1's opinion is more similar (which implies that also the trust they put on agent 2's opinion is more similar). Thus, if for instance

$$m_{22}(t) > (1 + \alpha) m_{12}(t), \tag{4}$$

then agent 1 manipulating agent 2 accelerates convergence. However, if $m_{22}(t) < m_{12}(t)$, it slows down convergence since manipulation increases the already existing tendency of opinions to oscillate. The more interesting case is the first one, though. We can write (4) as

$$(1 + \alpha)\mathrm{Hom}(\{1\}, t) + \mathrm{Hom}(\{2\}, t) > \alpha,$$

that is, manipulation accelerates convergence if there is sufficient aggregated homophily in the society and agent 1 does not exert too much effort.

The example shows that manipulation can speed up or slow down the convergence process. More important, it suggests that in a sufficiently homophilic society where exerting effort is rather costly, manipulation reducing homophily (i.e., across the cut, see Lemma 3) increases the speed of convergence. Notice that manipulation

increasing homophily (i.e., within one of the groups separated by the cut) is not possible in this simple setting since both groups are singletons. However, it seems plausible that it would slow down convergence in homophilic societies.[67]

3.4.4 Three-agents Example

Finally, let us consider an example with three agents to illustrate the results of this section. We use a utility model that is composed of the satisfaction function in Example 12 (i) and a cost function that combines fixed costs and quadratic costs of effort.

Example 16 (Three-agents society)**.** Take $\mathcal{N} = \{1, 2, 3\}$ and assume that

$$u_i\big(M(t), x(t); j, \alpha\big) = -\frac{1}{2}\sum_{k \neq i}\big(x_i(t) - x_k(t+1)\big)^2 - \big(\alpha^2 + 1/10 \cdot \mathbf{1}_{\{\alpha>0\}}(\alpha)\big)$$

for all $i \in \mathcal{N}$. Let $x(0) = (10, 5, 1)'$ be the vector of initial opinions and

$$M(0) = \begin{pmatrix} .6 & .2 & .2 \\ .1 & .4 & .5 \\ 0 & .6 & .4 \end{pmatrix}$$

be the initial trust matrix. Notice that this society is connected. The vector of initial long-run influence – and of long-run influence in the classical model without manipulation – is $\pi(\mathcal{N}; 0) = \pi_{\text{cl}} = (.12, .46, .42)'$ and the initial speed of convergence is measured by $\lambda_2(\mathcal{N}; 0) = \lambda_{2,\text{cl}} = .55$. At period 0, any agent selected to exert effort would do so. It is either $E(0) = (1; 3, 1.46), (2; 1, .6)$ or $(3; 1, 1.4)$. In expectation, we get $\mathbb{E}[\pi(\mathcal{N}; 1)] = (.2, .41, .39)'$ and $\mathbb{E}[\lambda_2(\mathcal{N}; 1)] = .21$. So, on average agent 1 profits from manipulation. Since initially the other agents almost did not listen to her and also her opinion was far apart from the others' opinions, she exerts significant effort when selected. In particular, the society is homophilic: taking the cut $(\{1\}, \{2, 3\})$, we get

$$\text{Hom}(\{1\}, 0) = .2 \text{ and } \text{Hom}(\{2, 3\}, 0) = .9.$$

So, since with probability one the manipulation is across the cut, the strong decrease in the (expected) second largest eigenvalue supports our suggestion from Section 3.4.3 that manipulation reducing homophily (i.e., across the cut) increases the speed of convergence.

[67] In the above example, increasing homophily is attained by increasing the weight of an agent on herself, which leads to an increase of the second largest eigenvalue in sufficiently homophilic societies.

At the next period, there is only manipulation if at the last period an agent other than agent 3 was selected to manipulate. In expectation, we get $\mathbb{E}[\pi(\mathcal{N};2)] = (.22, .41, .38)'$ and $\mathbb{E}[\lambda_2(\mathcal{N};2)] = .17$. Again, agent 1 profits on average from manipulation, but only slightly since opinions are already closer and since she is not as isolated as in the beginning. The convergence gets, on average, slightly faster as well.

Manipulation ends here, that is, with probability one no agent exerts effort from period 2 on, i.e $M(t) = M(2)$ for all $t \geq 2$. Hence, the expected influence of the agents' initial opinions on the consensus is

$$\mathbb{E}[\overline{\pi}(\mathcal{N};2)'] = \mathbb{E}[\pi(\mathcal{N};2)'\overline{M}(1)] = \mathbb{E}[\pi(\mathcal{N};2)'M(1)] = (.21, .41, .38).$$

Thus, the expected consensus that society reaches is

$$\mathbb{E}[x(\infty)] = \mathbb{E}[\overline{\pi}(\mathcal{N};2)']x(0) = 4.53.$$

Compared to this, the classical model gives $x_{\text{cl}}(\infty) = \pi'_{\text{cl}}x(0) = 3.88$ and hence, our model leads to an average long-run belief of society that is closer to the initial opinion of agent 1 since she is the one who (on average) gains influence due to manipulation.

3.5 The Wisdom of Crowds

We now investigate how manipulation affects the extent of misinformation in society. In this section, we assume that the society forms one minimal closed and aperiodic group. Clearly, societies that are not connected fail to aggregate information.[68] We use an approach similar to Acemoglu et al. (2010) and assume that there is a *true state* $\mu = (1/n)\sum_{i\in\mathcal{N}} x_i(0)$ that corresponds to the average of the initial opinions of the n agents in the society. Information about the true state is dispersed, but can easily be aggregated by the agents: uniform overall influence on the long-run beliefs leads to perfect aggregation of information.[69] Notice that, in general, agents cannot infer the true state from the initial information since they only get to know the information of their neighbors.

[68] However, as in Example 13, we can observe that manipulation leads to a connected society and thus such an event can also be viewed as reducing the extent of misinformation in the society.

[69] We can think of the initial opinions as being drawn independently from some distribution with mean μ. Then, uniform overall influence leads as well to optimal aggregation, the difference being that it is not perfect in this case due to the finite number of samples.

At a given period t, the *wisdom* of the society is measured by the difference between the true state and the consensus they would reach in case no more manipulation takes place:

$$\overline{\pi}(\mathcal{N};t)'x(0) - \mu = \sum_{i\in\mathcal{N}} \left(\overline{\pi}_i(\mathcal{N};t) - \frac{1}{n}\right) x_i(0).$$

Hence, $\|\overline{\pi}(\mathcal{N};t) - (1/n)\mathbb{I}\|_2$ measures the extent of misinformation in the society, where $\mathbb{I} = (1,1,\ldots,1)' \in \mathbb{R}^n$ is a vector of 1s and $\|x\|_2 = \sqrt{\sum_{k\in\mathcal{N}} |x_k|^2}$ is the standard Euclidean norm of $x \in \mathbb{R}^n$. We say that an agent i *undersells* (*oversells*) her information at period t if $\overline{\pi}_i(\mathcal{N};t) < 1/n$ ($\overline{\pi}_i(\mathcal{N};t) > 1/n$). In a sense, an agent underselling her information is, compared to her overall influence, (relatively) well informed.

Definition 12 (Extent of misinformation)**.** A manipulation at period t reduces the *extent of misinformation* in society if

$$\|\overline{\pi}(\mathcal{N};t+1) - (1/n)\mathbb{I}\|_2 < \|\overline{\pi}(\mathcal{N};t) - (1/n)\mathbb{I}\|_2,$$

otherwise, it (weakly) increases the *extent of misinformation*.

The next lemma describes, given some agent manipulates another agent, the change in the overall influence of an agent from period t to period $t+1$.

Lemma 4. *Suppose that $\mathcal{C}(0) = \{\mathcal{N}\}$ and that $\mathcal{N}$ is aperiodic. For $k \in \mathcal{N}$, at period t,*

$$\overline{\pi}_k(\mathcal{N};t+1) - \overline{\pi}_k(\mathcal{N};t) = \sum_{l=1}^{n} \overline{m}_{lk}(t)\big(\pi_l(\mathcal{N};t+1) - \pi_l(\mathcal{N};t)\big).$$

In case there is manipulation at period t, the overall influence of the initial opinion of an agent increases if the agents that overall trust her gain (on average) influence from the manipulation. Next, we provide conditions ensuring that a manipulation reduces the extent of misinformation in the society. First, manipulation should not be too cheap for the agent who is manipulating. Second, only agents underselling their information should gain overall influence. We say that $\overline{\pi}(\mathcal{N};t)$ is *generic* if for all $k \in \mathcal{N}$ it holds that $\overline{\pi}_k(\mathcal{N};t) \neq 1/n$.

Proposition 11. *Suppose that $\mathcal{C}(0) = \{\mathcal{N}\}$, $\mathcal{N}$ is aperiodic and that $\overline{\pi}(\mathcal{N};t)$ is generic. Then, there exists $\overline{\alpha} > 0$ such that $E(t) = (i;j,\alpha)$, $\alpha > 0$, reduces the extent of misinformation if*

(*i*) *$\alpha \leq \overline{\alpha}$, and*

(*ii*) $\sum_{l=1}^{n} \overline{m}_{lk}(t)\big(\pi_l(\mathcal{N};t+1) - \pi_l(\mathcal{N};t)\big) \geq 0$ *if and only if* k *undersells her information at period* t.

Intuitively, condition (ii) says that (relatively) well informed agents (those that undersell their information) should gain overall influence, while (relatively) badly informed agents (those that oversell their information) should lose overall influence. Then, this leads to a distribution of overall influence in the society that is more equal and hence reduces the extent of misinformation in the society – but only if i does not exert too much effort on j (condition (i)). Otherwise, manipulation makes some agents too influential, in particular the manipulating agent, and leads to a distribution of overall influence that is even more unequal than before. In other words, information aggregation can be severely harmed when for some agents manipulation is rather cheap.

We now introduce a true state of the world into Example 16. On average, manipulation reduces the extent of misinformation in each period and the society converges to a more precise consensus.

Example 17 (Three-agents society, cont'd)**.** Recall that $\mathcal{N} = \{1,2,3\}$ and that

$$u_i\big(M(t),x(t);j,\alpha\big) = -\frac{1}{2}\sum_{k\neq i}\big(x_i(t) - x_k(t+1)\big)^2 - \big(\alpha^2 + 1/10\cdot \mathbf{1}_{\{\alpha>0\}}(\alpha)\big)$$

for all $i \in \mathcal{N}$. Furthermore, $x(0) = (10,5,1)'$ and

$$M(0) = \begin{pmatrix} .6 & .2 & .2 \\ .1 & .4 & .5 \\ 0 & .6 & .4 \end{pmatrix}.$$

Hence, $\mu = (1/3)\sum_{i\in\mathcal{N}} x_i(0) = 5.33$ is the true state. The vector of initial overall influence is $\overline{\pi}(\mathcal{N};0) = \pi(\mathcal{N};0) = (.12,.46,.42)'$. Recall that in expectation, we obtain $\mathbb{E}[\overline{\pi}(\mathcal{N};1)] = \mathbb{E}[\pi(\mathcal{N};1)] = (.2,.41,.39)'$, $\mathbb{E}[\overline{\pi}(\mathcal{N};2)] = (.21,.41,.38)'$ and that there is no more manipulation from period 2 on. Thus,

$$\begin{aligned}\|\overline{\pi}(\mathcal{N};0) - (1/3)\mathbb{I}\|_2 = .268 &> \|\mathbb{E}[\overline{\pi}(\mathcal{N};1)] - (1/3)\mathbb{I}\|_2 = .161 \\ &> \|\mathbb{E}[\overline{\pi}(\mathcal{N};2)] - (1/3)\mathbb{I}\|_2 = .158.\end{aligned}$$

So, in terms of the expected long-run influence, manipulation reduces the extent of misinformation in society. And indeed, the agents reach the expected consensus $\mathbb{E}[x(\infty)] = 4.53$, which is closer to the true state $\mu = 5.33$ than the consensus they would have reached in the classical model of DeGroot, $x_{\text{cl}}(\infty) = 3.88$.

This confirms the intuition that manipulation has the most bite in the beginning, before potentially misleading opinions have spread. Furthermore, this example suggests that manipulation can have positive effects on information aggregation if agents have homogeneous preferences for manipulation.

3.6 Conclusion

We investigated the role of manipulation in a model of opinion formation where agents have beliefs about some question of interest and update them taking weighted averages of neighbors' opinions. Our analysis focused on the consequences of manipulation for the trust structure and long-run beliefs in the society, including learning.

We showed that manipulation can modify the trust structure and lead to a connected society, and thus, to consensus. Furthermore, we found that manipulation fosters opinion leadership in the sense that the manipulating agent always increases her influence on the long-run beliefs. And more surprisingly, this may even be the case for the manipulated agent. The expected change of influence on the long-run beliefs is ambiguous and depends on the agents' preferences and the social network.

We also showed that the trust structure of the society settles down and, if the satisfaction of agents does not directly depend on the trust, manipulation will come to an end and they reach a consensus (under some weak regularity condition). To obtain insights on the relation of manipulation and the speed of convergence, we provided examples and argued that in sufficiently homophilic societies where manipulation is rather costly, manipulation accelerates convergence if it decreases homophily and otherwise it slows down convergence.

Regarding learning, we were interested in the question whether manipulation is beneficial or harmful for information aggregation. We used an approach similar to Acemoglu et al. (2010) and showed that manipulation reduces the extent of misinformation in the society if manipulation is rather costly and the agents underselling their information gain and those overselling their information lose overall influence. Not surprisingly, agents for whom manipulation is cheap can severely harm information aggregation. Furthermore, our main example suggests that homogeneous preferences for manipulation favor a reduction of the extent of misinformation in society.

We should notice that manipulation has no bite if we use the approach of Golub and Jackson (2010). They studied large societies and showed that opinions converge to the true state if the influence of the most influential agent in the society is vanishing as the society grows. Under this condition, manipulation does not change

convergence to the true state since its consequences are negligible compared to the size of the society. In large societies, information is aggregated before manipulation (and possibly a series of manipulations) can spread misinformation. The only way manipulation could have consequences for information aggregation in large societies would be to enable agents to manipulate a substantial proportion of the society instead of only one agent. Relaxing the restriction to manipulation of a single agent at a time is left for future work.

We view our paper as first attempt in studying manipulation and misinformation in society. Our approach incorporated strategic considerations in a model of opinion formation à la DeGroot. We made several simplifying assumptions and derived results that apply to general societies. We plan to address some of the open issues in future work, e.g., extending manipulation to groups and allowing for more sophisticated agents.

3.A Appendix

Proof of Proposition 8

(i) Follows immediately since all minimal closed groups remain unchanged.

(ii) If agent i manipulates agent j, then $m_{ji}(t+1) > 0$ and thus, since $C' \ni j$ is minimal closed at period t, there exists a path at $t+1$ from l to i for all $l \in C'$. Since C is still minimal closed, it follows that $R(t+1) = R(t) \cup C'$, i.e., $\mathcal{C}(t+1) = \mathcal{C}(t) \backslash \{C'\}$.

(iii) (a) If agent i manipulates agent j, then it follows that $\sum_{l \in C \cup \{i\}} m_{kl}(t+1) = 1$ for all $k \in C$ since C is closed at t. Furthermore, since by assumption there is no path from i to k for any $k \in \cup_{C' \in \mathcal{C}(t) \backslash \{C\}} C'$ and by definition of R', $\sum_{l \in C \cup R' \cup \{i\}} m_{kl}(t+1) = 1$ for all $k \in R' \cup \{i\}$. Hence, it follows that $\sum_{l \in C \cup R' \cup \{i\}} m_{kl}(t+1) = 1$ for all $k \in C \cup R' \cup \{i\}$, i.e., $C \cup R' \cup \{i\}$ is closed.

Note that moreover, since by assumption there is no path from i to k for any $k \in \cup_{C' \in \mathcal{C}(t) \backslash \{C\}} C'$, there is a path from i to j (otherwise $R' \cup \{i\}$ was closed at t). Thus, since C is minimal closed and i manipulates j, there is a path from k to l for all $k, l \in C \cup \{i\}$ at $t+1$. Then, by definition of R', there is also a path from k to l for all $k \in C \cup \{i\}$ and $l \in R'$. Moreover, again by assumption and definition of R', there exists a path from k to l for all $k \in R'$ and all $l \in C$ (otherwise a subset of R' was closed at t).

Combined, this implies that the same holds for all $k, l \in C \cup R' \cup \{i\}$. Hence, $C \cup R' \cup \{i\}$ is minimal closed, i.e., $\mathcal{C}(t+1) = \mathcal{C}(t) \backslash \{C\} \cup \{C \cup R' \cup \{i\}\}$.

(b) If agent i manipulates agent j, then $m_{ji}(t+1) > 0$ and thus, since $C \ni j$ is minimal closed at period t, there exists a path at $t+1$ from l to i for all $l \in C$. Hence, by assumption there exists a path from agent j to k, but not vice versa since $C' \ni k$ is minimal closed. Thus, $R(t+1) = R(t) \cup C$, which finishes the proof.

Proof of Proposition 9

Suppose without loss of generality that $\mathcal{C}(t) = \{\mathcal{N}\}$. First, note that aperiodicity is preserved since manipulation can only increase the number of simple cycles. We can write

$$M(t+1) = M(t) + e_j z(t)',$$

where e_j is the j-th unit vector, and

$$\begin{aligned} z_k(t) &= \begin{cases} (m_{ji}(t) + \alpha) / (1+\alpha) - m_{ji}(t) & \text{if } k = i \\ (m_{jk}(t)) / (1+\alpha) - m_{jk}(t) & \text{if } k \neq i \end{cases} \\ &= \begin{cases} \alpha(1 - m_{ji}(t)) / (1+\alpha) & \text{if } k = i \\ -\alpha m_{jk}(t) / (1+\alpha) & \text{if } k \neq i \end{cases}. \end{aligned}$$

From Hunter (2005), we get

$$\begin{aligned} &\pi_k(\mathcal{N}; t+1) - \pi_k(\mathcal{N}; t) = -\pi_k(\mathcal{N}; t)\pi_j(\mathcal{N}; t+1) \sum_{l \neq k} z_l(t) r_{lk}(t) \\ &= \begin{cases} \alpha / (1+\alpha)\, \pi_i(\mathcal{N}; t)\pi_j(\mathcal{N}; t+1) \sum_{l \neq i} m_{jl}(t) r_{li}(t) & \text{if } k = i \\ \alpha / (1+\alpha)\, \pi_k(\mathcal{N}; t)\pi_j(\mathcal{N}; t+1) \left(\sum_{l \neq k} m_{jl}(t) r_{lk}(t) - r_{ik}(t) \right) & \text{if } k \neq i \end{cases}, \end{aligned}$$

which finishes the proof.

Proof of Corollary 7

We know that $\pi_k(C; t), \pi_k(C; t+1) > 0$ for all $k \in C$. Note that if i manipulates j, i.e., $\alpha > 0$, then it must be that $m_{ji}(t) < 1$ since otherwise $[M(t)](i; j, \alpha) = [M(t)](i; j, 0)$ and thus the agent would not have exerted effort. Thus, by Remark 3, $\sum_{l \in C \backslash \{i\}} m_{jl}(t) r_{li}(t) > 0$ and hence $\pi_i(\mathcal{N}; t+1) > \pi_i(\mathcal{N}; t)$, which proves part (i). Part (ii) is obvious. Part (iii) follows since $m_{jk}(t) = 1$ implies $\sum_{l \in C \backslash \{k\}} m_{jl}(t) r_{lk}(t) = 0$, which finishes the proof.

Proof of Lemma 2

By Proposition 8, we know that $\mathcal{C}(t) = \{\mathcal{N}\}$ for all $t \geq 0$, and furthermore, also aperiodicity is preserved. First, we show that the opinions converge to a consensus $x(\infty)$. Therefore, suppose to the contrary that the opinions (with positive probability) do not converge. This implies that there exists a periodic trust matrix $M^* \in \mathbb{R}^{n\times n}$ such that for some sequence of agents $\{i^*(t)\}_{t\geq 0}$ chosen to manipulate, $M(t) \to M^*$ for $t \to \infty$. Denote the decision of $i^*(t)$ at period t by $(j^*(t), \alpha^*(t))$. Notice that since $M(t)$ is aperiodic for all $t \geq 0$, i.e., $M(t) \neq M^*$ for all $t \geq 0$, this is only possible if there are infinitely many manipulations. (5)

Denoting by $x^*(t)$ the opinions and by $M^*(t)$ the trust matrix at period t in the above case, we get

$$\begin{aligned}
&\big\|[x^*(t)]\big(i^*(t); j^*(t), \alpha^*(t)\big) - [x^*(t)]\big(i^*(t); j^*(t), 0\big)\big\| \\
=&\big\|[M^*(t)]\big(i^*(t); j^*(t), \alpha^*(t)\big)x^*(t) - M^*(t)x^*(t)\big\| \\
&\to 0 \text{ for } t \to \infty,
\end{aligned}$$

and thus, by assumption,

$$\begin{aligned}
&v_{i^*}\big([M^*(t)]\big(i^*(t); j^*(t), \alpha^*(t)\big), x^*(t)\big) - v_{i^*}\big([M^*(t)]\big(i^*(t); j^*(t), 0\big), x^*(t)\big) \\
\to 0 < \underline{c} \leq c_{i^*}&(j^*(t), \alpha^*(t)) \text{ for } t \to \infty,
\end{aligned}$$

which is a contradiction to (5). Having established the convergence of opinions, it follows directly that $\|[x(t)](i; j, \alpha) - [x(t)](i; j, 0)\| \to 0$ for $t \to \infty$, any i selected at t and her choice (j, α). Hence, by assumption, $v_i\big([M(t)](i; j, \alpha), x(t)\big) - v_i\big([M(t)](i; j, 0), x(t)\big) \to 0 < \underline{c} \leq c_i(j, \alpha)$ for $t \to \infty$, any i selected at t and her choice (j, α), which shows that there exits an almost surely finite stopping time τ such that for all $t \geq \tau$, $E(t) = (i; \cdot, 0)$ for any i chosen to manipulate at t.

Furthermore, since $M(\tau)$ is aperiodic and no more manipulation takes place, agents reach a (random) consensus that can be written as

$$\begin{aligned}
x(\infty) = \pi(\mathcal{N}; \tau)'x(\tau) &= \pi(\mathcal{N}; \tau)'M(\tau)x(\tau - 1) \\
&= \pi(\mathcal{N}; \tau)'M(\tau - 1)M(\tau - 2)\cdots M(1)x(0) \\
&= \pi(\mathcal{N}; \tau)'\overline{M}(\tau - 1)x(0),
\end{aligned}$$

where the second equality follows from the fact that $\pi(\mathcal{N}; \tau)$ is a left eigenvector of $M(\tau)$ corresponding to eigenvalue 1, which finishes the proof.

Proof of Proposition 10

Suppose that the sequence $(\tau_k)_{k=1}^{\infty}$ of stopping times denotes the periods where the trust structure changes, i.e., at $t = \tau_k$ the trust structure changes the k-th time.

Notice that $\tau_k = +\infty$ if the k-th change never happens. By Proposition 8, it follows that when $\tau_k < +\infty$, either

(a) $1 \leq |\mathcal{C}(\tau_k + 1)| < |\mathcal{C}(\tau_k)|$ and $|R(\tau_k + 1)| > |R(\tau_k)|$, or

(b) $|\mathcal{C}(\tau_k + 1)| = |\mathcal{C}(\tau_k)|$ and $0 \leq |R(\tau_k + 1)| < |R(\tau_k)|$

holds. This implies that the maximal number of changes in the trust structure is finite, i.e., there exists $K < +\infty$ such that there are at most K changes in the structure and thus, almost surely $\tau_{K+1} = +\infty$. Hence, $\tau = \max\{\tau_k + 1 \mid \tau_k < +\infty\} < +\infty$, where $\tau_0 \equiv 0$, is the desired almost surely finite stopping time, which finishes part (i). Part (ii) follows from Lemma 2. The restriction to C of the matrices $M(t)$ in the computation of the consensus belief is due to the fact that $M(t)|_C$ is a stochastic matrix for all $t \geq 0$ since C is minimal closed at $t = \widehat{\tau}$, which finishes the proof.

Proof of Lemma 3

Suppose that $i \in S$. Since $\sum_{k\in S} m_{jk}(t) - \sum_{k\notin S} m_{jk}(t) \leq (<)1$, it follows that

$$\begin{aligned}
&\sum_{k\in S} m_{jk}(t) - \sum_{k\notin S} m_{jk}(t) \\
\leq (<) &\left(\sum_{k\in S} m_{jk}(t) - \sum_{k\notin S} m_{jk}(t)\right)/(1+\alpha) + \alpha/(1+\alpha) \\
= &\left(\sum_{k\in S\setminus\{i\}} m_{jk}(t) - \sum_{k\notin S} m_{jk}(t)\right)/(1+\alpha) + (m_{ji}(t)+\alpha)/(1+\alpha) \\
= &\sum_{k\in S} m_{jk}(t+1) - \sum_{k\notin S} m_{jk}(t+1)
\end{aligned}$$

and hence $\mathrm{Hom}(S; t+1) \geq (>)\mathrm{Hom}(S; t)$, which finishes part (i). Part (ii) is analogous, which finishes the proof.

Proof of Lemma 4

We can write

$$\begin{aligned}
\overline{\pi}_k(\mathcal{N}; t+1) &= \sum_{l=1}^{n} \overline{m}_{lk}(t)\pi_l(\mathcal{N}; t+1) \\
&= \sum_{l=1}^{n} \overline{m}_{lk}(t)\big(\pi_l(\mathcal{N}; t+1) - \pi_l(\mathcal{N}; t)\big) + \sum_{l=1}^{n} \overline{m}_{lk}(t)\pi_l(\mathcal{N}; t)
\end{aligned}$$

$$= \sum_{l=1}^{n} \overline{m}_{lk}(t)\big(\pi_l(\mathcal{N};t+1) - \pi_l(\mathcal{N};t)\big) + \underbrace{\sum_{l=1}^{n} \overline{m}_{lk}(t-1)\pi_l(\mathcal{N};t)}_{=\overline{\pi}_k(\mathcal{N};t)},$$

where the last equality follows since $\pi(\mathcal{N};t)$ is a left eigenvector of $M(t)$, which finishes the proof.

Proof of Proposition 11

Let $N_* \subseteq \mathcal{N}$ denote the set of agents that undersell their information at period t. Then, the agents in $N^* = \mathcal{N}\backslash N_*$ oversell their information and additionally, $N_*, N^* \neq \emptyset$. By Proposition 9, we have $\pi_k(\mathcal{N};t+1) - \pi_k(\mathcal{N};t) \to 0$ for $\alpha \to 0$ and all $k \in \mathcal{N}$ and thus by Lemma 4 we have

$$\overline{\pi}_k(\mathcal{N};t+1) - \overline{\pi}_k(\mathcal{N};t) \to 0 \text{ for } \alpha \to 0 \text{ and all } k \in \mathcal{N}. \tag{6}$$

Let $k \in N_*$, then by (ii) and Lemma 4, $\overline{\pi}_k(\mathcal{N};t+1) \geq \overline{\pi}_k(\mathcal{N};t)$. Hence, by (6), there exists $\overline{\alpha}(k) > 0$ such that

$$1/n \geq \overline{\pi}_k(\mathcal{N};t+1) \geq \overline{\pi}_k(\mathcal{N};t) \text{ for all } 0 < \alpha \leq \overline{\alpha}(k).$$

Analogously, for $k \in N^*$, there exists $\overline{\alpha}(k) > 0$ such that

$$1/n \leq \overline{\pi}_k(\mathcal{N};t+1) < \overline{\pi}_k(\mathcal{N};t) \text{ for all } 0 < \alpha \leq \overline{\alpha}(k).$$

Therefore, setting $\overline{\alpha} = \min_{k\in\mathcal{N}} \overline{\alpha}(k)$, we get for $0 < \alpha \leq \overline{\alpha}$

$$\begin{aligned}
\|\overline{\pi}(\mathcal{N};t) - (1/n)\cdot\mathbb{I}\|_2^2 &= \sum_{k\in\mathcal{N}} |\overline{\pi}_k(\mathcal{N};t) - 1/n|^2 \\
&= \sum_{k\in N_*} \underbrace{|\overline{\pi}_k(\mathcal{N};t) - 1/n|^2}_{\geq |\overline{\pi}_k(\mathcal{N};t+1) - 1/n|^2} + \sum_{k\in N^*} \underbrace{|\overline{\pi}_k(\mathcal{N};t) - 1/n|^2}_{> |\overline{\pi}_k(\mathcal{N};t+1) - 1/n|^2} \\
&> \sum_{k\in\mathcal{N}} |\overline{\pi}_k(\mathcal{N};t+1) - 1/n|^2 \\
&= \|\overline{\pi}(\mathcal{N};t+1) - (1/n)\cdot\mathbb{I}\|_2^2,
\end{aligned}$$

which finishes the proof.

Chapter 4

Strategic Communication in Social Networks

4.1 Introduction

Individuals form their beliefs and opinions on various economic, political and social issues based on information they receive from their social environment. This may include friends, neighbors and coworkers as well as political actors and news sources, among others. Typically, all these individuals have widely diverging interests, views and tastes, as can be seen in daily political discussions or in all kinds of bargaining situations. In election campaigns, politicians have incentives to argue solutions or proposals that differ from their beliefs. In budget allocation problems, the recipients of capital, e.g., ministries, local governments or departments of companies or universities, have incentives to overstate their capital requirement, while the other side is concerned with efficiency. Another example are court trials, where the accused has clearly incentives to misreport the events in question. And in marketing, firms might overstate the product quality to attract costumers.

When interests are conflicting, individuals will find it more advantageous not to reveal their true belief for strategic reasons. However, in the literature on communication in social networks, it is usually assumed that agents report their beliefs truthfully, see, e.g., DeGroot (1974); Golub and Jackson (2010); DeMarzo et al. (2003); Acemoglu et al. (2010); Förster et al. (2013). DeMarzo et al. (2003) state that this assumption is for simplicity, but that "[n]onetheless, in many persuasive settings, (e.g., political campaigns and court trials) agents clearly do have incentives to strategically misreport their beliefs."

The terms *belief* and *opinion* are usually employed as synonyms in the literature.

In this paper, we disentangle these two terms by introducing conflicting interests. The *belief* of an individual about some issue of common interest will be what she holds to be true given her information about the issue. On the other hand, her *opinion* (or *biased belief*) will be what is ought to be the answer to the issue given her bias.[70] We assume that when two individuals communicate, the receiver of information would like to get to know the belief of the sender about the issue as precisely as possible in order to refine her own belief, while the sender wants to spread his opinion, i.e., he would like the receiver to update her opinion with his opinion.

To illustrate this approach, consider an international meeting of politicians, e.g., the United Nations climate change conferences. The common issue of the decision-makers at these meetings is to find and to agree on the measures or actions to take in order to limit global warming. Each decision-maker holds a belief about which measures are to be taken by the global community to achieve this goal. However, the opinion they (intend to) support (communicate) in front of the other decision-makers often differs from this belief due to strategic reasons that depend on the local environment within their country. These reasons include local costs of adaption of the measures, the risk profile of the country, and the local public opinion. In other words, the opinion that a decision-maker intends to support is the ideal measure or action from her point of view.[71] During these meetings, politicians interact repeatedly with each other. When receiving information, they would like to do so as precisely as possible since the ideal action for each country depends on the fundamentals of global warming, while they intend to communicate their opinion when sending information in order to reach an outcome close to the ideal measure for their country.

An important question for society is how the presence of these conflicts influences information aggregation, long-run beliefs and opinions in society. We develop a framework of belief (opinion) dynamics where individuals with conflicting interests communicate strategically in a social network and update their beliefs with the obtained information.

More precisely, we consider a society represented by a social network of n agents.

[70] In this sense, her opinion is a personal judgement about the issue for strategic reasons or taste considerations.

[71] The 2009 United Nations climate change conference that took place in Copenhagen, Denmark, led to a political agreement on the goal of limiting global warming to no more than two degrees Celsius over the pre-industrial average. However, views on the measures to take remained widely diverging depending on local environments and therefore prevented a full-fledged legal agreement, see Bodansky (2010).

At time $t \geq 0$, each agent holds a belief $x_i(t) \in [0,1]$ about some common issue.[72] Furthermore, each agent has a bias $b_i \in \mathbb{R}$ that is common knowledge and that determines her opinion (biased belief) $x_i(t) + b_i$.[73] Each agent starts with an initial belief $x_i(0) \in [0,1]$ and meets (communicates with) agents in her social neighborhood according to a Poisson process in continuous time that is independent of the other agents.[74]

When an agent is selected by her associated Poisson process, she receives information from one of her neighbors (called the sender of information) according to a stochastic process that forms her social network.[75] The sender wants to spread his opinion, while the receiver wants to infer his belief in order to update her own belief. In equilibrium, this conflict of interest leads to noisy communication à la Crawford and Sobel (1982, henceforth CS): the sender sends one of finite messages that contains information about his belief, which is then interpreted by the receiver. In optimal equilibrium, communication is as informative as possible given the conflict of interest, i.e., the sender uses as many messages as possible and discriminates as finely as possible between different beliefs.[76]

The receiver updates her belief by taking the average of the interpretation of the sent message and her pre-meeting belief. Although simple, this updating rule reflects the idea that agents fail to adjust for repetitions and dependencies in information they hear several times due to the complexity of social networks, as argued by DeMarzo et al. (2003).[77]

Our framework induces a belief dynamics process as well as an opinion dynamics process. As a first observation, we note that we can concentrate our analysis on the belief dynamics process since both processes have the same convergence properties. We say that an agent's belief *fluctuates* on an interval if her belief will never leave

[72] We refer to DeMarzo et al. (2003) for a discussion about the representation of beliefs by a unidimensional structure.

[73] Notice that our notion is consistent with the literature in the sense that the terms belief and opinion coincide in absence of a bias.

[74] See Acemoglu et al. (2010, 2013), who use this timing in related models.

[75] Note that we model communication as directed. We want to allow for asymmetric communication since, e.g., an agent might obtain a lot of information from another agent, but this might not be the case vice versa. We can think of journalists whose information reach a large audience, who themselves only receive information from few people, though.

[76] Note that CS argue that the optimal equilibrium is particularly plausible in a situation like ours, where communication is repeated.

[77] Note that this updating rule has another appealing interpretation: if the initial beliefs were drawn independently from a normal distribution with equal mean and equal variance and if there was no conflict of interest, then this updating rule would be optimal. In view of this, we should think about the conflicts of interest as being rather small.

the interval and if this does not hold for any subinterval. In other words, the belief "travels" the whole interval, but not beyond.

In our main result, we show that for any initial beliefs, the belief dynamics process converges to a set of intervals that is *minimal mutually confirming*. Given each agent's belief lies in her corresponding interval, these intervals are the convex combinations of the interpretations the agents use when communicating. Furthermore, we show that the belief of an agent eventually fluctuates on her corresponding interval whenever the interval is proper, i.e., whenever it contains infinitely many elements (beliefs). As a consequence, the belief dynamics has a steady state if and only if there exists a minimal mutually confirming set such that all its intervals are degenerate, i.e., contain only a single point. We illustrate our results by several examples. Furthermore, we note that outcomes with a steady state must be constructed explicitly by choosing specific biases and configurations of the social network: as long as conflicts are small and some agents communicate with several different agents, outcomes with a steady state are non-generic.

The introduction of conflict of interest leads not only to persistent disagreement among the agents, but also to fluctuating beliefs and opinions, a phenomenon that is frequently observed in social sciences, see, e.g., Kramer (1971), who documents large swings in US voting behavior within short periods, and works in social and political psychology that study how political parties and other organizations influence political beliefs, e.g., Cohen (2003); Zaller (1992). At the same time, our result is surprising in view of the literature on communication in social networks: in most models, a strongly connected network leads to mutual consensus among the agents in the long-run. To this respect, Acemoglu et al. (2013) is the closest to our work, where the authors introduce stubborn agents that never change their belief, which leads to fluctuating beliefs when the other agents update regularly from different stubborn agents.

There exists a large literature on communication in social networks, using both Bayesian and non-Bayesian updating rules.[78] Apart from the various works that assume truthful communication, Büchel et al. (2012) study a model where agents act strategically in the sense that their stated belief differs from their true belief depending on their preferences for conformity. Acemoglu et al. (2014) study a model of Bayesian learning where the agents' objective is to form beliefs (acquire informa-

[78]In Bayesian and observational learning models communication is typically assumed to be truthful and agents converge to a mutual consensus, e.g., Banerjee and Fudenberg (2004); Gale and Kariv (2003); Acemoglu et al. (2011). Another stream of literature studies how observable behaviors spread in a population, e.g., López-Pintado (2008, 2012); Jackson and Yariv (2007); Morris (2000).

tion) about an irreversible decision that each agent has to make, eventually. In this setting, agents might want to misreport their information in order to delay the decisions of other agents. The authors show that it is an equilibrium to report truthfully whenever truthful communication leads to asymptotic learning, i.e., the fraction of agents taking the right decision converges to 1 (in probability) as the society grows. They also show that in some situations, misreporting can lead to asymptotic learning while truthful communication would not. However, also these models lead to mutual consensus under the condition that the underlying social network is strongly connected.

Several authors have proposed models to explain non-convergence of beliefs, usually incorporating some kind of homophily that leads to segregated societies and polarized beliefs.[79] Axelrod (1997) proposed such a model in a discrete belief setting, and later on Hegselmann and Krause (2002) and Deffuant et al. (2000) studied the continuous case, see also Lorenz (2005); Blondel et al. (2009); Como and Fagnani (2011). Golub and Jackson (2012) argue that the presence of homophily can substantially slow down convergence and thus lead to a high persistence of disagreement. While being able to explain persistent disagreement, these models fail to explain belief fluctuations in society.

Furthermore, our work is related to contributions on cheap-talk games. Hagenbach and Koessler (2010), Galeotti et al. (2013) and Ambrus and Takahashi (2008) extend the framework of CS to a multi-player (-sender) environment, but maintain the one-shot nature of the model.

The paper is organized as follows. In Section 4.2 we introduce the model and notation. Section 4.3 concerns the equilibrium in the communication stage. In Section 4.4 we look at the long-run belief dynamics. In Section 4.5 we conclude and discuss briefly some of our model choices. The proofs are presented in Appendix 4.A.

4.2 Model and Notation

We consider a set $\mathcal{N} = \{1, 2, \dots, n\}$, with $n \geq 2$, of agents who repeatedly communicate with their neighbors in a social network. At time $t \geq 0$, each agent $i \in \mathcal{N}$ holds a *belief* $x_i(t) \in [0, 1]$ about some common issue of interest. Furthermore, agent i has a *bias* $b_i \in \mathbb{R}$ that is common knowledge, i.e., her *biased belief* $x_i(t) + b_i$ is the

[79] An exception being Friedkin and Johnsen (1990), who study a variation of the model by DeGroot (1974) where agents can adhere to their initial beliefs to some degree. This leads as well to persistent disagreement among the agents.

ideal response to the issue from her point of view given her belief $x_i(t)$. We refer to $x_i(t) + b_i$ as her *opinion* at time t about the issue given her belief $x_i(t)$.

The *social network* is given by a stochastic matrix $P = (p_{ij})_{i,j\in\mathcal{N}}$, i.e., $p_{ij} \geq 0$ for all $i, j \in \mathcal{N}$ and $\sum_{j\in\mathcal{N}} p_{ij} = 1$ for all $i \in \mathcal{N}$. For agent i, p_{ij} is the probability to meet agent j, and $\mathcal{N}_i = \{j \in \mathcal{N} \mid p_{ij} > 0\}$ denotes i's *neighborhood*. Let $(\mathcal{N}, g)$ denote the directed graph where $g = \{(i, j) \mid p_{ij} > 0\}$ is the set of directed links induced by meeting probabilities $p_{ij} > 0$. Throughout the paper we will make the following assumption.

Assumption 1. (i) (Self-communication) Agents do not communicate with themselves, i.e., $p_{ii} = 0$ for all $i \in \mathcal{N}$.

(ii) (Connectivity) The graph $(\mathcal{N}, g)$ is strongly connected, i.e., for all $i, j \in \mathcal{N}$ there exists a directed path connecting i to j with links in g.

The first part states that "self-communication" is not possible. We make this assumption for simplicity, but it could be included as a possibility to abstain from communication. The second part guarantees that every agent "communicates" indirectly with every other agent, possibly through several links. We make this assumption for several reasons. First, it seems to be natural as evidence suggests that our societies are indeed connected, see, e.g., Watts (2003). And second, it is known to be a necessary condition for convergence of beliefs to a consensus. We want to exclude that beliefs fail to converge because agents are not connected.

Each agent $i \in \mathcal{N}$ starts with an initial belief $x_i(0) \in [0, 1]$. Agents meet (communicate) and update their beliefs according to an asynchronous continuous-time model. Each agent is chosen to meet another agent at instances defined by a rate one Poisson process independent of the other agents. Therefore, over all agents, the meetings occur at time instances t_s, $s \geq 1$, according to a rate n Poisson process. Note that by convention, at most one meeting occurs at a given time $t \geq 0$. Hence, we can discretize time according to the agent meetings and refer to the interval $[t_s, t_{s+1})$ as the s-th time slot. There are on average n meetings per unit of absolute time, see Boyd et al. (2006) for a detailed relation between the meetings and absolute time. At time slot s, we represent the beliefs of the agents by the vector $x(s) = (x_1(s), x_2(s), \ldots, x_n(s))'$.[80]

If agent $i \in \mathcal{N}$ is chosen at time slot s, $s \geq 1$ (probability $1/n$), she meets agent $j \in \mathcal{N}$ with probability p_{ij} and communicates with him. We assume that agent i

[80] We denote the transpose of a vector x by x'.

updates her belief with information she receives from agent j.[81] Agent j sends a *message* (or *signal*) $m \in \mathcal{M} := \{m_1, m_2, \ldots, m_L\}$ containing information about his belief $x_j(s-1)$, where $L \in \mathbb{N}$ is very large but finite, and which is interpreted by i as an estimate $y^{ij}(m)$ of $x_j(s-1)$.[82] Agent i then updates her belief by taking the average of this interpretation and her pre-meeting belief:

$$x_i(s) = \frac{x_i(s-1) + y^{ij}(m)}{2}.$$

If not stated otherwise, agent i will denote the agent that updates her belief (the *receiver* of information), and agent j will denote the agent with whom she communicates (the *sender* of information). We write $g(s) = ij$ if link (i, j) is chosen at time slot s.

Next, we specify how communication between agents takes place. We adapt the framework of Jäger et al. (2011) to conflicting interests and repeated communication. Suppose that $g(s) = ij$; we make the following assumption about the objectives of the agents.

Assumption 2 (Objectives)**.** Agent i's objective is to infer agent j's belief $x_j(s-1)$, while agent j's objective is to spread his opinion $x_j(s-1) + b_j$.

Thus, agent i's ideal interpretation is $y^{ij}(m) = x_j(s-1)$. For agent j, notice first that we can write i's updated opinion as the average of the biased interpretation $y^{ij}(m) + b_i$ and her pre-meeting opinion:

$$x_i(s) + b_i = \frac{(x_i(s-1) + b_i) + (y^{ij}(m) + b_i)}{2},$$

i.e., i updates her opinion with the biased interpretation $y^{ij}(m) + b_i$. Hence, agent j's ideal interpretation is $y^{ij}(m) = x_j(s-1) + (b_j - b_i)$ since in this case i updates her opinion with j's opinion:

$$x_i(s) + b_i = \frac{(x_i(s-1) + b_i) + (x_j(s-1) + b_j)}{2}.$$

Note that in absence of conflict of interest ($b_i = b_j$) j's ideal interpretation is equal to his belief, i.e., ideal interpretations coincide.

Formally, the agents' preferences are given by

$$u_i(x_j(s-1), y^{ij}(m)) = h(|x_j(s-1) - y^{ij}(m)|)$$

[81] Note that agent j does not update his belief. Together with the directed social network, this assumption allows for asymmetric communication.

[82] We know from CS that assuming a (sufficiently) large but finite number of messages represents only a restriction in absence of conflict of interest. Since we focus on conflicting interests, we take this assumption for analytical convenience.

and

$$u_j(x_j(s-1), y^{ij}(m)) = h(|x_j(s-1) + (b_j - b_i) - y^{ij}(m)|),$$

where $h : \mathbb{R}_+ \to \mathbb{R}$ is a continuous, concave and strictly decreasing function. Agent j wants to send a message m such that $i's$ interpretation is as close as possible to his ideal interpretation $x_j(s-1) + (b_j - b_i)$, while i wants to choose an interpretation that is as close as possible to j's belief $x_j(s-1)$.[83] Thus, the belief dynamics is well-defined since agent i optimally chooses an interpretation in $[0,1]$ whatever message she receives. A simple example are quadratic preferences.

Example 18 (Quadratic preferences)**.**

$$u_i(x,y) = -(x-y)^2 \text{ and } u_j(x,y) = -(x + (b_j - b_i) - y)^2.$$

Let F be an atomless distribution on $[0,1]$ with strictly positive and continuous density $f : [0,1] \to \mathbb{R}_+$. We impose the following assumption on the beliefs of agents about the other agents' beliefs.

Assumption 3 (Distribution of beliefs)**.** Prior to each round of communication, agent i believes that j's belief is distributed according to F on $[0,1]$.

This assumption reflects the idea that each round of communication is independent. Since time is continuous and agents' meetings are independent, they do not know how many times others have updated their belief in a given period of time. And moreover, since agents need to coordinate on the distribution in equilibrium, at least after repeated communication, it seems plausible to keep it constant. Or, if we think of the initial beliefs as being drawn independently from a commonly known distribution F, then Assumption 3 means that agents believe that after updating their beliefs they are still distributed according to F.

We employ Bayesian Nash equilibrium and exclude the possibility of any prior commitment of the agents. In this signaling game, a strategy for the sender j is a measurable function

$$m^{ij} : [0,1] \to \mathcal{M}$$

that assigns a message to each possible belief and for the receiver i, it is a function

$$y^{ij} : \mathcal{M} \to [0,1]$$

that assigns an interpretation to each possible message. We refer to the interpretation of message m_l as $y_l = y^{ij}(m_l)$ and to the set of beliefs that induces m_l as

[83] We can also interpret $-u_k(x_j(s-1), y^{ij}(m))$ as the loss from communication, see Jäger et al. (2011).

$C_l = (m^{ij})^{-1}(m_l) = \{x \in [0,1] : m^{ij}(x) = m_l\}$ when there is no confusion. A Bayesian Nash equilibrium of the game consists of strategies (m^{ij}, y^{ij}) such that

(i) for each message $m_l \in \mathcal{M}$,

$$y_l \in \operatorname*{argmax}_{y\in[0,1]} \int_{C_l} u_i(x,y)F(dx), \text{ and}$$

(ii) for each belief $x \in [0,1]$,

$$m^{ij}(x) \in \operatorname*{argmax}_{m\in\mathcal{M}} u_j\big(x, y^{ij}(m)\big).$$

Condition (i) says that for each of the messages, agent i chooses an interpretation that maximizes her expected utility under Assumption 3, i.e., upon receiving message m_l she chooses an interpretation y_l that maximizes her expected utility conditional on j's belief being distributed according to F on $C_l = (m^{ij})^{-1}(m_l)$. Condition (ii) says that for each belief agent j chooses a message that maximizes his utility.

We assume without loss of generality that whenever two messages lead to the same interpretation, then agent j only sends the message with the lower index. We say that a message m_l is *induced* (*used*) in equilibrium if $C_l = (m^{ij})^{-1}(m_l)$ has positive measure, and otherwise we assume that $C_l = \emptyset$.[84] Thus, we can restrict our attention to the messages that are induced in equilibrium and their interpretations, which are distinct.[85] Throughout the paper we assume that Assumption 1, 2 and 3 hold.

4.3 Communication Stage

In this section we characterize, given $g(s) = ij$, how agent j communicates with agent i. First, notice that the ideal (optimal) interpretations, $x + (b_j - b_i) = \operatorname{argmax}_{y\in\mathbb{R}} u_j(x,y)$ and $x = \operatorname{argmax}_{y\in\mathbb{R}} u_i(x,y)$, are unique and strictly increasing in j's belief x. Furthermore, ideal interpretations differ under conflict of interest.

[84] Notice that there are equilibria where some C_l is a non-empty null set. But, since null sets play no role for the expected utility of agent i, we can change j's strategy on a null set without affecting i's strategy.

[85] Notice that we do not consider mixed strategies. Receiver i's best reply to a message induced in a (mixed) equilibrium is unique since $y \mapsto u_i(x,y)$ is strictly concave. For sender j, the restriction to pure strategies comes without loss of generality since his best reply is F–almost everywhere unique as we will see in the next section.

Thus, we know from the analysis in CS that the number of messages induced in equilibrium is bounded.

Suppose j uses messages $m \in \mathcal{M}|_{L(ij)} := \{m_1, m_2, \ldots, m_{L(ij)}\}$ in equilibrium that lead to distinct interpretations $(y_l)_{l=1}^{L(ij)}$. Then, given j holds belief $x_j(s-1)$, he sends a message that maximizes his utility, i.e.,

$$\begin{aligned} m^{ij}(x_j(s-1)) &\in \operatorname*{argmax}_{m \in \mathcal{M}|_{L(ij)}} u_j\big(x_j(s-1), y^{ij}(m)\big) \\ &= \operatorname*{argmax}_{m \in \mathcal{M}|_{L(ij)}} h(|x_j(s-1) + (b_j - b_i) - y^{ij}(m)|) \\ &= \operatorname*{argmin}_{m \in \mathcal{M}|_{L(ij)}} |x_j(s-1) + (b_j - b_i) - y^{ij}(m)|, \end{aligned}$$

where the last equality follows since h is strictly decreasing. Note that this choice is not uniquely defined if $x_j(s-1)+(b_j-b_i)$ has equal distance to two interpretations, but since the set of such beliefs forms a null set with respect to F, we can assume without loss of generality that j sends the message with the lowest index in this case.[86] Hence, we can identify j's strategy in equilibrium with a partition $(C_l)_{l=1}^{L(ij)}$ of $[0,1]$, where

$$C_l = (m^{ij})^{-1}(m_l) = \{x \in [0,1] : m^{ij}(x) = m_l\} = [c_{l-1}, c_l)$$

is such that $0 = c_0 < c_1 < \cdots < c_{L(ij)} = 1$. Note that c_l refers to the belief where j is indifferent between sending message m_l and m_{l+1}. So, in equilibrium he partitions the unit interval and only communicates the element of the partition his belief is from.

Upon receiving message m_l, i will choose an interpretation $y_l = y^{ij}(m_l)$ that maximizes her expected utility conditional on j's belief being distributed according to F on C_l, i.e.,

$$y_l = \operatorname*{argmax}_{y \in [0,1]} \int_{C_l} u_i(x,y) F(dx) = \operatorname*{argmax}_{y \in [0,1]} \int_{c_{l-1}}^{c_l} h(|x-y|) F(dx).$$

Note that y_l is the unique best Bayesian estimator of C_l (under Assumption 3). The number of messages induced in equilibrium is bounded under conflict of interest: we show that the distance between any two interpretations induced in equilibrium is larger than the distance $|b_j - b_i|$ between the ideal interpretations of the agents. Only the equilibrium with one message always exists: in this equilibrium, j's strategy is

[86] Null sets play no role for the expected utility of agent i and therefore changing j's strategy on a null set does not change i's strategy.

given by $C_1 = [0,1]$ and i uses the best Bayesian estimator (under Assumption 3) of the unit interval,

$$y_1 = \underset{y\in[0,1]}{\operatorname{argmax}} \int_0^1 h(|x-y|)F(dx).$$

We refer to the finite upper bound on the number of messages (or the "size" of the partition) induced in equilibrium by $L(ij)$. We call the equilibrium using $L(ij)$ messages *optimal equilibrium* since it is *most informative* in the sense that it uses the finest partition.[87] Furthermore, this equilibrium is *essentially unique* in the sense that all equilibria using $L(ij)$ messages induce F–a.s. (almost surely) the same partition. As the receiver's interpretation of a given partition element is unique, this implies that in all equilibria the relation between the sender's belief and the receiver's interpretation is a.s. the same, see CS. And, following their argumentation, we assume that agents coordinate on this equilibrium.[88]

In absence of conflicting interests, the same result holds since we only allow for a finite number of messages. Agents use the maximal number of messages $L(ij) = L$ in optimal equilibrium. Since we do not want to restrict the game under conflict of interest, we assume $L \geq \max\{L(ij) \mid b_i \neq b_j\}$.

The following proposition summarizes our findings.

Proposition 12. *Suppose that $g(s) = ij$.*

(i) If $b_i \neq b_j$, then there exists an upper bound $1 \leq L(ij) < 1/|b_j - b_i| + 1$ on the number of messages in equilibrium.

(ii) The game has an essentially unique optimal equilibrium (m^{ij}, y^{ij}) in which agent j uses $L(ij)$ (L if $b_i = b_j$) messages and his strategy is given by a partition $(C_l)_{l=1}^{L(ij)}$, where $C_l = (m^{ij})^{-1}(m_l) = [c_{l-1}, c_l)$ is such that $0 = c_0 < c_1 < \cdots < c_{L(ij)} = 1$ and

$$|c_l + (b_j - b_i) - y_l| = |c_l + (b_j - b_i) - y_{l+1}| \text{ for } l = 1, 2, \ldots, L(ij) - 1.$$

Furthermore, agent i's strategy is given by interpretations

$$y_l = y^{ij}(m_l) = \underset{y\in[0,1]}{\operatorname{argmax}} \int_{c_{l-1}}^{c_l} h(|x-y|)F(dx) \text{ for } l = 1, 2, \ldots, L(ij).$$

[87] Notice that it is ex-ante pareto-superior to all other equilibria, see CS.

[88] They argue that this equilibrium seems to be particularly plausible in situations where communication is repeated, that is, in our case.

All proofs can be found in Appendix 4.A. We denote the optimal equilibrium when $g(s) = ij$ by the triple $\mathcal{E}^{ij} = (L(ij), C^{ij}, Y^{ij})$, where $C^{ij} = (c_1, c_2, \ldots, c_{L(ij)-1})$ denotes j's strategy and $Y^{ij} = (y_1, y_2, \ldots, y_{L(ij)})$ denotes i's strategy.

A choice of the distribution F that is prominent in the sense that agents are likely to be able to coordinate on it is the uniform distribution. And surprisingly, this allows us to explicitly compute the equilibrium outcome.

Corollary 10. *Suppose that $g(s) = ij$ and that $F = U(0,1)$ is the uniform distribution.*

(i) If $b_i \neq b_j$, then there exists a finite upper bound

$$L(ij) = \max\{l \in \mathbb{N} \mid 1/(2l) > |(l-1)(b_j - b_i)|\}$$

on the number of messages in equilibrium.

(ii) The game has an essentially unique optimal equilibrium (m^{ij}, y^{ij}) in which agent j uses $L(ij)$ (L if $b_i = b_j$) messages and his strategy is given by a partition $(C_l)_{l=1}^{L(ij)}$, where $C_l = (m^{ij})^{-1}(m_l) = [c_{l-1}, c_l)$ is such that

$$c_l = l/L(ij) - 2l(L(ij) - l)(b_j - b_i).$$

Furthermore, agent i's strategy is given by interpretations

$$y_l = y^{ij}(m_l) = (2l-1)/(2L(ij)) - \big((2l-1)L(ij) - 2(l^2 - l) - 1\big)(b_j - b_i)$$

for $l = 1, 2, \ldots, L(ij)$.

The next example illustrates how such equilibria can look like.

Example 19. Consider $\mathcal{N} = \{1, 2\}$, the vector of biases $b = (0, 1/20)'$ and the uniform distribution $F = U(0,1)$. The first agent is not biased, while the second is biased to the right.

When $g(s) = 12$, then $L(12) = 3$ messages are induced in optimal equilibrium and strategies are $C^{12} = (4/30, 14/30)$ and $Y^{12} = (2/30, 9/30, 22/30)$. This means that if, for instance, agent 2's belief is below $c_1 = 4/30$, then he sends message m_1 and agent 1 interprets this as $y_1 = 2/30$. When $g(s) = 21$, then as well $L(21) = 3$ messages are induced in optimal equilibrium and strategies are $C^{21} = (16/30, 26/30)$ and $Y^{21} = (8/30, 21/30, 28/30)$. Both equilibria are depicted in Figure 4.1.

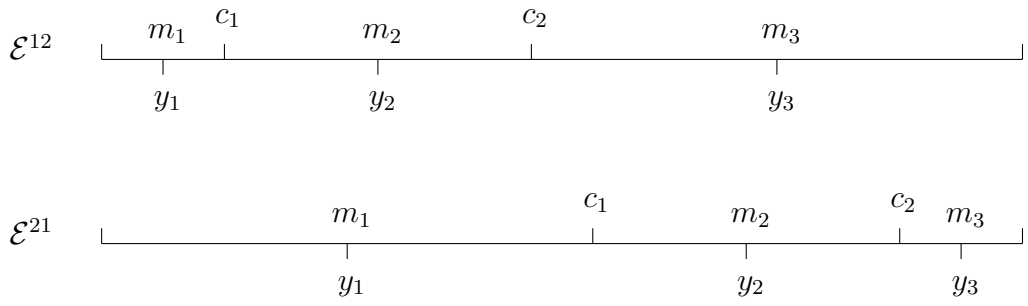


Figure 4.1: Optimal equilibria in Example 19.

4.4 Belief Dynamics

In this section we study the long-run behavior of the belief dynamics. At each time slot s, a pair of agents $g(s) = ij$ is selected according to the social network and communicates by employing the optimal equilibrium $\mathcal{E}^{ij}$. Agent i adopts the average of her pre-meeting belief and the equilibrium outcome of communication (her interpretation) as her updated belief. Hence, the belief dynamics $\{x(t)\}_{t\geq 0}$ defines a Markovian stochastic process. Note that we can define as well the opinion dynamics process $\{x(t) + b\}_{t\geq 0}$, where $b = (b_1, b_2, \ldots, b_n)'$ denotes the vector of biases.

Remark 5. The opinion dynamics $\{x(t) + b\}_{t\geq 0}$ is obtained by a translation of the state space of the belief dynamics $\{x(t)\}_{t\geq 0}$. Hence, both processes have the same properties in terms of convergence.

In the following, we will focus on the belief dynamics. The next example suggests that conflicting interests might prevent society from finding a consensus and instead lead to fluctuating beliefs.

Example 20. Consider $\mathcal{N} = \{1, 2, 3\}$, the vector of biases $b = (0, 1/25, -1/15)'$ and the uniform distribution $F = U(0, 1)$. Furthermore, all agents hold the same initial belief $x_i(0) = 1/2$ and the social network is given by

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix},$$

i.e., each possible pair of agents is chosen with probability 1/6 at a given time slot. This leads to the following equilibria in the communication stage:

- $\mathcal{E}^{12} = (4, (6/600, 108/600, 306/600), (3/600, 57/600, 207/600, 453/600))$,
- $\mathcal{E}^{13} = (3, (360/600, 560/600), (180/600, 460/600, 580/600))$,
- $\mathcal{E}^{21} = (4, (294/600, 492/600, 594/600), (147/600, 393/600, 543/600, 597/600))$,
- $\mathcal{E}^{23} = (2, (428/600), (214/600, 514/600))$,
- $\mathcal{E}^{31} = (3, (40/600, 240/600), (20/600, 140/600, 420/600))$,
- $\mathcal{E}^{32} = (2, (172/600), (86/600, 386/600))$.

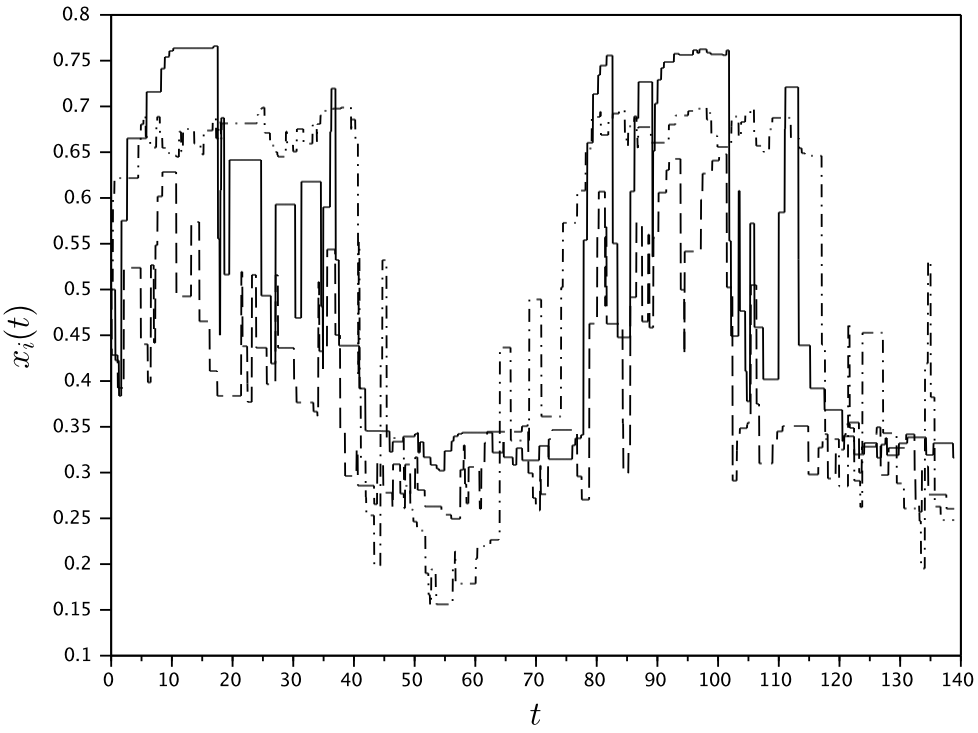


Figure 4.2: Long-run belief dynamics in Example 20. The solid line represents agent 1, the dashed line agent 2 and the dashed-dotted line agent 3.

The number of messages induced in equilibrium varies depending on the pair of agents selected to communicate. Agents 1 and 2 use four messages when communicating. The agents with the largest conflict of interest, 2 and 3, only use two messages in equilibrium, though. When looking at the long-run belief dynamics, we find that beliefs do not converge although agents started with identical beliefs.

Instead, the beliefs keep fluctuating forever. In particular, each belief fluctuates on some subinterval of $[0, 1]$. Agent 1's belief fluctuates on $[180/600, 460/600]$, agent 2's belief on $[147/600, 393/600]$, and agent 3's belief on $[86/600, 420/600]$. Figure 4.2 depicts one outcome of the long-run belief dynamics.

Note that the boundaries of the subintervals on which the beliefs fluctuate in the above example are related to the interpretations used by the agents when receiving information. In the following, we want to characterize the asymptotic behavior of the belief dynamics. First, we formalize what we mean by *fluctuation*. We say that an interval is *proper* if it contains infinitely many elements (beliefs).

Definition 13 (Fluctuation)**.** We say that the belief of an agent $i \in \mathcal{N}$ *fluctuates* on the closed and proper interval $\mathcal{I} \subseteq [0, 1]$ at time slot s if a.s. $x_i(s') \in \mathcal{I}$ for all $s' \geq s$, but for any closed subinterval $\mathcal{I}' \subsetneq \mathcal{I}$ this does not hold.

In other words, fluctuation on some interval means that the agent's belief never leaves the interval again, but still it "travels" the whole interval. Next, we define the concept of *mutually confirming intervals*. For $j \in \mathcal{N}_i$, let

$$Y^{ij}\big|_{\mathcal{I}_j} = \{y \in Y^{ij} \mid y = y^{ij}(m^{ij}(x)) \text{ for some } x \in \mathcal{I}_j\}$$

denote the restriction of Y^{ij} to the interpretations that correspond to messages sent when j's belief is in $\mathcal{I}_j$.

Definition 14 (Mutually confirming intervals)**.** We say that a set of intervals $\{\mathcal{I}_i\}_{i\in\mathcal{N}}$ is *mutually confirming* if, for all $i \in \mathcal{N}$,

$$\mathcal{I}_i = \operatorname{conv}\left(\cup_{j\in\mathcal{N}_i} Y^{ij}\big|_{\mathcal{I}_j}\right).$$

We say that a set of intervals $\{\mathcal{I}_i\}_{i\in\mathcal{N}}$ is *minimal mutually confirming* if there does not exist a mutually confirming set $\{\mathcal{I}'_i\}_{i\in\mathcal{N}}$ such that $\mathcal{I}'_i \subseteq \mathcal{I}_i$ for all $i \in \mathcal{N}$ and $\mathcal{I}'_i \subsetneq \mathcal{I}_i$ for at least one $i \in \mathcal{N}$.

Mutually confirming intervals are the convex combinations of the interpretations of the messages sent when communicating, given each agent's belief lies in her corresponding interval. The next theorem shows that the belief dynamics converges to a minimal mutually confirming set of intervals. Furthermore, we show that the belief of an agent eventually fluctuates on her corresponding interval whenever the interval is proper.

Theorem 2. *(i) For any vector of initial beliefs $x(0) \in [0, 1]^n$, the belief dynamics $\{x(t)\}_{t\geq 0}$ converges to a minimal mutually confirming set of intervals $\{\mathcal{I}_i\}_{i\in\mathcal{N}}$, and*

(ii) there exists an a.s. finite stopping time τ on the probability space induced by the belief dynamics process such that the belief of agent $i \in \mathcal{N}$ fluctuates on $\mathcal{I}_i$ at time slot $s = \tau$ if $\mathcal{I}_i$ is proper.

Theorem 2 implies that if all intervals of a minimal mutually confirming set are *degenerate*, i.e., contain only a single point, then the belief dynamics process has a steady state.

Corollary 11. *The belief dynamics $\{x(t)\}_{t\geq 0}$ has a steady state x^* if and only if there exists a minimal mutually confirming set of intervals $\{\mathcal{I}_i\}_{i\in\mathcal{N}}$ such that $\mathcal{I}_i$ is degenerate for all $i \in \mathcal{N}$. In this case, $x^* = (\mathcal{I}_i)_{i\in\mathcal{N}}$.*

When each agent communicates only with one other agent, there is a steady state for sure. The next example shows that this is also possible if some agent communicates with several agents.

Example 21. Consider $\mathcal{N} = \{1, 2, 3\}$, the vector of biases $b = (0, 37/600, -26/600)'$ and the uniform distribution $F = U(0, 1)$. Furthermore, all agents hold the same initial belief $x_i(0) = 1/2$ and the social network is given by

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

i.e., agent 1 is connected to all other agents, while these agents only listen to agent 1. This leads to the following equilibria in the communication stage:

- $\mathcal{E}^{12} = (3, (26/300, 126/300), (13/300, 76/300, 213/300))$,
- $\mathcal{E}^{13} = (3, (152/300, 252/300), (76/300, 202/300, 276/300))$,
- $\mathcal{E}^{21} = (3, (174/300, 274/300), (87/300, 224/300, 287/300))$,
- $\mathcal{E}^{31} = (3, (48/300, 148/300), (24/300, 98/300, 224/300))$.

All equilibria induce three messages in equilibrium. The vector of beliefs $x^* = (76/300, 87/300, 98/300)'$ is a steady state of the process. Note that since agent 1 communicates with two different agents, it is key that the interpretation $y = 76/300$ is part of both equilibria when she is selected to update her belief. Figure 4.3 depicts an outcome where beliefs converge to this steady state.

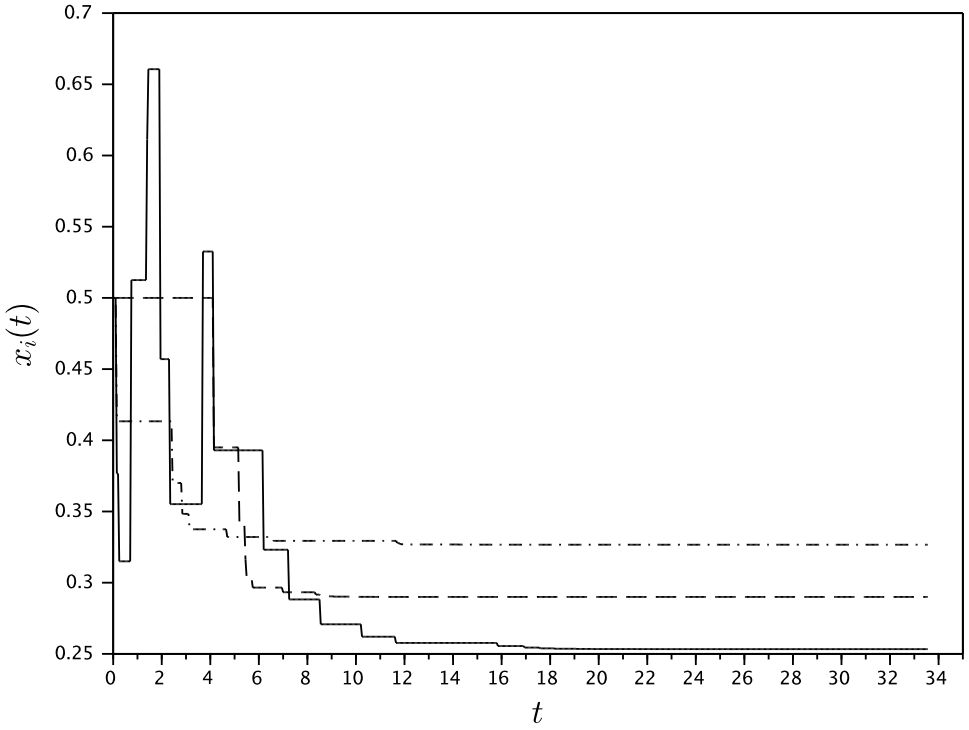


Figure 4.3: Long-run belief dynamics in Example 21. The solid line represents agent 1, the dashed line agent 2 and the dashed-dotted line agent 3.

The above example shows that the belief dynamics might converge in certain cases. However, such an outcome must be constructed explicitly by choosing specific biases and network configurations. The network needs to be sparse since each time an agent communicates with several agents, we need to find biases such that some interpretation is part of all equilibria. And additionally, we must ensure that these common interpretations are mutually confirming. In particular, a steady state is not stable with respect to the biases.

Remark 6. If conflicts of interest are small enough such that in optimal equilibrium agents send more than one message (no "babbling") and some agents communicate with at least two different agents, then outcomes with a steady state are non-generic.

4.5 Discussion and Conclusion

We introduce conflicting interests into a model of long-run belief dynamics. Our analysis is motivated by numerous examples such as political campaigns or court trials, where conflicts between different individuals are clearly present. We consider a society represented by a strongly connected network, agents meet (communicate) pairwise with their neighbors and exchange information strategically.

We disentangle the terms *belief* and *opinion*, previously employed as synonyms in the literature: the belief of an individual about some issue of common interest is what she holds to be true given her information, and her opinion is what is ought to be the answer to the issue given her bias. We assume that when two individuals communicate, the receiver of information would like to learn the true belief of the sender about the issue as precisely as possible in order to refine her own belief, while the sender wants to spread his opinion, i.e., he would like the receiver to update her opinion with his opinion.

This conflict of interest prevents the agents from revealing their true belief in equilibrium, and instead it leads to noisy communication à la CS: the sender sends one of finite messages that contains information about his belief, which is then interpreted by the receiver. In optimal equilibrium, communication is as informative as possible given the conflict of interest, i.e., the sender uses as many messages as possible. The receiver updates her belief by taking the average of the interpretation of the sent message and her pre-meeting belief.

In our main result, we show that the belief dynamics process converges to a set of intervals that is *minimal mutually confirming*. Given each agent's belief lies in her corresponding interval, these intervals are the convex combinations of the interpretations the agents use when communicating. Furthermore, we show that the belief of an agent eventually *fluctuates* on her corresponding interval whenever the interval is proper. As a consequence, the belief dynamics has a steady state if and only if there exists a minimal mutually confirming set such that all its intervals are degenerate.

We remark that outcomes with a steady state are non-generic as long as conflicts of interest are small and some agents communicate with several different agents. Hence, we can conclude that the introduction of conflict of interest leads not only to persistent disagreement among the agents, but also to fluctuating beliefs and opinions.

Though frequently observed in social science, the phenomenon of fluctuation is barely studied in the literature on communication in social networks, the only

exception being Acemoglu et al. (2013). While their result is very close to ours, they achieve it with a different approach. Instead of conflicting interests, they introduce stubborn agents that never change their belief into a model of belief dynamics. This also leads to fluctuating beliefs when the other agents update regularly from different stubborn agents. In our model, a stubborn agent would be an agent that only communicates with herself.

Finally, we would like to comment briefly on some of our model choices. Our paper presents a first attempt to enrich a model of belief dynamics with a framework of communication that incorporates conflicting interests. We use a simple averaging rule to update beliefs. Such a mechanism presents a natural starting point and has been intensively argued in the literature. However, it would be interesting to see whether (partially) Bayesian updating rules would generate similar results.

Due to the repeated nature of communication, the sender does not receive a signal prior to communication that is drawn from a commonly known probability distribution as in classical cheap-talk games. Instead she holds a belief resulting from her initial belief and previous communication. Therefore, apart from the number of messages, agents also need to coordinate on the distribution of the sender's beliefs. Viewing this distribution as the outcome of a coordination process suggests that it should be kept constant over time. And furthermore, if we think of the initial beliefs as being drawn independently from a probability distribution, then it seems likely that agents can coordinate on this distribution. However, when agents are (partially) Bayesian, it might be desirable to also allow them to update their beliefs about the other agents' beliefs. We leave these issues for future work.

4.A Appendix

Proof of Proposition 12

Suppose that agent j uses L' messages in equilibrium. We know that agent j's strategy is given by a partition $(C_l)_{l=1}^{L'}$, where $C_l = (m^{ij})^{-1}(m_l) = [c_{l-1}, c_l)$ is such that $0 = c_0 < c_1 < \cdots < c_{L'} = 1$ and

$$|c_l + (b_j - b_i) - y_l| = |c_l + (b_j - b_i) - y_{l+1}| \text{ for } l = 1, 2, \ldots, L' - 1.$$

And furthermore, agent i's strategy is given by interpretations

$$y_l = y^{ij}(m_l) = \underset{y \in [0,1]}{\operatorname{argmax}} \int_{c_{l-1}}^{c_l} h(|x - y|) F(dx) \text{ for } l = 1, 2, \ldots, L'.$$

Next, we show that there is an upper bound on the number of messages induced in equilibrium under conflict of interest, i.e., $b_i \neq b_j$. Let $y_l < y_{l+1}$ be two interpretations induced in equilibrium. Then, c_l satisfies $|c_l+(b_j-b_i)-y_l| = |c_l+(b_j-b_i)-y_{l+1}|$. Hence, since h is strictly decreasing,

$$y_l < \operatorname*{argmax}_{y\in[0,1]} u_j(c_l, y) = \operatorname*{argmax}_{y\in[0,1]} h(|c_l + (b_j - b_i) - y|) = c_l + (b_j - b_i) < y_{l+1}, \quad (7)$$

i.e., at c_l, the point where j is indifferent between interpretations y_l and y_{l+1}, he would prefer an interpretation to be implemented that lies strictly between these two interpretations. On the other hand, the same is true for i given she knew that j's opinion is exactly c_l:

$$\begin{aligned} y_l = \operatorname*{argmax}_{y\in[0,1]} \int_{c_{l-1}}^{c_l} h(|x-y|)F(dx) &< \operatorname*{argmax}_{y\in[0,1]} h(|c_l - y|) \\ &= c_l \\ &< \operatorname*{argmax}_{y\in[0,1]} \int_{c_l}^{c_{l+1}} h(|x-y|)F(dx) \\ &= y_{l+1}. \end{aligned} \quad (8)$$

It follows from (7) and (8) that

$$|y_l - y_{l+1}| > |c_l + (b_j - b_i) - c_l)| = |b_j - b_i|$$

and hence, there exists a maximal number of messages $\bar{L} < 1/|b_j - b_i| + 1$ that is feasible in equilibrium.

Note that if $b_i = b_j$, then there is no bound due to the biases and thus, the number of messages in equilibrium is bounded by $\bar{L} = L$. We know already that there always exists an equilibrium with only one message.

Altogether, agent j uses $1 \leq L(ij) = \bar{L}$ messages in optimal (i.e., most informative) equilibrium and moreover, this equilibrium is essentially unique since i's interpretations are unique and j's strategy is F–a.s. unique and thus all equilibria induce F–a.s. the same partition, which finishes the proof.

Proof of Corollary 10

Consider the optimal equilibrium $\mathcal{E}^{ij} = (L(ij), C^{ij}, Y^{ij})$, $C^{ij} = (c_1, c_2, \ldots, c_{L(ij)-1})$ and $Y^{ij} = (y_1, y_2, \ldots, y_{L(ij)})$. Since $F = U(0,1)$ and h is strictly decreasing, we get

$$y_l = \operatorname*{argmax}_{y\in[0,1]} \int_{c_{l-1}}^{c_l} h(|x-y|)dx = \operatorname*{argmin}_{y\in[0,1]} \int_{c_{l-1}}^{c_l} |x-y|dx = (c_l + c_{l-1})/2.$$

Thus, j's strategy satisfies

$$\begin{aligned}
&|c_l + (b_j - b_i) - y_l| = |c_l + (b_j - b_i) - y_{l+1}| \\
\Leftrightarrow &|c_l + (b_j - b_i) - (c_l + c_{l-1})/2| = |c_l + (b_j - b_i) - (c_{l+1} + c_l)/2| \\
\Leftrightarrow &|(c_l - c_{l-1})/2 + (b_j - b_i)| = |(c_l - c_{l+1})/2 + (b_j - b_i)|.
\end{aligned}$$

By monotonicity of the c_l, this yields

$$c_{l+1} = 2c_l - c_{l-1} + 4(b_j - b_i) \text{ for } l = 1, 2, \ldots, L(ij) - 1.$$

And by the boundary condition $c_0 = 0$, it follows that

$$c_l = lc_1 + 2l(l-1)(b_j - b_i) \text{ for } l = 1, 2, \ldots, L(ij).$$

The other boundary condition, $c_{L(ij)} = 1$, implies that $c_1 = 1/L(ij) - 2(L(ij) - 1)(b_j - b_i)$ and hence,

$$\begin{aligned}
c_l &= lc_1 + 2l(l-1)(b_j - b_i) \\
&= l\big(1/L(ij) - 2(L(ij) - 1)(b_j - b_i)\big) + 2l(l-1)(b_j - b_i) \\
&= l/L(ij) - 2l(L(ij) - l)(b_j - b_i). \qquad (9)
\end{aligned}$$

Hence,

$$\begin{aligned}
y_l &= (c_l + c_{l-1})/2 \\
&= (2l-1)/(2L(ij)) - l(L(ij) - l)(b_j - b_i) - (l-1)(L(ij) - l + 1)(b_j - b_i) \\
&= (2l-1)/(2L(ij)) - \big(l(2L(ij) - 2l + 1) - (L(ij) - l + 1)\big)(b_j - b_i) \\
&= (2l-1)/(2L(ij)) - \big((2l-1)L(ij) - 2(l^2 - l) - 1\big)(b_j - b_i).
\end{aligned}$$

Suppose that $b_i \neq b_j$. Since this is an optimal equilibrium, $L' = L(ij)$ is the largest number of messages such that the strategy determined by (9) is feasible, which, by monotonicity, is the case if and only if

$$\begin{aligned}
&\begin{cases} c_1 = 1/L' - 2(L'-1)(b_j - b_i) > 0 \\ c_{L'-1} = (L'-1)/L' - 2(L'-1)(b_j - b_i) < 1 \end{cases} \\
\Leftrightarrow &1/(2L') > |(L'-1)(b_j - b_i)|. \qquad (10)
\end{aligned}$$

Thus,

$$L(ij) = \max\{l \in \mathbb{N} \mid 1/(2l) > |(l-1)(b_j - b_i)|\}. \qquad (11)$$

Note that (10) has only finitely many positive integer solutions, among them $L' = 1$, and thus, (11) is well-defined, which finishes the proof.

Proof of Theorem 2

To prove the theorem, we first construct a homogeneous Markov chain $\{\widetilde{x}(s)\}_{s\in\mathbb{N}} = \{(\widetilde{x}_i(s))_{i\in\mathcal{N}}\}_{s\in\mathbb{N}}$ in discrete time with finite n-dimensional state space $\mathcal{A} = \times_{i\in\mathcal{N}} A_i$, where A_i denotes the set of states for agent i. We know that we can replace the time-continuous belief dynamics process $\{x(t)\}_{t\geq 0}$ by the time-discrete process $\{x(s)\}_{s\in\mathbb{N}}$, where $x(s)$ is the vector of beliefs at time slot s. In the following, we also simplify the state space of the process. We find a partition of the unit interval such that it is enough to know in which element of the partition each agent's belief is.

Let $i \in \mathcal{N}$ and $C^i = \cup_{j\in\mathcal{N}_i} C^{ij}$ denote the set of points for which some agent $j \in \mathcal{N}_i$ is indifferent between two messages when communicating with i. Furthermore, $Y^i = \cup_{j\in\mathcal{N}_i} Y^{ij}$ denotes the set of agent i's interpretations. We assume without loss of generality that the set $C^i \cup Y^i$ consists of rational numbers for all $i \in \mathcal{N}$.[89] Then, there exists a lowest common denominator d of the set $\cup_{i\in\mathcal{N}} C^i \cup Y^i$.

This allows us to define the partition $C_d = \{k/d \mid 0 \leq k \leq d\}$ of $[0,1]$, where each partition element (without loss of generality) is an interval $[(k-1)/d, k/d)$, $k = 1, 2, \ldots, d$. This partition distinguishes the beliefs of the agents finely enough to keep track of how the belief dynamics process evolves as we will show. Take $i \in \mathcal{N}$, $j \in \mathcal{N}_i$ and suppose that

$$x_i(s-1) \in [(k_i-1)/d, k_i/d) \text{ and } x_j(s-1) \in [(k_j-1)/d, k_j/d),$$

$1 \leq k_i, k_j \leq d$. By construction of the partition, there exists $1 \leq l \leq L(ij)$ such that $x_j(s-1) \in [c_{l-1}, c_l)$, i.e., C_d is fine enough to determine the message $m^{ij}(x_j(s-1))$ sent in equilibrium by agent j. Moreover, also by construction, there exists $1 \leq \bar{k} \leq d-1$ such that the interpretation of this message is $y^{ij}\big(m^{ij}(x_j(s-1))\big) = \bar{k}/d$. And since $x_i(s-1) \in [(k_i-1)/d, k_i/d)$, it follows that

$$\begin{aligned} x_i(s) = 1/2(x_i(s-1) + \bar{k}/d) &\in [(k_i - 1 + \bar{k})/(2d), (k_i + \bar{k})/(2d)) \\ &\subseteq [(\lceil (k_i + \bar{k})/2 \rceil - 1)/d, \lceil (k_i + \bar{k})/2 \rceil/d), \end{aligned}$$

i.e., C_d is also fine enough to determine i's updated belief and altogether, it is fine enough to keep track of the belief dynamics process.

Therefore, we can identify the continuous state space $[0,1]^n$ of $\{x(s)\}_{s\in\mathbb{N}}$ with the finite state space $\mathcal{A} = A^n = \{a_1, a_2, \ldots, a_d\}^n$ of $\{\widetilde{x}(s)\}_{s\in\mathbb{N}}$, where $a_k \equiv [(k-1)/d, k/d)$, $k = 1, 2, \ldots, d$. In other words, a state $a \in \mathcal{A}$ specifies for each agent the partition element of C_d her belief is in at time slot s.

[89] If some number is irrational, then we can approximate it arbitrarily well by a rational number, e.g., using the method of continued fractions.

Let $\bar{x}(a_k) = (2k-1)/(2d)$ denote the average value of $[(k-1)/d, k/d)$ and furthermore, let $\tilde{y}^{ij}(a_k) = y^{ij}\big(m^{ij}(\bar{x}(a_k))\big)$ denote i's interpretation of j's message when j's belief is in $[(k-1)/d, k/d)$. We define the transition probabilities of $\{\widetilde{x}(s)\}_{s\in\mathbb{N}}$ as follows:

$$\mathbb{P}[\widetilde{x}(s) = (a^{-i}, a_l) \mid \widetilde{x}(s-1) = a] = 1/n \sum_{\substack{j\in\mathcal{N}_i: \\ (a^i,a^j)\in B^{ij}(l)}} p_{ij} \tag{12}$$

for all $a \in \mathcal{A}$ and $l \in \{1, 2, \ldots, d\}$, where

$$B^{ij}(l) = \{(a_k, a_{k'}) \in A^2 \mid 1/2[\bar{x}(a_k) + \tilde{y}^{ij}(a_{k'})] \in [(l-1)/d, l/d)\}$$

is the set of all pairs of individual states (a^i, a^j) such that agent i changes from state a^i to state $a^{i'} = a_l$ given that she updates from agent j who is in state a^j. All other transition probabilities (i.e., those where more than one component changes) are assumed to have zero probability. By construction, the following result holds.

Lemma 5. *$\{\widetilde{x}(s)\}_{s\in\mathbb{N}}$ is a homogeneous Markov chain with finite state space $\mathcal{A}$ and transition probabilities given by (12), and, in particular, at any time slot s,*

$$\widetilde{x}(s) = (a_{k_1}, a_{k_2}, \ldots, a_{k_n})' \text{ if and only if } x(s) \in \times_{i\in\mathcal{N}}[(k_i-1)/d, k_i/d).$$

Furthermore, for a set of states $Z \subseteq \mathcal{A}$, let $Z|_k = \{a \in A \mid \exists z \in Z : z_k = a\}$ denote the set of all possible values the k-th component of states in Z can take. Then, the following holds.

Lemma 6. *If $Z \subseteq \mathcal{A}$ is a recurrent communication class of $\{\widetilde{x}(s)\}_{s\in\mathbb{N}}$, then $\{\mathcal{I}_i\}_{i\in\mathcal{N}}$ is a minimal mutually confirming set of $\{x(t)\}_{t\geq 0}$, where*

(i) $\mathcal{I}_i = \bigcup_{k:\, a_k\in Z|_i}[(k-1)/d, k/d]$ if $|Z|_i| \geq 2$, and

(ii) $\mathcal{I}_i = (k-1)/d$ or $\mathcal{I}_i = k/d$ if $Z|_i = \{a_k\}$.

Proof. Suppose that Z is a recurrent communication class of $\{\widetilde{x}(s)\}_{s\in\mathbb{N}}$, i.e., the Markov chain will never leave this class and each state $z \in Z$ is visited infinitely often by $\{\widetilde{x}(s)\}_{s\in\mathbb{N}}$. We show that $\{\mathcal{I}_i\}_{i\in\mathcal{N}}$ is a minimal mutually confirming set of $\{x(t)\}_{t\geq 0}$.

Note that for $i \in \mathcal{N}$ and each individual state $z^i \in Z|_i$, it is $\widetilde{x}_i(s) = z^i$ for infinitely many time slots s. Let

$$Y_Z^i = \bigcup_{j\in\mathcal{N}_i} \bigcup_{z^j\in Z|_j} \tilde{y}^{ij}(z^j)$$

denote the set of all interpretations of agent i when $\widetilde{x}(s) \in Z$. Note that if $a_k, a_{k'} \in Z|_i$ for $k < k'$, then also $a_{k''} \in Z|_i$ for all $k < k'' < k'$. Thus, if $|Z|_i| \geq 2$,

$$\mathcal{I}_i = \bigcup_{k\colon a_k \in Z|_i} [(k-1)/d, k/d] = \operatorname{conv}(Y_Z^i)$$

since all intervals $[(k-1)/d, k/d]$ in the union are visited by i. On the other hand, if $Z|_i = \{a_k\}$, then i always uses the same interpretation when in Z, either $(k-1)/d$ or k/d. Hence, $\mathcal{I}_i = (k-1)/d = \operatorname{conv}(Y_Z^i)$ or $\mathcal{I}_i = k/d = \operatorname{conv}(Y_Z^i)$. Altogether, we have $\mathcal{I}_i = \operatorname{conv}(Y_Z^i)$ for all $i \in \mathcal{N}$. And furthermore, note that

$$\begin{aligned} Y_Z^i = \bigcup_{j \in \mathcal{N}_i} \bigcup_{z^j \in Z|_j} \tilde{y}^{ij}(z^j) &= \bigcup_{j \in \mathcal{N}_i} \bigcup_{a_k \in Z|_j} \tilde{y}^{ij}(a_k) \\ &= \bigcup_{j \in \mathcal{N}_i} \bigcup_{a_k \in Z|_j} y^{ij}\big(m^{ij}(\bar{x}(a_k))\big) \\ &= \bigcup_{j \in \mathcal{N}_i} \bigcup_{k\colon a_k \in Z|_j} y^{ij}\big(m^{ij}((2k-1)/(2d))\big) \\ &= \bigcup_{j \in \mathcal{N}_i} Y^{ij}\big|_{\mathcal{I}_j}\,, \end{aligned}$$

where the last equality follows from the definition of $\mathcal{I}_j$. Hence, we get $\mathcal{I}_i = \operatorname{conv}\left(\bigcup_{j \in \mathcal{N}_i} Y^{ij}|_{\mathcal{I}_j}\right)$ for all $i \in \mathcal{N}$, i.e., we have shown that $\{\mathcal{I}_i\}_{i \in \mathcal{N}}$ is a mutually confirming set of $\{x(t)\}_{t \geq 0}$ and furthermore, it is also minimal since by assumption Z is a recurrent communication class of $\{\widetilde{x}(s)\}_{s \in \mathbb{N}}$, which finishes the proof. □

Since the state space of $\{\widetilde{x}(s)\}_{s \in \mathbb{N}}$ is finite, there exists an a.s. finite stopping time τ such that for any initial state $\widetilde{x}(0) \in \mathcal{A}$,

$$\widetilde{x}(\tau) \in \{a \in \mathcal{A} \mid \exists Z \ni a \text{ recurrent communication class of } \{\widetilde{x}(s)\}_{s \in \mathbb{N}}\}.^{90}$$

So, suppose that $\widetilde{x}(\tau) \in Z$. We show that this implies that the original chain converges to the minimal mutually confirming set $\{\mathcal{I}_i\}_{i \in \mathcal{N}}$ defined in Lemma 6.

If $|Z|_i| \geq 2$, then part (i) of Lemma 6 implies that $x_i(\tau) \in \mathcal{I}_i$. Furthermore, since Z is a recurrent communication class of $\{\widetilde{x}(s)\}_{s \in \mathbb{N}}$, the boundaries of $\mathcal{I}_i$ are used infinitely often as interpretations by i and thus, $x_i(\tau)$ fluctuates on $\mathcal{I}_i$.

On the other hand, if $Z|_i = \{a_k\}$, agent i uses only a single interpretation when updating since $(k-1)/d$ and k/d cannot be both interpretations by choice of the partition C_d. This implies that, without loss of generality, $x_i(t) \to k/d = \mathcal{I}_i$ for $t \to \infty$, and hence, $x(t) \to \{\mathcal{I}_i\}_{i \in \mathcal{N}}$ for $t \to \infty$, which finishes the proof.

[90] We refer, e.g., to Brémaud (1999) for this result.

Chapter 5

Concluding Remarks

The objective of this thesis was to contribute to the literature on non-Bayesian social influence models by shedding light on three particular aspects of social influence: *anonymous influence*, *manipulation*, and *conflicting interests*. In a dynamic framework, we analyzed how these aspects affect long-run beliefs and opinions in society.

First, we studied influence processes modeled by *OWA operators*, which are the only *anonymous* aggregation functions. As one would expect, an aggregation model is anonymous if all agents use these functions. We characterized influential coalitions, showed that cyclic terminal classes cannot exist due to anonymity and characterized terminal states. Our main result provides a necessary and sufficient condition for convergence to consensus. Moreover, we extended our model to *decomposable* aggregation functions. In particular, this allows to combine OWA operators with the classical approach of ordinary weighted averages. It turned out that our previous condition on convergence to consensus is still sufficient in this generalized setting. We also analyzed the *speed of convergence* to terminal classes as well as *probabilities of absorption* by different terminal classes. For anonymous models, we were able to reduce the computational demand substantially compared to the general case. Furthermore, we applied our results to *fuzzy linguistic quantifiers* and showed that if agents use in some sense similar quantifiers and not too many agents deviate from these quantifiers, the society will eventually reach a consensus.

Second, we introduced the possibility of *manipulation* into the model by DeGroot (1974). We showed that manipulation can modify the *trust structure* and lead to a connected society, and thus, to consensus. Furthermore, we found that manipulation fosters *opinion leadership* in the sense that the manipulating agent always increases her influence on the long-run beliefs. And more surprisingly, this may even be the case for the manipulated agent. The expected change of influence

on the long-run beliefs is ambiguous and depends on the agents' preferences and the social network. We also showed that the trust structure of the society settles down and, if the satisfaction of agents does not directly depend on the trust, manipulation will come to an end and they reach a consensus (under some weak regularity condition). Regarding learning, we were interested in the question whether manipulation is beneficial or harmful for *information aggregation*. We used an approach similar to Acemoglu et al. (2010) and showed that manipulation reduces the extent of misinformation in the society if manipulation is rather costly and the agents underselling their information gain and those overselling their information lose overall influence. Not surprisingly, agents for whom manipulation is cheap can severely harm information aggregation. Furthermore, our main example suggests that homogeneous preferences for manipulation favor a reduction of the extent of misinformation in society.

Finally, we introduced conflicting interests into a model of non-Bayesian belief dynamics. We disentangled the terms *belief* and *opinion*: the belief of an individual about some issue of common interest is what she holds to be true given her information, and her opinion is what is ought to be the answer to the issue given her bias. We assumed that when two individuals communicate, the receiver of information would like to learn the true belief of the sender about the issue, while the sender wants to spread his opinion. This conflict of interest leads to noisy communication à la Crawford and Sobel (1982) in equilibrium. In our main result, we showed that the belief dynamics process converges to a set of intervals that is *minimal mutually confirming*. Furthermore, we showed that the belief of an agent eventually *fluctuates* on her corresponding interval whenever the interval is proper. We remarked that outcomes with a steady state are non-generic as long as conflicts of interest are small and some agents communicate with several different agents. Hence, we can conclude that the introduction of conflict of interest leads not only to persistent disagreement among the agents, but also to fluctuating beliefs and opinions. Remarkably, though frequently observed in social science, the phenomenon of fluctuation is barely studied in the literature on communication in social networks, the only exception being Acemoglu et al. (2013).

Our results contribute to the understanding of social influence and the evolution of beliefs and opinions in our societies. In the following, we broadly discuss areas for future research on this topic.

In Chapter 2, we presented a theoretical framework to study the phenomenon of anonymous influence. We understand this as a starting point and further research will be required to develop a deeper understanding of this important topic. One

possible avenue of research is to generalize the model to extended OWA operators, which do not depend on the number of agents and are therefore more flexible. For instance, this would allow to introduce a network into the model such that agents only get to know (directly) the beliefs of neighbors. Moreover, the implications of anonymity for learning and spread of misinformation could be explored.

The latter also constitute a promising area for future research on their own. After DeMarzo et al. (2003) and Golub and Jackson (2010) had studied learning in the DeGroot model, Acemoglu et al. (2010, 2013) shifted the focus more toward spread of misinformation. They showed that forceful agents, who almost do not change their own beliefs, can prevent efficient information aggregation and furthermore, that stubborn agents, who never change their beliefs, can completely prevent information aggregation. We have contributed to this stream of literature by studying manipulation within the framework of the DeGroot model. One insight from these works is that somehow powerful agents can severely harm information aggregation, and in some cases even prevent it. Since this seems to be less likely in models where agents are at least partially Bayesian, it will be necessary to study manipulation in partially Bayesian models in order to gain a deeper understanding of the issue. Notice however that there is only little room left for manipulation when agents are fully rational. The degree of rationality of the agents seems therefore to be a key determinant in future research.

Another important avenue of research deals with the question whether society reaches a consensus or whether disagreement persists. Many classical models of dynamic social influence, both Bayesian and non-Bayesian, lead to mutual consensus among the agents in the long-run under the condition that the network is strongly connected, see, e.g., DeGroot (1974); Acemoglu et al. (2010); Gale and Kariv (2003). Since this result seems not to be plausible in many situations in real life, several authors have proposed models to explain non-convergence of beliefs, see, e.g., Axelrod (1997); Golub and Jackson (2012); Acemoglu et al. (2013). However, in most of these models the reason for persisting disagreement is some kind of homophily, which either leads to segregated societies or substantially slows down convergence and thus leads to a high persistence of disagreement. To this respect, Acemoglu et al. (2013) is an exception, where the authors introduce stubborn agents that never change their belief, which leads not only to persistent disagreement, but also to fluctuating beliefs when the other agents update regularly from different stubborn agents. We provided another explanation for persistent disagreement and fluctuation of beliefs in Chapter 4: conflict of interest. Although there are now several different explanations for non-convergence of beliefs, the picture stays incomplete and more insights

are necessary to complement it.

In particular, further investigation on conflict of interest seems to be promising. In our contribution, we introduced agents with conflicting agents that, broadly speaking, are rational within periods, but naïve across periods and showed that this leads to non-truthful communication and fluctuating beliefs. On the other hand, Acemoglu et al. (2014) have studied a model of Bayesian learning where agents might want to misreport their information in order to delay the decisions of other agents, i.e., they have also introduced a kind of conflict of interest. However, their model still leads to mutual consensus under the condition that the underlying social network is strongly connected. Thus, further research will be necessary to clarify the role of conflict of interest in belief formation processes.

Bibliography

Acemoglu, D., K. Bimpikis, and A. Ozdaglar (2014). Dynamics of information exchange in endogenous social networks. *Theoretical Economics 9*(1), 41–97.

Acemoglu, D., G. Como, F. Fagnani, and A. Ozdaglar (2013). Opinion fluctuations and disagreement in social networks. *Mathematics of Operations Research 38*(1), 1–27.

Acemoglu, D., M. Dahleh, I. Lobel, and A. Ozdaglar (2011). Bayesian learning in social networks. *The Review of Economic Studies 78*(4), 1201–36.

Acemoglu, D. and A. Ozdaglar (2011). Opinion dynamics and learning in social networks. *Dynamic Games and Applications 1*(1), 3–49.

Acemoglu, D., A. Ozdaglar, and A. ParandehGheibi (2010). Spread of (mis)information in social networks. *Games and Economic Behavior 70*(2), 194–227.

Ambrus, A. and S. Takahashi (2008). Multi-sender cheap talk with restricted state spaces. *Theoretical Economics 3*(1), 1–27.

Asavathiratham, C. (2000). The influence model: A tractable representation for the dynamics of networked Markov chains. *Ph.D. thesis*. Massachusetts Institute of Technology.

Austen-Smith, D. and J. R. Wright (1994). Counteractive lobbying. *American Journal of Political Science 38*(1), 25–44.

Axelrod, R. (1997). The dissemination of culture a model with local convergence and global polarization. *Journal of conflict resolution 41*(2), 203–26.

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

Bala, V. and S. Goyal (2001). Conformism and diversity under social learning. *Economic Theory 17*(1), 101–20.

Banerjee, A. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics 107*(3), 797–817.

Banerjee, A. and D. Fudenberg (2004). Word-of-mouth learning. *Games and Economic Behavior 46*(1), 1–22.

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

Blondel, V. D., J. M. Hendrickx, and J. N. Tsitsiklis (2009). On Krause's multi-agent consensus model with state-dependent connectivity. *Automatic Control, IEEE Transactions on 54*(11), 2586–97.

Bodansky, D. (2010). The Copenhagen climate change conference: a postmortem. *American Journal of International Law 104*(2), 230–40.

Borm, P., R. van den Brink, and M. Slikker (2002). An iterative procedure for evaluating digraph competitions. *Annals of Operations Research 109*, 61–75.

Boyd, S., A. Ghosh, B. Prabhakar, and D. Shah (2006). Randomized gossip algorithms. *Information Theory, IEEE Transactions on 52*(6), 2508–30.

Brémaud, P. (1999). *Markov Chains*. Springer.

Büchel, B., T. Hellmann, and S. Klößner (2012). Opinion dynamics under conformity. Center for Mathematical Economics Working Papers 469, Bielefeld University.

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

Calvó-Armengol, A. and M. Jackson (2009). Like father, like son: social network externalities and parent-child correlation in behavior. *American Economic Journal: Microeconomics 1*(1), 124–50.

Chandrasekhar, A., H. Larreguy, and J. Xandri (2012). Testing models of social learning on networks: evidence from a framed field experiment. Mimeo, Massachusetts Institute of Technology.

Choi, S., D. Gale, and S. Kariv (2012). Social learning in networks: a quantal response equilibrium analysis of experimental data. *Review of Economic Design 16*(2-3), 135–57.

Cohen, G. (2003). Party over policy: the dominating impact of group influence on political beliefs. *Journal of personality and social psychology 85*(5), 808–22.

Como, G. and F. Fagnani (2011). Scaling limits for continuous opinion dynamics systems. *The Annals of Applied Probability 21*(4), 1537–67.

Corazzini, L., F. Pavesi, B. Petrovich, and L. Stanca (2012). Influential listeners: an experiment on persuasion bias in social networks. *European Economic Review 56*(6), 1276–88.

Cornelis, C., P. Victor, and E. Herrera-Viedma (2010). Ordered weighted averaging approaches for aggregating gradual trust and distrust. In *XV congreso Español sobre tecnologías y lógica fuzzy, Proceedings*, pp. 555–60. Ghent University, Department of Applied Mathematics and Computer Science.

Crawford, V. and J. Sobel (1982). Strategic information transmission. *Econometrica 50*(6), 1431–51.

Deffuant, G., D. Neau, F. Amblard, and G. Weisbuch (2000). Mixing beliefs among interacting agents. *Advances in Complex Systems 3*, 87–98.

DeGroot, M. (1974). Reaching a consensus. *Journal of the American Statistical Association 69*(345), 118–21.

DeMarzo, P., D. Vayanos, and J. Zwiebel (2003). Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics 118*(3), 909–68.

Ellison, G. (1993). Learning, local interaction, and coordination. *Econometrica 61*(5), 1047–71.

Förster, M., M. Grabisch, and A. Rusinowska (2013). Anonymous social influence. *Games and Economic Behavior 82*, 621–35.

French, J. R. P. (1956). A formal theory of social power. *Psychological Review 63*(3), 181–94.

Friedkin, N. E. (1991). Theoretical foundations for centrality measures. *American journal of Sociology 96*(6), 1478–504.

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

Gale, D. and S. Kariv (2003). Bayesian learning in social networks. *Games and Economic Behavior 45*(2), 329–46.

Galeotti, A., C. Ghiglino, and F. Squintani (2013). Strategic information transmission networks. *Journal of Economic Theory 148*(5), 1751–69.

Galeotti, A. and S. Goyal (2009). Influencing the influencers: a theory of strategic diffusion. *The RAND Journal of Economics 40*(3), 509–32.

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

Golub, B. and M. O. Jackson (2012). How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics 127*(3), 1287–338.

Grabisch, M., J.-L. Marichal, R. Mesiar, and E. Pap (2009). *Aggregation Functions*. Number 127 in Encyclopedia of Mathematics and its Applications. Cambridge University Press.

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

Grabisch, M. and A. Rusinowska (2011). Influence functions, followers and command games. *Games and Economic Behavior 72*(1), 123–38.

Grabisch, M. and A. Rusinowska (2013). A model of influence based on aggregation functions. *Mathematical Social Sciences 66*(3), 316–30.

Gullberg, A. T. (2008). Lobbying friends and foes in climate policy: the case of business and environmental interest groups in the European Union. *Energy Policy 36*(8), 2964–72.

Hagenbach, J. and F. Koessler (2010). Strategic communication networks. *The Review of Economic Studies 77*(3), 1072–99.

Harary, F. (1959). A criterion for unanimity in French's theory of social power. In D. Cartwright (Ed.), *Studies in social power*, pp. 168–82. University of Michigan.

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

Hunter, J. J. (2005). Stationary distributions and mean first passage times of perturbed Markov chains. *Linear Algebra and its Applications 410*, 217–43.

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press.

Jackson, M. O. and L. Yariv (2007). Diffusion of behavior and equilibrium properties in network games. *The American Economic Review 97*(2), 92–8.

Jäger, G., L. Metzger, and F. Riedel (2011). Voronoi languages: equilibria in cheap-talk games with high-dimensional types and few signals. *Games and Economic Behavior 73*(2), 517–37.

Jiang, H. and R. Eastman (2000). Application of fuzzy measures in multi-criteria evaluation in GIS. *International Journal of Geographical Information Science 14*(2), 173–84.

Kramer, G. (1971). Short-term fluctuations in US voting behavior, 1896–1964. *American Political Science Review 65*(1), 131–43.

López-Pintado, D. (2008). Diffusion in complex social networks. *Games and Economic Behavior 62*(2), 573–90.

López-Pintado, D. (2012). Influence networks. *Games and Economic Behavior 75*(2), 776–87.

López-Pintado, D. and D. Watts (2008). Social influence, binary decisions and collective dynamics. *Rationality and Society 20*(4), 399–443.

Lorenz, J. (2005). A stabilization theorem for dynamics of continuous opinions. *Physica A: Statistical Mechanics and its Applications 355*(1), 217–23.

Malczewski, J. and C. Rinner (2005). Exploring multicriteria decision strategies in GIS with linguistic quantifiers: a case study of residential quality evaluation. *Journal of Geographical Systems 7*(2), 249–68.

Mäs, M. (2010). *The diversity puzzle*. ICS dissertation series. Ridderprint Offsetdrukkerij B.V., Ridderkerk.

McKelvey, R. D. and T. R. Palfrey (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior 10*(1), 6–38.

Morris, S. (2000). Contagion. *The Review of Economic Studies 67*(1), 57–78.

Mueller-Frank, M. (2010). A fixed point convergence theorem in Euclidean spaces and its application to non-Bayesian learning in social networks. *Working paper, University of Oxford*.

van den Brink, R. and R. Gilles (2000). Measuring domination in directed networks. *Social Networks 22*(2), 141–57.

Van Dijk, T. A. (2006). Discourse and manipulation. *Discourse Society 17*(3), 359–83.

Watts, D. (2003). Six degrees: the science of a connected age. *New York City, NY: WW Norton and Company*.

Yager, R. (1988). On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on Systems, Man and Cybernetics 18*(1), 183–90.

Yager, R. and J. Kacprzyk (1997). *The Ordered Weighted Averaging Operators*. Kluwer Academic Publishers.

Zadeh, L. (1983). A computational approach to fuzzy quantifiers in natural language. *Computers and Mathematics with Applications 9*(1), 149–84.

Zaller, J. (1992). *The nature and origins of mass opinion*. Cambridge University Press.