UNIVERSITÉ LAVAL

Three Essays in Labor Economics and the Economics of Networks

Thèse

Elfried Faton

Doctorat en économique
Philosophiæ doctor (Ph. D.)

Québec, Canada

Three Essays in Labor Economics and the Economics of Networks

Thèse

Elfried G. F. FATON

Sous la direction de:

Bernard Fortin, directeur de recherche
Vincent Boucher, codirecteur de recherche

Résumé

Cette thèse porte sur l'influence des interactions sociales et des structures de réseaux sur divers enjeux économiques. De manière spécifique, la thèse fournit de nouveaux résultats expliquant l'impact des interactions sociales sur l'effort, la performance, la productivité au travail d'individus, ainsi que leurs croyances sur des sujets variés. En particulier, le chapitre 1 expose de nouveaux résultats empiriques au sujet des variables expliquant l'effort, la qualité des soins offerts et la performance des professionnels de santé maternelle et néo-natale d'un pays en développement (le Bénin). Ces résultats sont obtenus dans un contexte où ils reçoivent des salaires fixes indépendants de leur performance. De plus, le chapitre 2 complète les conclusions du premier chapitre en précisant certains résultats clés concernant la productivité de ces professionnels de santé. Pour ce faire, une mesure de leurs pouvoirs de négociation individuels en milieu de travail est proposée. Quant au chapitre 3, il se positionne davantage dans la littérature sur la théorie de la formation d'opinions en réseaux. Il développe des résultats nouveaux sur la convergence des croyances et l'atteinte d'un consensus au sein d'un réseau d'individus. Plus spécifiquement, il évalue l'influence de certains biais cognitifs sur le processus de mise à jour des croyances. Les résultats de la thèse se résument comme suit.

Le chapitre premier utilise une approche en jeux non-coopératifs pour mettre en lumière l'existence d'un mécanisme de substituabilité stratégique des professionnels de santé maternelle et néo-natale en milieu de travail au Bénin. D'une part, les résultats du chapitre suggèrent que, dans le but de produire un certain niveau de qualité de soins aux patients de leur formation sanitaire, certains professionnels de santé (altruistes) augmentent leur effort afin de compenser la qualité de soins insuffisante produite par leurs collègues. D'autre part, grâce à certaines informations fournies dans les données utilisées, une méthode probabiliste simple est décrite dans ce chapitre, pour tenir compte des variabilités éventuelles dans le poids des interactions entre collègues.

Le chapitre 2, quant à lui, s'intéresse également professionnels de santé maternelle et néo-natale. Toutefois, il propose une autre théorie permettant de mieux comprendre certains mécanismes qui sous-tendent la substituabilité stratégique obtenue à l'équilibre dans le chapitre 1. Plus précisément, ce chapitre présente une approche par équilibre de négociation *à la* Nash, afin d'expliquer comment certaines caractéristiques individuelles déterminent le pouvoir de

négociation de ces travailleurs et, par la même occasion, leur part de travail. Les résultats obtenus montrent que certaines caractéristiques sociales telles que l'éducation, l'expérience et le nombre d'enfants des travailleurs déterminent leur pouvoir de négociation au travail et ainsi donc, leur productivité.

Enfin, le chapitre 3 explore l'impact de certains biais cognitifs sur les propriétés de convergence et de consensus en réseau connues jusque-là, en ce qui a trait au modèle naïf d'apprentissage de Degroot. Ainsi, en présence d'un *biais de confirmation* et d'un *biais de supériorité relative des extrémistes*, le chapitre démontre que même dans un réseau apériodique et fortement connecté, les croyances ne convergent pas nécessairement vers un consensus. En plus de cela, ce chapitre développe quelques caractéristiques des structures de réseau *à priori* et des vecteurs de croyances initiales qui affectent l'existence d'un consensus. Globalement, ce dernier chapitre de la thèse propose une interprétation nouvelle de quelques mécanismes clés à la base d'enjeux sociaux tels que le radicalisme politique, les comportements extrémistes en société, ou encore la non-convergence des croyances au sein de divers réseaux d'individus.

Abstract

This thesis is about the influence of social interactions and network structure on various economic outcomes. Specifically, the thesis presents new findings explaining how social interactions shape individual outcomes like their effort, performance and productivity in the workplace, as well as their beliefs on miscellaneous social matters. Specifically, Chapter 1 gives new empirical results on some variables affecting the effort, quality of healthcare provided, and performance of maternal and child health (MCH) workers from a developing country (Benin). The results are obtained in a context of fixed salaries irrespective of workers' performance. In addition, Chapter 2 complements the results in Chapter 1, by explaining some of its main results on workers' productivity, in light of their bargaining power in the workplace. As for Chapter 3, it stands in the theory of opinion formation in a network. This chapter gives new results on the convergence of individual beliefs and reaching a consensus within a network when we consider a few cognitive biases in individuals' behavior. More specifically, the results of this thesis are summarized as follows.

Chapter 1 uses a non-cooperative game approach to bring to light the existence of strategic substitutability in the workplace of MCH workers in Benin. Particularly, the paper suggests that, to provide collectively a certain quality of healthcare in their health facility, some workers (altruists) increase their effort to compensate for the failure of their peers in offering a good quality of care. Moreover, using some relevant information in the data, the chapter also proposes a simple probability-based method to account for some variability in the strength of interactions among colleagues.

Chapter 2 on the other hand, focuses on the same MCH workers, and proposes a new theory to understand better some mechanisms behind the equilibrium expressed by the strategic substitutability obtained in Chapter 1. More specifically, the chapter presents a simple Nash-bargaining approach to establish how individual characteristics mold their bargaining power and consequently their workload share. The results show that workers social characteristics like their education, experience and number of children determine their bargaining power in the workplace, and thus their productivity.

Finally, Chapter 3 explores how some cognitive biases affect convergence and consensus prop-

erties known up to now in an average-based model of opinion formation. In particular, when accounting for a *confirmation bias* and an *extremist relative superiority bias*, the chapter reveals that, in an *a priori* strongly connected and aperiodic network, beliefs do not necessarily converge to a consensus. Furthermore, some typical features of *a priori* networks and vectors of initial beliefs which influence the existence of a consensus are given. Overall, the chapter proposes a new understanding of some mechanisms behind social issues like political radicalism, extreme behaviors and the non-convergence of opinions within a network.

Table des matières

Liste des tableaux

Liste des figures

To The One who Loved me first,

To my parents, Mathias Faton
and Lydie Yonlonfoun

You never know how much you
really believe anything until its
truth or falsehood becomes a
matter of life and death to you.

C.S. Lewis

Remerciements

Cette thèse n'aurait pu être réalisée sans le soutien de plusieurs personnes qui à jamais auront laissé une marque importante dans ma vie. De prime abord, je remercie infiniment mon directeur de thèse Bernard Fortin qui a accepté de diriger ma thèse et dont le précieux soutien et les encouragements m'ont permis d'aboutir à ce résultat. J'aimerais également lui témoigner ma profonde gratitude pour les bourses et soutiens financiers qui m'ont été accordés dans le processus de rédaction de cette thèse. Ma profonde reconnaissance va également à Vincent Boucher qui a consenti très tôt à se joindre à la direction de ma thèse et grâce à qui j'ai pu élargir les horizons de mes travaux. Je tiens à le remercier chaleureusement pour son implication, ses encouragements et sa positivité à toute épreuve dans les moments les plus durs. Ils ont été pour moi de véritables bouffées d'air frais dans les moments cruciaux. Je remercie également les professeurs Léonard Wantchékon, Sylvain Dessy et Guy Lacroix qui ont contribué à rendre cette aventure possible. Merci également aux professeurs Charles Bellemare, Bruce Shearer et Myra Yazbeck pour avoir accepté d'évaluer cette thèse.

J'aimerai également adresser mes remerciements au professeur Philippe Barla, directeur actuel du département d'économique de l'Université Laval, ainsi qu'aux directeurs du CRREP (précédemment CIRPÉE) et du troisième cycle qui se sont succédés pendant la durée de mes études. Merci pour le soutien et les bourses qui m'ont été accordées pendant mes études. Ma reconnaissance va également à "ma gang" de promotion Marie-Albertine, Bodel et Carolle. Merci d'avoir rendu nos veilles d'études aussi intéressantes. Merci également à mes collègues : Gilles, Rokhaya, Jean-Armand, Aimé, Sétou, André-Marie, Thomas, Marius, Mélissa, Antoine, Mbéa, Isaora et Steve.

Le résultat de cette thèse n'aurait pu être obtenu sans le soutien moral et les efforts de plusieurs autres personnes chères à mon cœur. Je ne saurai commencer sans témoigner ma reconnaissance aux deux personnes qui ont chacune consenti d'énormes sacrifices et des efforts considérables pour que j'accède très tôt à une bonne éducation : mon papa, Faton Mathias et ma maman, Yonlonfoun Lydie. Merci également à mes sœurs Placide, Maïlys et Iris, à ma grande amie Pamela.

Partir dans le but poursuivre une thèse à l'étranger, c'est certes s'éloigner de tous ceux qu'on connait, notre famille, nos amis, mais c'est aussi rencontrer et se rapprocher de nouvelles personnes. J'aimerai remercier ma grande famille de Québec, qui ne fait que s'élargir jusqu'aux extrémités de la terre. Merci à mes bons amis Bertha, Adèle, Christiane, Sara, Franck, Aurore, Nora, Élodie, Dimitri, Mariam, Chancelle et Golda. Ma reconnaissance va également à Iris, Jean-Philippe ainsi qu'aux couples Michel et Diane, Yves et Alice, Miché et Armelle, pour leur soutien considérable dans les périodes les plus difficiles. Merci pour leurs paroles de connaissance et d'encouragement qui m'ont aidé à persévérer jusqu'au bout de cette thèse. Merci à toute ma grande famille de l'ÉVA, ainsi qu'à tous les membres de la chorale Chœur et Christ, et notre maestro Jessy. Vous êtes une véritable famille pour moi !

Enfin, j'aimerai adresser mes remerciements à la personne la plus importante de ma vie, Celui par qui je suis, mon meilleur Ami, le plus fidèle de tous, Celui sans qui rien de tout ceci n'aurait pu être : mon Unique Seigneur et Sauveur, Jésus-Christ. Je veux refléter qui tu es à jamais !

Avant-propos

Cette thèse est composée de trois chapitres représentant chacun des articles à soumettre pour publication dans des revues à comité de lecture.

Le chapitre premier est un article co-rédigé avec mon directeur de thèse Bernard Fortin. Cet article, dont je suis l'auteur principale, fait actuellement l'objet de révisions dans le but d'être soumis dans une revue scientifique traitant des questions relatives à l'économie du travail et à l'économie du développement.

Le second chapitre est un article dont je suis l'unique auteur. Il est également en cours de révision. Une version plus élaborée sera soumise dans une revue scientifique spécialisée en économie du travail.

Le dernier chapitre est un article dont je suis également l'auteur unique. Une version de cet article sera très prochainement produite et soumise dans une revue scientifique à comité de lecture spécialisée en économie du comportement ou en économie politique.

Introduction

In recent decades, successive studies on the influence of social interactions on various economic outcomes have been unveiled. This continuously growing literature on the economics of networks has revealed, across several fields, that most individuals are influenced by their peers in their behaviors, beliefs and choices. For instance, in the domain of labor, previous authors have addressed workers' performance and productivity compared with their peers', and they find convincing empirical results (see Falk and Ichino (2006), Mas and Moretti (2009), Barankay (2012), Gill and Prowse (2012), Beugnot et al. (2019)). On another hand, in a very different field, some authors have analyzed the process of opinion formation among individuals connected in a network, and find interesting results (see DeMarzo et al. (2003), Golub and Jackson (2010), Golub and Jackson (2012)). Yet, there still is a lot to do in the broad field of the economics of networks. My thesis focuses on two distinct areas in this field. Firstly, it makes several empirical contributions to what has yet been found on the role played by social interactions in determining workers' productivity. Second, in a theoretical framework, it discusses how cognitive biases shape individual beliefs and opinion formation in a network, over time.

Concerning the first area addressed in this thesis, a part of the literature have often shown that under favorable remuneration conditions, working with highly productive peers can have a positive effect on workers' own productivity. For instance, Falk and Ichino (2006), Mas and Moretti (2009), and several others, give empirical evidence for a positive peer influence on workers' productivity. However, most of the literature on the subject address questions related to co-workers' influence on workers' productivity, mainly, in fair working environments, and often under performance-based remuneration schemes. In this thesis, however, I address similar questions, but in a context of a particular developing country where workers are less privileged than their counterparts from developed regions who are usually the ones depicted in the literature. Specifically, I study a case of maternal and child health (hereafter MCH) workers, including doctors, midwives, nurses and mostly nursing auxiliaries. These MCH workers are predominantly women and paid fixed salaries, with a positive probability of not even receiving these salaries on time.

Such a context is interesting to address for a few compelling reasons. First, the domain of healthcare is usually perceived as an area in which most workers have very altruistic and generous values. Therefore, while accounting for social interactions in their workplace, exploring what influences their productivity in a less "fair" working context, provides a reference framework for workers in other fields who also face similar "imperfections" in their working environment. Second, concerning these so-called imperfections, such as workers' salary arrears, it is also useful to assess their part in these workers' productivity. Lastly, given the high proportion of nursing auxiliaries among the healthcare staff and their significant role in most developing countries, it is relevant to explore their productivity while considering all the intertwined mechanisms at play.

Specifically, using data from Benin health workers, the empirical part of this thesis contributes to the literature by providing some answers to a few questions including the following.

- What is the behavioral response of workers to their peers productivity in an environment where their salaries are fixed whatever their performance, and where they may not receive these salaries in full, or may have salary arrears ?
- What individual characteristics may influence workers' productivity in a framework in which social interactions in the workplace are accounted for ?
- What are a few peer characteristics likely to influence workers' productivity ?
- What individual characteristics influence their power to bargain when it comes to working in order to fulfill a certain share of the total demand for healthcare in their health facility ?
- How does individual bargaining power affect their workload share ?

To answer most of these questions, the thesis uses a spatial autoregressive (SAR) model founded on a non-cooperative game approach. A cooperative game reasoning is also used to address a few other concerns. Besides, the thesis proposes a simple probabilistic method to account for some variability in peers influence weights, when these weights are not fully known while estimating a SAR model.

Furthermore, in the second area of this thesis which concerns opinion formation in networks, I propose an intuitive way to introduce a few cognitive biases in a naive (average-based) learning model. I also analyze their impact on beliefs over time. More specifically, when individuals are subject to a confirmation bias and an extremist relative superiority bias, I analyze conditions "*a priori*" for the convergence of beliefs and for their convergence to a consensus within their network. In particular, I address the convergence of beliefs to extreme or near extreme values. I also explain some underlying mechanisms of political radicalism, the prevalence of extreme behaviors in some networks and the fact that opinions do not necessarily converge in real-world networks.

In the remainder of this thesis, Chapter 1 and Chapter 2 investigate the productivity of maternal and child care workers from Benin in terms of quality and quantity of healthcare respectively. Chapter 3, on the other hand, explores how some known properties of convergence and consensus in an average-based model of opinion formation, change when we account for a few cognitive biases.

Chapitre 1

Strategic Substitutability in the Workplace : an Empirical Evidence

Elfried Faton and Bernard Fortin

1.1 Résumé

Cet article analyse l'effort et la productivité des professionnels de soins maternels et néo-nataux, dans un contexte où ils reçoivent des salaires fixes. Dans ce contexte particulier, nous évaluons les effets de pairs en milieu du travail. Nos résultats exposent une évidence empirique de substituabilité stratégique entre collègues, lorsque ceux-ci interagissent en groupe pour produire (collectivement) une certaine qualité de soin. Plus spécifiquement, l'article suggère un mécanisme de compensation du faible niveau d'effort des uns par leurs collègues plus altruistes. Les données empiriques que nous utilisons proviennent de la Banque Mondiale. De plus, dans cet article, dû au fait que la pondération réelle de l'influence des pairs n'est pas donnée, nous proposons l'usage d'une matrice espérée d'interactions sociales pour tenir compte des possibles variations dans le niveau individuel d'influence des pairs. Cette matrice espérée d'interactions sociales est enfin utilisée dans un modèle de type spatial autorégressif (SAR) pour estimer les effets de pairs.

Mots clés : salaires fixes, effets de pairs, substitut stratégique, matrice espérée d'interactions sociales, soins maternels et néo-nataux, pays en développement.

Codes JELF : C31, I15, J24.

1.2 Abstract

This paper investigates effort and productivity of maternal and child health (MCH) workers, in a context where they are subject to fixed wages. In that context, we explore peer effects in effort at work. We find an empirical evidence of strategic substitutability in the workplace, for co-workers interacting in a group to produce (collectively) a defined level of healthcare quality. Specifically, the paper suggests that, to maintain a given quality of healthcare in their health facility, altruistic workers increase their effort, to compensate for their shirking peers' failure to provide a good quality of healthcare. We use the World Bank's empirical data on health workers and health facilities in a developing country. As the real interaction weights among co-workers are not fully known in the data, we propose an expected social interaction matrix to account for possible variations in peer influence. Finally, peer effects are estimated using the expected interaction matrix in a Spatial Autoregressive (SAR) model structure.

Keywords: fixed wage, peer effect, strategic substitute, expected social interaction matrix, maternal and child care, developing country.

JEL codes: C31, I15, J24.

---

0. We are grateful to Vincent Boucher, Rokhaya Dieye, Gilles Koumou, participants in "Seminaire du jeudi midi" at Departement d'Économique (Université Laval), and participants at the 2018 International Association for Applied Econometrics (IAAE) conference for useful comments and support. We also thank participants at 2015 SIER conference for their comments. Our gratitude goes to the team of the World Bank's Result-Based Financing (RBF) project directed by Christophe Lemière for our early access to the database. Any remaining errors are our own.

1.3 Introduction

Within recent decades, an extensive literature in economics has been developed to address the fact that people's choice and behavior are often influenced by their peers. These peers influence, also known as *peer effects*, are explored in various areas like education (Calvo-Armengol et al. (2009), Imberman et al. (2012), Boucher et al. (2014), etc.), health (Eisenberg et al. (2014), Fortin and Yazbeck (2015)) and several others including labor (Falk and Ichino (2006), Mas and Moretti (2009), Rosaz et al. (2012), Beugnot et al. (2019), Cornelissen et al. (2017)). Specifically in labor economics, Kandel and Lazear (1992), Falk and Ichino (2006), Mas and Moretti (2009) and several other authors find evidence that workers are also influenced by their peers in their productivity. In general, they find a positive effect of an increase in workers' productivity on their co-workers' productivity. This is interpreted in the literature as strategic complementarity, and often happens when workers' salaries are positively related with their performance at work. On the other hand, strategic substitutability in effort or quality, as portrayed further in this paper, implies that variations in workers' effort or quality tend to be negatively related to their peers'.

In this paper, we investigate effort and productivity at work, for Maternal and Child Health (MCH) Workers, in a context where they are subject to fixed wages. More specifically, we study the presence of peer effect among co-workers, by exploring empirically workers' behavioral response to peer effort, in a fixed wages context. To do so, we introduce the microeconomic foundations, explaining MCH workers' choice of effort, to provide collectively a certain level of quality healthcare. Each worker's chosen effort at (Nash) equilibrium is then used to obtain a spatial autoregressive (SAR) model. We use a Generalized Instrumental Variables (GIV) strategy introduced by Kelejian and Prucha (1998) and readapted by Bramoullé et al. (2009) to estimate the parameters of the model. Peer effects in effort is analyzed, and we find an empirical evidence of negative peer effects, interpreting strategic substitutability among MCH workers. Our results suggest two possible mechanisms. First, some MCH workers tend to "free ride," as their peers increase effort. Second, they are some altruistic workers who tend to increase effort, in response to shirking peers, to guarantee a defined level of quality healthcare collectively in their health facility.

Our study is done in a context where workers interact in a health facility (HF), to produce a certain level of quality of healthcare. Each worker receives a fixed wage, regardless of the quality of healthcare they produce (individually). Wages are rather based on workers' characteristics, like their education, job category (doctor, midwife, nurse or nursing auxiliary), and experience in maternal and child care. Besides, some workers may receive bonuses, based on seniority. However, some overdue wages and bonuses often characterize workers' job environments. Furthermore, in the context of the developing country we explore, Benin, there is a general shortage of qualified health workers. This, combined with several other issues, reduces employers' ability (HFs) to fire their personnel unless in extremely rare situations (See Chaudhury et al. (2006), Dizon-Ross et al. (2017)). Considering such a unique job environment, we expected our analyze of workers' effort to entail some particular mechanisms.

Throughout the study, effort refers to any action taken by MCH workers to produce a certain level of healthcare quality for mothers and children. Therefore, effort may for instance refer to action taken by a health worker to maintain or to improve their skills and knowledge in maternal and child care. This improvement may happen through learning (e.g., learning from interactions and discussions with peers, studying older or newer methods or discoveries in medicine) or practice (treating more patients, or finding solutions to difficult cases). It is necessary to think of productivity in terms of quality of healthcare given to patients, rather than quantity of patients treated.

In the paper, we account for the fact that individuals do not directly observe their peers' real effort. Yet, they can observe the resulting quality of healthcare they produce.[1] Therefore, the negative peer effect we estimate, which we interpret as a strategic substitutability in effort, happens through the effect of peer productivity on effort.[2] Briefly, after observing several indicators of the quality of healthcare provided by their peers, MCH workers "guess" their peers real effort, and finally choose their own effort accordingly.

Our paper makes a number of contributions to the literature. First, it contributes to the literature on social interactions and peer effects, by introducing an empirical evidence of negative peer effect in groups. In fact, this is to our knowledge, a first empirical evidence of strategic substitutability in the workplace. Previous research on peer effects in workers' productivity or quality, often find a positive relation between workers and peers outcome (see Mas and Moretti (2009), Cornelissen et al. (2017)). However, the important characteristic of the context studied here, which explains our result is that workers are paid fixed wages. Whether workers are compensated through a fixed wage scheme, or a performance-based salary or even a mixed remuneration scheme, their response in terms of effort usually changes. For example, Fortin et al. (2010) find that, among physicians, those paid under a fee-for-service scheme have in average a higher productivity (they provide more services) than those paid

1. In general, co-workers have a way to observe the quality of healthcare provided by their peers through diverse situations. First, they may observe their peers quality directly through interactions and discussions with them. While discussing with peers, they may find out how good or bad their knowledge about various clinical situations is. Second, they may improve their "guess" on their co-worker's quality through rates of returns and readmissions of patients after being treated. Specifically, concerning low quality co-workers, workers may also observe cases of clinical errors, medical negligence, or medical malpractice. Moreover, workers may sometimes observe peers' effort through their rates of absenteeism or how focused they are at work.

2. More precisely, when individuals see an increase in their peers productivity (which results from an increase in their effort), they tend to reduce their effort (such that the quality produced does not decrease below a certain level specific to each HF). Likewise, a decrease in peers' effort, means a decrease in their productivity (quality), which leads some workers to compensate by producing more effort.

under a mixed remuneration scheme. In their paper which focuses on physicians in Quebec (Canada), the mixed remuneration scheme refers to a mix between a fixed wage and a fee-for-services. In addition, according to Ake and Colin (2012), although simple to administer, the salary or fixed wage scheme encourages shirking through less time and effort per patient. Moreover, in a brief literature review on physicians productivity, Fortin et al. (2008) highlight that physicians paid under fixed salaries scheme tend to reduce their effort and productivity in comparison with physicians paid under the fee-for-service scheme. In a more recent paper, Flory et al. (2016) also find that in relative performance pay contracts, workers tend to increase their productivity in comparison with fixed wage schemes. The fixed wage scheme, principally used for MCH workers in our context, may also result in shirking at work. [3]

Second, the paper contributes to the literature on social interactions, by proposing a method accounting for variations in the strength of interactions among peers, when the *real* strength of interaction is unknown. In the paper, we use information on MCH workers' schedules to propose an intuitive way of measuring peer weights, in the interaction matrix. Specifically, knowing the adjacency matrix, we use the probability of interactions among peers inside the same HF, to build an *expected* social interaction matrix. [4] That is the matrix used for our empirical estimations on the SAR model.

Third, the paper contributes to the limited literature on effort at work, and quality of healthcare by maternal and child health workers, in the job environments described above. Besides investigating peer effects in that context, we also explore some individual characteristics which influence workers' effort. For instance, we find that beyond eighteen years of tenure in a HF, each additional year of tenure have a negative effect on workers' effort to provide a good quality of healthcare to their patients.

Finally, we propose a few policy recommendations particular to the complex job environment studied. These recommendations follow some specific microeconomic foundations underlying the SAR model we estimate. In fact, among the recommendations often proposed to improve the quality of maternal healthcare in developing countries, little to none of them are based on theoretical microeconomics models of workers' quality or effort. Instead, several research on the domain, often emphasize the importance of increasing the numbers of health workers to

3. Our context is specific to the case of Benin. In a 2009 report on the characteristics of the salaries of healthcare professionals from Benin, Codo and Agueh Onambele (2009) give more information and statistics on the structure of healthcare professionals salaries that was in place until year 2008 in Benin. This structure was still unchanged during years 2010–2011 when the data we use were collected.

4. In a network with groups' structure, when real interaction weights are unknown (to the econometrician), several authors often use an equal ratio $\frac{1}{n_g-1}$, where n_g is the cardinal of each group g, as a standard measure of peers' weights. See Lee (2007), Bramoullé et al. (2009), Boucher et al. (2014) among other papers. However, in this paper, although the peers' weights are not known either (to the econometrician), the data gives information on the workers schedules. This refinement allows us to propose measure of the expected interaction weights which would be more rational than a simple equal-weight assignment. For more details, go to section 1.5.3.

reduce maternal morbidity. Yet, if the effort of some skilled birth attendants [5] Is affected by their peers' effort, only increasing the numbers of birth attendants without any consideration of their skills and individual characteristics, may lead to none or less improvement of the quality of maternal and child healthcare. Hence, instead of investing bigger amounts (of money) to increase the numbers of birth attendants, it is better for public decision makers to understand the underlying motivations which influence workers in providing higher effort for higher healthcare quality. Besides, the offer of quality healthcare workers, is inelastic in general.

We use the World Bank's data on health workers and health facilities in Benin. [6] These data are part of the baseline study on the Result-Based Financing (RBF) program in Benin's health sector. For peer effects identification purpose, we limited the data to samples composed of one, three and more MCH workers. [7] Therefore, the resulting data gather information on 386 MCH workers from 128 HFs over the initial 250 HFs represented in the initial data. Furthermore, around 93% of the HFs are public, and 95% of MCH workers are women.

The remaining of the paper is organized as follows. Section 1.4 explains the context in which this study takes place and why we are interested in a developing country. Section 1.5 develops the model, its microeconomic foundations and the estimation method. Section 1.6 describes the data and section 1.7 focuses on empirical developments, results and discussions. Finally, section 1.8 concludes.

1.4 Context of the Study

In this study, we address a context, in which MCH workers are subject to fixed wages, independent from their performance or effort at work. Specifically, the data is composed of a majority of women, who are principally from the public sector. Specifically, in the data we use, approximately 92% of MCH workers work in the public sector, about 3% in the private sector, and the remaining 5% work either in semi-public, religious health centers or NGOs. Besides, more than 40% MCH workers in the data have unpaid wages and bonuses, and most of them (95%) feel wronged and unsatisfied with their salary. This is a pretty common situation in the public sector of several developing countries, particularly in sub-Saharan Africa.

Furthermore, health workers in developing countries are scarce and insufficient to provide a good coverage of the healthcare needs of their population. Therefore, they are very rarely fired

5. In the literature, the term skilled birth attendants is often used to determine health workers who have appropriate training on maternal care. They often do not even consider the fact that, some workers may have appropriate training (e.g., midwives), but still provide a low quality of healthcare.

6. Data were collected by the World Bank from December 2010 to 2011. However, they form a unique database.

7. Prior research using SAR-like models as Davezies et al. (2006) for instance give detailed explanation on the groups size requirements for the identification of peer effects.

from their job. As Chaudhury et al. (2006) describe it in their paper on public education and health workers in some developing regions, in the public health sector of Benin, the system of workers hiring, or promotions also is mainly based on workers educational qualifications and seniority. Another major, but not much documented, aspect in hiring, and promotions on Benin job market in general and the public sector in particular is the importance of social capital (i.e., personal connections) and patronage, instead of performance.[8] Additionally, there is an imperfect monitoring of workers' effort at work. Also, during the period targeted in our data (2010–2011), workers' remuneration still follow a fixed-wage scheme in Benin.[9] Such practices must, undoubtedly, affect workers effort and the quality of healthcare they provide.[10]

In the seminal paper by Shapiro and Stigltitz (1984), a necessary condition to prevent shirking of workers in their model is to allow positive unemployment. Allowing unemployment in their model, authors state that paying the efficient wage to workers can prevent them from shirking. On the opposite side, they state, and it is obvious, that in a model where there is no unemployment and imperfect monitoring, workers can choose to shirk. This is due to the quasi-null penalty of shirking in the absence of unemployment. Intuitively if there is no unemployment in the model, even if a worker is caught shirking, they can automatically be rehired. We borrow this intuition from Shapiro and Stigltitz (1984) and state that, even in a model where employer cannot fire nor give severe sanctions to a shirking worker as they wish (due to workers penury), workers can also choose to shirk. Additionally, if they are not timely or well paid, the motivation of workers to shirk is even greater. Our motivation to address the matter of MCH workers' performance in developing countries also resides in the disturbing statistics on maternal and child health in those regions.

Developing countries experience the highest rates of maternal and neonatal deaths. According to WHO, UNICEF, UNFPA and the World Bank's 2010 estimates of maternal deaths, a range of $230,000$ to $398,000$ maternal deaths occurred in 2010, which correspond to a rate of 170 to 300 deaths per $10,0000$ births.[11] 99% of these deaths have occurred in developing countries. Furthermore, while the lifetime risk of maternal deaths in developed regions in 2010 was 1 in $3,800$, the rate in sub-Saharan Africa was as high as 1 death in 39 births. That is almost one hundred times higher than the rate in developed regions, the highest rate of all regions

8. Although no official statistics on this practice in Benin exist, it remains an important issue as it is in other countries in sub-Saharan Africa. Lewis (2006) gives few examples of corruption, patronage and what they called the "purchasing of public positions" in some countries including countries like Ghana, Uganda and Ethiopia in Africa, and others like Dominican Republic outside Africa.

9. Workers could also benefit from bonuses. However, none of their remuneration was performance-based (for more details see Codo and Agueh Onambele (2009)).

10. For instance, concerning patronage, although their results do not specifically target the health sector, Colonnelli et al. (2018) address some negative impact of patronage in the public sector on the quality of the services provided.

11. See WHO (2012).

worldwide. In Benin specifically, the lifetime risk of maternal deaths in 2010 was of 1 death in 59 births. These statistics show the importance of addressing a subject related to maternal and child health in the context of a developing country.

A key strategy promoted by the millennium development goals (MDG) to reduce child and maternal mortality and morbidity and improve maternal and child health is to increase access to skilled care during pregnancy and delivery. In fact, the link between maternal deaths and skilled birth attendants no longer needs to be demonstrated.[12] Wagstaff and Claeson (2004) estimated that the full use of existing health interventions (at 99% coverage) could avert 63% of child deaths and 74% of maternal deaths. Most studies linking maternal mortality and morbidity to skilled birth attendants (principally in journals specialized in obstetrics and/or medicine) mainly focus on the importance of increasing the number of skilled birth attendants to lower the rates of maternal mortality and morbidity (see Carlough and McCall (2005)).

Yet, increasing the number of skilled birth attendants in developing countries is already a difficult goal to achieve because of the very small elasticity of the health labor supply among other reasons. A significant investment in money (which most of those countries do not have) is required as well as enough preparation time to build more facilities and to increase healthcare staff through education. Meanwhile, thousands of maternal and child deaths continue to occur. To address these issues, it would be important that public decision makers find more feasible ways (in the short run) of improving the condition of maternal and child health in these countries. In that line, another objective in this paper is to propose attainable short-term policies on how to improve the quality of healthcare based on healthcare staff's productivity. We do it by exploring how combinations (of productivity and individual characteristics) of healthcare staff per HF explain the healthcare quality provided in HFs.

1.5 Model

In this section, our model, which accounts for peer influence, is explained. The section is broken down in four major parts. First, the underlying microeconomic foundations of our model are detailed in Section 1.5.1, then we describe our setup in Section 1.5.3.[13] Next, Section 1.5.4 describes some useful adjustments made for the empirical estimations on the model, and finally Section 1.5.5 explains our approach in estimating its parameters.

12. In WHO (2002) the term *skilled attendant* refers exclusively to "people with midwifery skills (for example midwives, doctors and nurses) who have been trained to proficiency in the skills necessary to manage normal deliveries and diagnose, manage or refer obstetric complications."

13. Calvo-Armengol et al. (2009), Blume et al. (2013), and Boucher and Fortin (2015) use a similar approach to develop the economic foundations of their models.

1.5.1 Microfoundations

The goal of this section is to explain the main motivations behind the chosen level of effort of health workers at work, and to establish its link with the objective of their employer (HF).[14] Prior studies have highlighted the fact that workers' effort, depends on several aspects; some of which are related to their working environment, and/or their characteristics.[15] In this paper, we consider a case in which workers' effort influence the quality of healthcare they provide to their patients.

Specifically, let e_i and a_i denote respectively the level of effort provided by an MCH worker i and their ability (to provide adequate care to patients). We note $q_i(e_i, a_i)$ the quality of healthcare produced by a worker i. For now, we leave unspecified the functional form of $q_i(e_i, a_i)$. We come back on this later. Let g index the HF where i works. There are n_g MCH workers working together in g. Then, the overall quality of healthcare provided by all workers in g, noted Q_g, is a quantity depending on two major inputs: the capital (K_g) and the labor (L_g).[16] For simplicity, we take its functional form as a simple Cobb-Douglas production function as follows:

$$Q_g = Q(q_1, ..., q_{n_g}, K_g) = AK_g^{r_0} \prod_{i=1}^{n_g} q_i^{r_i} \tag{1.1}$$

The parameters $r_0, ..., r_{n_g}$ correspond to the output elasticity of each input. That is, the local measure of percentage response of the overall quality of care produced, to a 1% change in capital input, and individual labor inputs (quality of healthcare by workers). Concretely, we could think of a situation where the HF attributes the same r_i to workers of the same category (doctors, midwives, nurses, nursing auxiliaries), so that $r_i \in \{R_d, R_m, R_n, R_{aux}\}$. So that, the quality of healthcare in an HF, would increase more with an increase in a doctor's quality, than it would with an increase in a nurse's quality (or vice versa). We arbitrarily impose constant return to scale, that is $\sum_i r_i = 1$. The form (1.1) imposes constant elasticity of substitution (CES) between inputs, but in reality it is of very little importance in our developments. In

14. For small maternities or dispensaries, the employer is the HF. However, for bigger HF with several healthcare departments, the department of maternal and child care is the main employer. Therefore, in the remaining of this paper, the term health worker, or worker, refers only to maternal and child health (MCH) workers.

15. The literature on workers' effort is various and rich. Among those studies, workers' effort or performance is sometimes linked to their working conditions and satisfaction, or at other times to their remuneration. See for instance Fortin et al. (2008),Fortin et al. (2010), (Ake and Colin, 2012), for studies on effort and remuneration.

16. Here, because of the cognitive nature of MCH workers tasks, the labor L_g refers principally to q_i, $i = 1, ..., n_g$. To keep things simple, the term capital (K_g) here, refers to the physical capital, as well as any other HF-related resources useful in the production of a *quality* healthcare. These may include for instance, infrastructure (their availability and condition), any functional equipment and materials (beds, delivery tables, etc.), clinical equipment (stethoscope, sphygmomanometer, forceps, etc.), other resources (water supply, electricity, government subsidies, etc.).

fact, we consider equation (1.1) only later, while discussing the mechanisms, in the section 1.7 on empirical results. In reality, our results hold generally, even without specifying the form of the HF's production function $Q(.)$ and the values of parameters $r_0, ..., r_{n_g}$.

Within the framework of this paper, we impose the following properties on the production function on $q_i(.)$:

Assumption 1.1. *$q_i(e_i, a_i)$ is twice differentiable on $\mathbb{R}_+ \times \mathbb{R}_+^*$ such that:*
(i) $\frac{\partial q_i()}{\partial e_i} > 0$, $\frac{\partial q_i()}{\partial a_i} > 0$ *; and*
(ii) $\frac{\partial^2 q_i()}{\partial e_i^2} \leq 0$*;*
(iii) $q_i(0, a_i) = 0, \ \forall a_i > 0$

Assumption 1.1*(i)* means that all else being equal, the quality of maternal and child care produced increases with workers' effort, and their ability. However, through the assumption 1.1*(ii)*, this increase stays constant or becomes less and less important (in magnitude), for higher levels of effort. Therefore, the individual's production function $q_i(.)$ is concave, thus ensuring that the marginal product of effort is constant or decreasing. In fact, at first, a very little increase in effort may produce a high increase in workers' quality, but from a certain moment forward, making very high effort, may sometimes lead to little improvements in quality.[17] Finally, the last assumption means that, with no effort, a worker's perceived quality is null from their employer's point of view.[18]

Choice of Effort

Up to now, we have explained how workers' quality is related to their whole HF's quality. However, workers do not directly choose the quality of the healthcare they provide. Instead, it is their chosen level of effort at work that determines their quality. In what follows, we discuss workers' preferences for effort at work, and how it is related to peers' effort. Consider an HF g where worker i interacts with $n_g - 1$ other co-workers. Let $\boldsymbol{e} = (e_1, ..., e_i, ..., e_{n_g})'$ a vector indicating the level of effort provided by the workers, $\mathbf{X} = (\mathbf{x}_1', ..., \mathbf{x}_i', ..., \mathbf{x}_{n_g}')'$ a matrix giving the workers' socio-economic characteristics (observable to anyone), and $\boldsymbol{\eta} =$

17. To justify the case of a decreasing marginal product, consider for example a doctor who, at first, provides little effort at work. For instance, they may provide healthcare for few patients, adopt little or no updating of their knowledge in clinical care. Surely, if they start providing healthcare to more patients, they may (sometimes) face more complicated clinical issues, which may lead them to read more on those cases, and in turn, them may improve radically their quality. However, for a doctor who is already making high effort, and consequently provides already a high quality, they may not be able to increase their quality as much as the previous doctor, while increasing their effort. This argument holds for several clinical healthcare issues. For example, a doctor is unable to cure cancer, only by making more effort through research, or caring for more cancer patients. In fact, they may never find a cure, although they increase substantially their effort.

18. The case $e_i = 0$ is extreme and quasi-impossible in reality for health workers. Yet, assuming it may happen, if a midwife never provides care for patients, nor makes any effort at work, the part of their quality in the HF's overall quality is null.

$(\eta_1, ..., \eta_{n_g})'$ a vector giving the workers' characteristics that are private information.[19] Later, for econometric purpose, we posit independence between these characteristics: $\mathbb{E}(\eta_i/\mathbf{x}_i) = 0, \; \forall \; i$.

A worker i has preferences over effort, represented by a utility function u_i. i chooses their level of effort $e_i \in \mathbb{R}_+$ to maximize u_i. u_i depends on i's chosen level of effort, some of their observable and private characteristics, and some characteristics of their workplace.[20] In particular, this utility function u_i is a sum of a private utility u_{pi}, which does not account for peers' externalities, and a social utility u_{si}. It is then defined as:

$$u_i(e_i, \boldsymbol{e_{-i}}) = \underbrace{(f_g + \mathbf{x}_i\upsilon + \eta_i)\, e_i - \frac{e_i^2}{2}}_{u_{pi}(e_i)} + \underbrace{e_i \mathbf{W}_i \mathbf{x}\tau + \beta e_i \mathbf{W}_i \boldsymbol{e}}_{u_{si}(ei, \boldsymbol{e_{-i}})} \tag{1.2}$$

$\boldsymbol{e_{-i}} = (e_1, ..., e_{i-1}, e_{i+1}, ..., e_n)$ represents the levels of effort chosen by workers other than i, and $\mathbf{W}_i$ is the i^{th} line of the $n \times n$ matrix $\mathbf{W}$ composed of all weights w_{ij} (strengths of interaction).[21] We can easily verify that the utility function $u_i()$ is continuous and strictly concave in e_i. In equation (1.2), f_g is the same for each individual in a same HF (or group) g but can vary from one group to another.[22] It represents the marginal utility of effort resulting from the characteristics of the health facility g.[23] $\mathbf{x}_i$ is a $k-$vector of some k observable characteristics of individual i, while η_i represents unobservable characteristics. Both $\mathbf{x}_i$ and η_i influence individual i's utility. At this stage we do not make any i.i.d. or constant variance assumption on the η_is. Further in this work, we will give more details on this decision. The form $\frac{e_i^2}{2}$ represents the cost that individual i bears when they make effort e_i (disutility of effort). In the absence of any peer, increasing effort is rewarding for the worker as long as it does not exceed a certain level $\bar{E}_{g0i}$.[24]

19. The characteristics $\mathbf{X}$ may include, for instance, their education, experience, titles or responsibilities in the HF, hours worked, any information related to their remuneration, or how much they feel recognition/consideration from their employer and peers. $\boldsymbol{\eta}$ on the other hand may represent private characteristics which influence their chosen level of effort, but are known by themselves only. Here, we consider $\boldsymbol{\eta}$ as a vector which aggregates all private characteristics of the worker which influence their marginal utility of effort.

20. In equation 1.1 above, the labor is not totally independent of the technology and material present in the HF. In fact, if we consider that the service produced is the quality of care, its interdependence with the job environment and available technology is an underlying hypothesis.

21. Recall that $w_{ii} = 0$ and $w_{ij} = 0$ if individual j is not in the same group as individual i, so that $\mathbf{W}_i\boldsymbol{e}$ and $\mathbf{W}_i\mathbf{x}$ give respectively a weighted average level of effort and a $k \times 1$ vector of (weighted) average characteristics of individual i's peers only.

22. f_g accounts for the characteristics of the health facility that affect everyone in the HF the same. It could be for instance its capital K_g or the type of HF (dispensary, maternity, private clinic, confessional hospital, hospital, etc.).

23. Thus, a health facility with higher standards in infrastructure or material endowments (higher f_g) could induce a higher motivation for its MCH workers. That could lead to an increase of the utility of such workers, in comparison with workers with the exact same individual characteristics in other less endowed HF.

24. As we consider the partial derivative $\frac{\partial u_{pi}}{\partial e_i}$, and given f_g, ν, $\mathbf{x}_i$, and η_i, we derive $\bar{E}_{g0i} = f_g + \mathbf{x}_i\nu + \eta_i$

As for the *social* component u_{si} of the utility, it is a sum of social components related to each peer in their HF (which can be seen as externalities). Let $u_{si,j}$ denote the social utility that individual i gains from interacting with their peer j. Then, we have:

$$u_{si}(e_i, \boldsymbol{e_{-i}}) = \sum_{\substack{j=1 \\ j \neq i}}^{n_g} u_{si,j}(e_i, e_j) \tag{1.3}$$

If i interacts with j, then they could receive up to $\bar{u}_{si,j}(e_i, e_j) = e_i \mathbf{x}_j \tau + \beta e_i e_j$. The component $\bar{u}_{si,j}(e_i, e_j)$ is composed of an exogenous component (the individual characteristics of peer j, $\mathbf{x}_j$) and an endogenous component (accounting for the effect that the chosen level of effort of j has on i's own level of effort). To account for some variability in the strength of interaction between peers coming from their frequency of interactions with peers at work, we affect a weight w_{ij} to each peer's social component.[25] Consequently, it is given by:

$$u_{si,j}(e_i, e_j) = w_{ij} e_i \left(\mathbf{x}_j \tau + \beta e_j\right) \tag{1.4}$$

Through the weight w_{ij}, equation (1.4) ensures that peers with whom an individual interacts more also have a bigger impact on their marginal social utility of effort. It also guarantees that, when interactions are stronger, workers are more sensitive to their peers individual characteristics or level of effort. In particular, equation (1.4) captures the fact that, although a worker may interact with peers who have the same individual characteristics and effort levels, these peers can impact them differently, depending on the strength of their interactions with them.

The optimal level of effort e_i^* exerted by each MCH worker i at equilibrium is obtained by resolving the following optimization problem for each MCH worker:

$$\underset{e_i}{max} \;\; u_i(e_i, \boldsymbol{e_{-i}}) \tag{1.5}$$

Proposition 1.1. *Within any health facility g, given $|\beta| < \dfrac{1}{\| \boldsymbol{W} \|}$, a Nash equilibrium for problem (1.5) exists and the best response function of effort of a MCH worker i, is uniquely*

25. Note that without introducing the weight w_{ij}, the marginal utility of effort that i receives from interacting with each peer is the same for peers with identical characteristics. However, we need to make a distinction between the magnitude of the peer effect when i spends more time with a colleague versus when they spend less time with another one. Yet, given β constant, we can only account for such a distinction by considering various weights w_{ij}. We, as econometricians, do not observe the real frequency of interactions among peers. Therefore, we use a measure of strength based on workers' probabilities of interactions with peers as mentioned later in equation (1.13). This gives an expectation of the strengths of interactions used to define an expected social interaction matrix.

given by:

$$e_i^* = f_g + \beta W_i \boldsymbol{e}^* + \boldsymbol{x}_i \upsilon + W_i \boldsymbol{x} \tau + \eta_i \tag{1.6}$$

The vector of effort level at Nash equilibrium is thus given by $\boldsymbol{e}^* = (e_1^*, ..., e_n^*)$.

Proof. See appendix A.1. □

Remark 1.1. *We notice that* $\frac{\partial^2 u_i}{\partial e_j \partial e_i} = \beta w_{ij}$ *takes the sign of* β*, which is the endogenous peer effect coefficient. Consequently, if* $\beta > 0$ *levels of effort of* i *and* j *are complements; otherwise if* $\beta < 0$ *they are substitutes.*

Worker's Resulting Quality

Given the equilibrium choice of effort e_i^* each worker i makes, their associated quality of healthcare observed by the employer and their peers is $q_i^* = q_i(e_i^*, a_i)$. For convenience, we propose the linear production function below, much easier to handle:

$$q_i(e_i, a_i) = \nu e_i + \psi a_i, \quad \psi, \nu > 0 \tag{1.7}$$

This is the simplest possible form of $q_i()$. It also ensures constant returns to scale. In addition, it allows us to make easier interpretations while discussing the mechanisms in section 1.7. [26]

1.5.2 A Proxy for Effort

The effort exerted by an MCH worker is not directly measurable. Therefore, we use instead an indicator of workers' skills and clinical knowledge as a proxy for effort, and denoted y_i. This proxy variable, y_i, is derived from a score obtained by MCH workers during a test of skills and knowledge in maternal and child care. [27] y_i equals the natural logarithm of the proportion of right response options mentioned by the MCH worker on the test. [28] Then, we consider:

$$y_i = \nu e_i^* + \psi a_i + \mu_i \tag{1.8}$$

26. The results obtained in the empirical section do not change whatever the form of $q_i()$. In fact, one can choose for instance a Cobb-Douglas form $q_i(e_i, a_i) = a_i^{\psi} e_i^{\nu}$ or any other form of production function to respect assumption 1.1*(iii)*. However, in that case, one must be careful while interpreting the main parameters in equation (1.6), in relation with the quality of healthcare. Also, note that equation (1.7) can be seen as a log-linearized form of $a_i^{\psi} e_i^{\nu}$ when $a_i,\ e_i > 0$.

27. As it is the case for any approximated measure, we acknowledge the potential existence of a measurement error issue often due to the use of a proxy variable. Although this potential existence of measurement error in the dependant variable mainly affects the standard errors, their existence in the explanatory variables usually induce more severe issues in regressions such as attenuation bias for instance. However, using IV methods with good independent instruments uncorrelated with the error terms often mitigate the problem.

28. We give the intuition and details about the computation of the proxy variable y_i and additional details on the raw test score used in its computation, and henceforth noted $MCHWscore_i$, the empirical measure of y_i and its interpretation in the section 1.7.1.

where μ_i is a random term assumed to be i.i.d. with zero mean, within each group, and variance σ^2_μ.[29] In this equation, ν and ψ are assumed to be positive because, all else being equal, an increase in the level of effort, or in the ability of individual i are likely to have a positive effect on their real skills and knowledge. Note that ability is not directly measurable either. However, we assume that it will be accounted for through individual characteristics, such as the level of education, or the position in the HF (doctor, midwife, nurse or nursing auxiliary). In fact, we expect that the differences in position of MCH workers within an HF also capture a part of the differences in their ability. Specifically, a doctor or a midwife is expected, on average, to show higher ability than a simple nurse or a nursing auxiliary. Similarly, we expected on average that a nurse has higher ability than a nursing auxiliary.

We use equation (1.8) for each individual $j \neq i$ in the same group as i, and we assume that each parameter or variable, but e_i^*, is known for all i. Consequently, provided $\nu \neq 0$, we deduce that each e_j^* should respect the mathematical form:

$$e_j^* = \frac{1}{\nu} y_j - \frac{\psi}{\nu} a_j - \frac{1}{\nu}\mu_i \tag{1.9}$$

We cannot have $\nu = 0$ because it would otherwise mean that there is no relationship whatsoever between y_i and e_i^*, which contradicts the main idea that y_i is used as proxy for e_i^*. In principle, if we interpret ν as the correlation between the two, it should take a value closer to 1.[30]

As we replace equation (1.9) in (1.6), we obtain another analytical form of e_i^* which depends on y_js instead of unobservable e_j^*s.

$$e_i^* = f_g + \frac{\beta}{\nu}\mathbf{W}_i\mathbf{y} + \mathbf{x}_i\upsilon + \mathbf{W}\mathbf{x}\tau - \frac{\beta\psi}{\nu}\mathbf{W}_i\mathbf{a} - \frac{\beta}{\nu}\mathbf{W}_i\boldsymbol{\mu} + \eta_i \tag{1.10}$$

Then, we replace (1.10) in equation (1.8), and we obtain the following:

$$y_i = \nu f_g + \beta\mathbf{W}_i\mathbf{y} + \mathbf{x}_i\upsilon\nu + \psi a_i + \mathbf{W}_i\mathbf{x}\tau\nu - \beta\psi\mathbf{W}_i\mathbf{a} - \beta\mathbf{W}_i\boldsymbol{\mu} + \mu_i + \eta_i\nu \tag{1.11}$$

As we rewrite $\tilde{\mathbf{x}} = (\mathbf{x}, \mathbf{a})$, the final form of the model which can be empirically estimated is then:

$$y_i = \alpha_g + \beta\mathbf{W}_i\mathbf{y} + \tilde{\mathbf{x}}\gamma + \mathbf{W}_i\tilde{\mathbf{x}}\delta + \epsilon_i \tag{1.12}$$

29. The proxy y_i as defined here is clearly a measure of workers' quality. Then, replacing equation (1.7) in (1.8), we obtain $y_i = q_i^* + \mu_i$.

30. Although ν is not really a correlation coefficient in the present case, it can tell us about the correlation between y_i and e_i^* depending on whether it is close to 0 or 1. However, it must not be near 0 in this case.

where $\alpha_g = \nu f_g$, $\gamma = (\upsilon'\nu, \psi)'$, $\delta = (\tau'\nu, -\beta\psi)'$ and $\epsilon_i = \mu_i + \eta_i\nu - \beta\mathbf{W}_i\boldsymbol{\mu}$.
$\boldsymbol{\epsilon}$ captures unobservable correlated effects between individuals and their peers of the same group.

1.5.3 Setup of the interaction matrix

Let $N = \{1, .., n\}$ a set of n MCH workers, and $\mathbf{A} = (a_{ij})$ an $n \times n$ adjacency matrix which indicates the presence or absence of interactions between workers across all HFs. Each component a_{ij} of $\mathbf{A}$ takes value 1 if individuals i and j $(i \neq j)$ are in the same group (HF) and 0 otherwise. Hence, $\mathbf{A}$ is block diagonal and symmetric. More specifically, we take all diagonal elements as null $(a_{ii} = 0)$, meaning that an individual is not considered to interact with (or influence) themselves. In reality, peers assigned to a same HF may never interact in the case their working schedules do not overlap. And even if they do overlap, the intensity of their interactions may vary according to these schedules. Let $\mathbf{T}$ designate the matrix whose components t_{ij} are random variables taking value 1 if i and j $(i \neq j)$ interact at least once a week and 0 otherwise. [31] $\mathbf{T}$ is the *real* adjacency matrix among all workers and may be distinct from $\mathbf{A}$. Then, the social interaction matrix to use should be $\tilde{\mathbf{T}}$ such that $\tilde{t}_{ij} = 0$ *iff* $t_{ij} = 0$, and $\tilde{t}_{ij}(\neq 0)$ gives the strength of interactions needed in equation (1.4). However, given that we neither observe $\mathbf{T}$ nor $\tilde{\mathbf{T}}$, but $\mathbf{A}$ and some other variables giving information on workers schedules, we present next the *matrix of expected interactions* $\mathbb{E}(\mathbf{T})$ used to approximate $\tilde{\mathbf{T}}$. [32]

A matrix of expected interactions

In the data, we only observe information about their average daily working hours and the number of days worked during a week. This information allows us to compute $\mathbf{W} = \mathbb{E}(\mathbf{T})$. By definition, $\mathbb{E}(t_{ij}) = \mathbb{P}(t_{ij} = 1)$. [33] To determine $\mathbf{W} = (w_{ij})$, consider the following two events:

T^1_{ij}: i and j work together (interact) at the same moment in a day;

T^2_{ij}: i and j work the same day at least once a week

Then, we posit $\mathbb{P}(t_{ij} = 1) = \mathbb{P}(T^1_{ij} \cap T^2_{ij})$, and using Kolmogorov definition of conditional probability, we have:

$$\mathbb{P}(t_{ij} = 1) = \mathbb{P}(T^1_{ij} | T^2_{ij})\mathbb{P}(T^2_{ij})$$

To simplify the notations, we further denote $\mathbb{P}(T^2_{ij}) = p_{ij}$. Its detailed formula is given later in equation (1.14). Conditional on working the same day at least once a week, we compute the probability of two peers interacting at the same moment as the ratio $\dfrac{h_{ij}}{\sum_k h_{ik}}$.
h_{ij} is the duration of interaction between i and j during a given day. Then, w_{ij} is computable

31. $\mathbf{T}$ is a refinement of $\mathbf{A}$, such that $t_{ii} = 0$ as well.
32. $\mathbb{E}$ represents the symbol for the mathematical expectation.
33. $\mathbb{E}(t_{ij}) = 1 \times \mathbb{P}(t_{ij} = 1) + 0 \times \mathbb{P}(t_{ij} = 0)$, where $\mathbb{P}$ represents the symbol for statistical probability.

such that:

$$w_{ij} = p_{ij} \frac{h_{ij}}{\sum_k h_{ik}} \tag{1.13}$$

Consequently, $w_{ij} \in [0, 1]$ for all couple (i, j) of individuals and, *in general*, $\sum_j w_{ij} \neq 1$ and $w_{ij} \neq w_{ji}$.[34] Hence, $\mathbf{W}$ is not row-normalized, and not necessarily symmetric.[35] Overall, defined as in equation (1.13), w_{ij} is the expected average strength of interactions between i and j we use in equation (1.4).

$\frac{h_{ij}}{\sum_k h_{ik}}$ also quantifies the average daily strength of interactions between two co-workers i and j (provided that they work on the same day). This fraction gives the ratio of the time spent between two co-workers comparatively to the sum of time spent with all co-workers of the same HF. Therefore, it implicitly allows us to make a clear distinction between the strength of interactions between two individuals interacting in a small HF, and those interacting in a larger one. This is straightforward. For instance, consider that a few individuals, let's say five, interact during a same given period within their HF, versus many more individuals (say a hundred) interacting during the same period in another HF. Then, those in the small group are likely to have a greater influence on each other, than those in the bigger group.

In most of the empirical studies in the networks and social interaction literature, authors usually construct a *binary* (adjacency) matrix whose intuition is similar to that of the *contiguity matrix* used in the spatial econometric literature. The binary matrix is then row-normalized, to obtain a matrix which accounts for average peer weights within the network.[36] The usual measure of peer influence taken in these studies, uses the degree of each node i (deg_i); which is simply the number of peers an individual interacts with in the network. Explicitly, for each line i of the adjacency matrix, each positive value is replaced by the fraction $\frac{1}{deg_i}$ (average peer weight). This in turn results in a social interaction matrix with identical values within a same group. While estimating peer effects in such models, the value obtained for the coefficient measuring peer effects, which is an *average* endogenous effect, does not help distinguish the *real* individual effect of each peer. In contrast with models using same peer weights and measuring average peer effects, other studies advocate for considering that each peer usually does not have the same influence over individuals; which can help in improving

34. Note that $w_{ij} = 0$ *iff* i and j are not in the same HF. Otherwise, even if they work only once a week, the probability of them interacting at least once a week is always greater than 0. Additionally, given that for any set of k co-workers $\{j_1, j_2, ..., j_k\}$ from any HF, events $\{t_{ij_1} = 1\}, \{t_{ij_2} = 1\}, ..., \{t_{ij_k} = 1\}$ are not independent, we cannot always have $\sum_j \mathbb{P}(t_{ij} = 1) = 1$.

35. In fact, $\mathbf{W}$ is symmetric *iff* $\sum_k h_{ik} = \sum_k h_{jk} \ \forall \ i, j$; which is not always the case. These quantities are the same *iff* i and j have the same total duration of interactions across peers. A trivial case is when both workers serve 24 hours, 7 days a week. Which means they live together in the HF. However, it is not the case generally across HFs.

36. See for instance, Davezies et al. (2006), Lee (2007),Bramoullé et al. (2009), and Boucher et al. (2014).

the information on the real effect of each peer. Several methods are used in the literature to differentiate peer effects within a network. For instance, Calvo-Armengol et al. (2009), use the Katz-Bonacich centrality to account for the differences in peer influence.[37] They interpret these differences as the strength of the friendship or ties between individuals.[38] On the other hand, Patacchini et al. (2017) and Dieye and Fortin (2017) propose some specifications of heterogeneous peer effects which account for possible dissimilarities within groups, based on specific characteristics.[39]

Simply put, an interaction matrix **W** with distinct weights among peers, is more intuitive, given that individuals usually do not attribute an identical value to the actions of each of their peers. Besides, even in group settings, symmetry is a very strong assumption. In general, individuals in a group do not necessarily reciprocate the same subjective value to one another. In like manner, our model is founded on the claim that the *relative time* spent with a co-worker, which is the time spent with that co-worker at work over the total time of interactions with all co-workers from the same HF, is appropriate to quantify this subjective value. Essentially, our model relies on the following assumptions.

Assumption 1.2. *Heterogeneous peer weights:*

(i) *Ceteris paribus*, people who spend more time together are more likely to influence each other than people who spend less time together.[40]

(ii) *Ceteris paribus*, the more co-workers individuals share time with simultaneously in an HF, the less these co-workers influence them individually.[41]

Computed Probability of Interactions p_{ij}

We compute the probability p_{ij} by using some basic notions in enumeration. We observe the number of days worked per week and the hours worked during a typical working day.[42]

37. In our paper, as opposite to Calvo-Armengol et al. (2009), individuals interact in groups. Consequently, the Bonacich centrality is basically the same for every individual within the same group, but it could vary from one group to another depending on the size of the groups.

38. Distant friends often have less influence on their peers than closer friends do.

39. Patacchini et al. (2017) propose a model accounting for heterogeneous peer effects, with respect to individuals' education attainments. Dieye and Fortin (2017) on the other hand, explore gender-related heterogeneous peer effects in obesity.

40. This assumption is supported by Patacchini et al. (2017) who find in their study that, to a certain extent, peers tend to be more influential, when their friendship last longer (more than a year).

41. This assumption is mainly supported by most of the studies on peer effects until now. In fact, when some authors use $\frac{1}{deg_i}$ as peer weight in their studies, they are implicitly making this assumption. We capture this, by taking the denominator part $\sum_j h_{ij}$. We use the form $\frac{h_{ij}}{\sum_j h_{ij}}$ instead of just using the ratio $\frac{1}{n_g-1}$ (where n_g is the number of co-workers an individual of an HF g has), to account for the heterogeneity in starting and finishing hours among peers at work, as observed in the data.

42. The days worked by individuals in a week (Monday, Tuesday, etc.) are not observed in the data.

Consider two individuals i and j who work in the same HF for d_i and d_j days respectively. During a given week, they have at most $min(d_i, d_j)$ possibilities to work a same day of the week. We calculate the probability that their schedules overlap at least once (p_{ij}) in a given 7-days working week. Then, the general form of this probability is given by:

$$p_{ij} = \frac{\sum_{k=1}^{min(d_i,d_j)} C_{max(d_i,d_j)}^{k} C_{7-max(d_i,d_j)}^{min(d_i,d_j)-k}}{C_7^{min(d_i,d_j)}} \tag{1.14}$$

Note that it is also possible to use other specifications to capture the strength of interactions among peers. For instance, p_{ij} could be the probabilities of interactions between i and j at least twice, three days, four, five, six or seven days during a week.

1.5.4 More on the Model

We consider a series of G health facilities. Each HF represents a group of MCH workers. For each group $g \in \{G_1, , G_G\}$ we have a set of n_g individuals (MCH workers) interacting . We consider a set of L observable characteristics (e.g. ,number of children, level of education, experience in the domain of maternal and child care, etc.); and for each individual i of group g, their individual characteristics are noted $x_{gil}, l = 1, ..., L$. Each of these characteristics influences the preference of the individual to increase or decrease their level of effort, and consequently to acquire more knowledge or less. We assume that all individuals have the similar utility u_i specified in equation (1.2). $\mathbf{W}$ still denotes the MCH workers *weighted* interaction matrix and w_{ij} its elements. $\mathbf{W}$ is a block diagonal matrix (not symmetric) defined as $diag\left(\mathbf{W}_{G_1}, \mathbf{W}_{G_2}, ..., \mathbf{W}_{G_G}\right)$ where each $\mathbf{W}_g$ is an $n_g \times n_g$ matrix. The components of $\mathbf{W}$ are defined as in equation (1.13) as following:

$$w_{ij} = \begin{cases} p_{ij} \frac{h_{ij}}{\sum_{j=1}^{n_g} h_{ij}} & \text{if } i, j \in g \\ 0 & \text{if } j = i \text{ or } \ \ i \in g \ and \ j \notin g \end{cases}$$

Let y_{gi} be the skills and knowledge score of an MCH worker i of group g and $\mathbf{x}_{gi}$ their $1 \times l$ vector of exogenous characteristics. $\mathbf{X}_{ng}$ denotes the $n_g \times l$ matrix of exogenous characteristics of all MCH workers in group g. The structural form of our model can be rewritten as follows:

$$y_{gi} = \alpha_g + \beta \sum_{j \in g} w_{ij} y_{gj} + \mathbf{x}_{gi}\gamma + \sum_{j \in g} w_{ij} \mathbf{x}_{gj} \delta + \epsilon_{gi} \tag{1.15}$$

As detailed later in section 1.7.1 in this paper, $y_{gi} = log(MCHWscore_{gi})$ and βw_{ij} measures the actual endogenous peer effect from i's peer j. It can also be interpreted as an elasticity.[43] Here, β cannot strictly be interpreted as an average peer effect because of the variability of the weights w_{ij}.

Handling unobserved peers

Our data is composed of small samples of at most five MCH workers in each group. In reality, the links in a group (or HF) should resemble the following figure 1.1. The figure shows an example of HF composed of five MCH workers among which only three are observed in the sample. The nodes 1, 2, and 3 represent the observed individuals, while the unobserved ones are 4 and 5. As we can see, many real links are not observed in the data. This means that the outcomes observed for individuals 1, 2 and 3 also reflect the effects of their unobserved peers 4 and 5. Thus, to make sure that our estimates only reflect the effects of the observed peers, a correction method should be implemented.

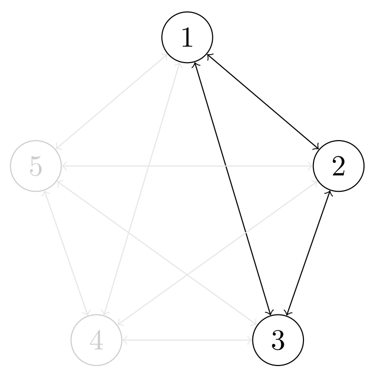


Figure 1.1 – Observed versus unobserved links in an HF

We use a correction method inspired by the one in Boucher et al. (2014) account for unobserved individuals. Their method requires that we know the real numbers n_g of MCH workers (observed and unobserved combined) in each group g. Let n'_g be the number of observed MCH workers in the sub-sample of the g^{th} group and g' the set of observed links in the g^{th} group. We denote $g" = g \setminus g'$ the set of unobserved links of the group g. For simplicity in the computations, we make the following assumption:

Assumption 1.3. *For each individual i their peers represented in the sample are a good representation of all peers in the HF population such that:*

The ratio of the sum of interaction hours with peers from the sample $\left(\sum_{j=1}^{n'_g} h_{ij}\right)$ over the sum of interaction hours with peers from the whole population in the HF $\left(\sum_{j=1}^{n_g} h_{ij}\right)$

43. $\beta w_{ij} = \frac{\partial y_{gi}}{\partial y_{gj}} = \frac{\partial MCHWscore_{gi}/MCHWscore_{gi}}{\partial MCHWscore_{gj}/MCHWscore_{gj}} = \frac{\partial q^*_{gi}/q^*_{gi}}{\partial q^*_{gj}/q^*_{gj}}$, with $q^*_{gi} = a^{\psi}_{gi} e^{*\nu}_{gi}$. βw_{ij} measures the sensibility of the healthcare quality q^*_{gi} provided by a worker i from group g to a 1% variation of a peer j's quality q^*_{gj}. All the coefficients $w_{ij}\delta_l$ can also be interpreted as the percentage point variation in q^*_{gi} due to a variation of 1 unit of the variable (individual or contextual).

is equal to the ratio of the number of observed peers over the real number of peers in the HF. This is equivalent to:

$$\frac{\sum_{j=1}^{n'_g} h_{ij}}{\sum_{j=1}^{n_g} h_{ij}} = \frac{n'_g - 1}{n_g - 1}$$

Consequently, if we write $w'_{ij} = p_{ij}\frac{h_{ij}}{\sum_{j=1}^{n'_g} h_{ij}}$, then we have $w_{ij} = \frac{n'_g-1}{n_g-1}w'_{ij}$ and the sampling-correction model gives:

$$y_{gi} = \tilde{\alpha_g} + \beta\frac{n'_g - 1}{n_g - 1}\sum_{j\in W'_g} w'_{ij}y_{gj} + \mathbf{x}_{gi}\gamma + \frac{n'_g - 1}{n_g - 1}\sum_{j\in W'_g} w'_{ij}\mathbf{x}_{gj}\delta + \epsilon_{gi} \tag{1.16}$$

where $\tilde{\alpha_g} = \alpha_g + \beta\frac{n'_g-1}{n_g-1}\sum_{j=n'_g+1}^{n_g} w'_{ij}y_{gj} + \frac{n'_g-1}{n_g-1}\sum_{j=n'_g+1}^{n_g} w'_{ij}\mathbf{x}_{gj}\delta$. $\tilde{\alpha_g}$ is the total group fixed effect that accounts for all unobserved individuals 'influence on all individuals observed in their group. The model (1.16) is the final model we estimate in this study.

1.5.5 Estimation Method

The form of the equation 1.12 reminds us of the common spatial autoregressive (SAR) model frequently used in spatial econometrics. Spatial models like the SAR model are widely used among social interactions and network economists, and several papers like Bramoullé et al. (2009), Davezies et al. (2006) address issues of identification in these types of models.

Peer effects in the literature: some econometric issues

Several issues hinder the identification of social interactions effects (endogenous and exogenous) [44]. The problem of identification which has once been one of the main issues in the studies on peer effects, spatial correlation and other simultaneous equations type models has already been well solved in the literature. Although previous studies assess the existence of social effects, they also find it very challenging to separate the effects resulting from peers' outcome (endogenous or peer effects) from those resulting from peers' exogenous characteristics (exogenous effects). [45] Manski (1993a) is one of the firsts to address the subject of identification of endogenous from exogenous effects. He uses a linear model where he considers the individual outcome/characteristics when assessing his/her group's mean outcomes/characteristics. Moffit (2001) in his model instead ignores the individual outcome/characteristics in his/her group's mean outcome/characteristics. Furthermore, he assumes a case of groups with equal

44. See Manski (1993a) for an introductory literature on endogenous, exogenous and correlated effects.
45. For instance, Manski (1993a) used a linear model and could identify social effects (endogenous + exogenous) but not separately. Moffit (2001) used a simultaneous linear equations model with groups of same sizes and was unable to identify endogenous and exogenous effects either.

sizes.[46] They both find that identification is impossible in these cases. They identify the reflection problem[47] (Manski, 1993a) to contribute in the difficulty of identification. Lee (2007) also addresses the reflection problem and tries to find a solution for identification. He uses a Spatial Autoregressive (SAR) model where the spatial weight matrix has a zero diagonal to count for group interactions.[48] He demonstrates that identification of the interaction effects is possible when there is a sufficient variation in the size of the groups specified in the model.[49] However, he specifies that large group sizes may induce weak identification. Additionally, by using a similar model[50] as in Lee (2007), Davezies et al. (2006) show that the presence of at least three different group sizes can help identification. Boucher et al. (2014) suggest another way to handle the reflection problem without using group size variations. They use at least one contextual variable they first excluded from their model as an instrument and find robust results. For estimation purpose, we adapt our sample to conform ourselves to what is done in Davezies et al. (2006) and Lee (2007). We use at least three size variation and we work with groups of sample size equal or greater than three.[51]

Another problem highlighted in the literature assessing social interaction effects is the problem of correlated unobservables. Correlated unobservables can occur when one or many variables are correlated with observed regressors and may influence observed outcomes that are absent in the data (Manski (1993a); Moffit (2001); Lee (2007)). However, the results on identification in Lee (2007) mentioned above still hold in the presence of such an issue. For this to happen, correlated unobservables are considered in the group fixed effects. Bramoullé et al. (2009) also consider them as a network fixed effect or either assume their absence for simplification. They demonstrate a clearly stated proposition for identifying social effects from correlated effects for different network structures using a linear-in-means model. As for group structure (transitive networks with undirected links), they show that the presence of at least two group sizes and a precise condition on the coefficients of the model guaranties identification of social effects. We also use some of these results by specifying group fixed effects in the model.[52]

46. Moffit (2001) considers G groups of Ng individuals each in his model. For further details on the model, see Moffit (2001)

47. The reflection problem Manski (1993a) is also known as simultaneity problem (see Moffit (2001)).

48. The use of a zero diagonal matrix means that in opposition to Manski's linear model (linear-in-expectations), the individual outcome or characteristics are not considered in the mean outcome or mean characteristics of his/her group.

49. There must be no overlapping of the groups. It means that every individual in a group interacts with the members of his group only.

50. They use a linear-in-means model which is almost the same as a SAR model where the weight matrix is no more a spatial weight matrix, but rather an interaction matrix with a zero diagonal.

51. See the section on the empirical model for more developments on the model we use.

52. See the section on the methodology.

The method

Due to the violation of the exogeneity of the regressors, an Ordinary Least Squares estimation may not be the best choice. Indeed, the endogeneity of the y_{gj}s in the model (1.12) would render OLS estimates inconsistent. As a result, we need to use another method that will give convergent and consistent estimates of the model.

The fact that we also have a very high number of groups in our data could also drive a problem of incidental parameters (as mentioned by Neyman and Scott (1948)) in our model because of the group-fixed effects α_g. Thus, to avoid this incidental parameters problem we have to get rid of the fixed effects before estimating the model (1.16). Equation (1.16) has the following matrix formulation within each group g, $\forall g = 1, ..., G$: $\mathbf{Y}_{ng} = \beta_0 \mathbf{W}_{ng}\mathbf{Y}_{ng} + \mathbf{Z}_{ng}\boldsymbol{\lambda} + \boldsymbol{l}_{n_g}\alpha_g + \boldsymbol{\epsilon}_{ng}$ where $\mathbf{Z}_{ng} = (\mathbf{X}_{ng}, \mathbf{W}_{ng}\mathbf{X}_{ng})$, $\boldsymbol{\lambda} = (\boldsymbol{\gamma}', \boldsymbol{\delta}')'$, $\boldsymbol{\epsilon}_{ng} = \boldsymbol{\eta}_{ng}\nu$, and $\boldsymbol{l}_{ng}$ is the unit vector of size $n_g \times 1$. To get rid of the group-fixed effects, we use a within equation in which we subtract the group mean outcome equation from the outcome equation (1.16). The projection matrix $\mathbf{J}_{ng} = \mathbf{I}_{ng} - \frac{1}{n_g}\boldsymbol{l}_{n_g}\boldsymbol{l}'_{n_g}$ multiplied to the group vector $\mathbf{Y}_{ng}$ of scores gives the derivation of $\mathbf{Y}_{ng}$ from its group mean $\bar{\mathbf{Y}}_g$. We have $\mathbf{J}_{ng}\boldsymbol{l}_{ng}\alpha_g = \alpha_g(\mathbf{I}_{ng} - \frac{1}{n_g}\boldsymbol{l}_{n_g}\boldsymbol{l}'_{n_g})\boldsymbol{l}_{ng} = \mathbf{0}_{n_g}$. Then, the model in difference is:

$$\mathbf{J}_{ng}\mathbf{Y}_{ng} = \beta_0 \mathbf{J}_{ng}\mathbf{W}_{ng}\mathbf{Y}_{ng} + \mathbf{J}_{ng}\mathbf{Z}_{ng}\boldsymbol{\lambda} + \mathbf{J}_{ng}\boldsymbol{\epsilon}_{ng} \tag{1.17}$$

This model can be consistently estimated using a generalized instrumental variable method.

Generalized IV method

We use the generalized IV method by Kelejian and Prucha (1998) (henceforth noted GIV) also used by Bramoullé et al. (2009) to estimate peers endogenous and contextual effects.[53] We verify that the hypothesis of independence between $\mathbf{I}$, $\mathbf{W}$, $\mathbf{W}^2$ and $\mathbf{W}^3$ for the identification of the coefficients suggested by Bramoullé et al. (2009) is fulfilled. Therefore, following Bramoullé et al. (2009) we can use the matrices ($\mathbf{JWX}$, $\mathbf{JW}^2\mathbf{X}$, etc.) as valid instruments for $\mathbf{JWY}$ in the first step of the instrumental variable strategy. We complete the method in two steps. In the first step ,we estimate a 2SLS of equation 1.17 using the matrix of ($\mathbf{JX}$,$\mathbf{JWX}$, $\mathbf{JW}^2\mathbf{X}$) as instruments for the initial model with the endogenous component $\mathbf{JWY}$. Let $\boldsymbol{\omega} = (\beta, \boldsymbol{\lambda}')'$ and $\hat{\boldsymbol{\omega}}_1 = (\hat{\beta}_1, \hat{\boldsymbol{\lambda}}'_1)'$ the estimates of the coefficients from the first step. We then use in the second step ($\widehat{\mathbf{JWY}}, \mathbf{JX}, \mathbf{JWX}$) as instruments, with $\widehat{\mathbf{JWY}} = \mathbf{JW}(\mathbf{I} - \hat{\beta}_1\ \mathbf{W})^{-1}(\mathbf{Z}\hat{\boldsymbol{\lambda}}_1 + \boldsymbol{\epsilon})$. The model estimated in the second step is thus just identified.

53. Caeyers and Fafchamps (2016), show that the estimation strategy used by Bramoullé et al. (2009) does not suffer from exclusion bias.

1.6 Data

We use data from 386 health workers performing maternal and neonatal healthcare in Benin, a country in sub-Saharan Africa. These data are provided by the World Bank and were collected from 2010 to 2011 from 128 health districts.[54] The data are composed of sixteen sub-datasets related to each health facility administration, finance, clinical information, patients and staff. The current data we use are made from a merging process of two of these sixteen sub-datasets.

The first sub-dataset is composed of 386 MCH workers, and is the principal dataset we use to define a measure of the knowledge of protocol in maternal and neonatal care of targeted health workers. Each MCH worker in the dataset gave their responses to a special questionnaire asking questions about what to do when facing patients in diverse clinical situations (all related to pregnancy or birth protocol).[55] One significant advantage with the use of these data for the study is that it gives us a unique and uniform measure for the assessment of the proxy of productivity we denoted above as y_i for all MCH workers.

The second sub-dataset addresses more general information on each health worker. All individual characteristics like age, marital status, level of education, experience, number of children, and unpaid salary are found in that dataset.[56] However, the second dataset gives information for other workers who do not necessarily provide maternal care, and thus are not listed in the first dataset. This second dataset gives information on only 493 MCH workers from the first dataset (before groups deletion). Consequently, approximately 12% of the information needed is missing in this second dataset. Before restricting the data to health facilities with more than two workers represented in the sample, the sample sizes for each health facility varied from 1 to 5. Most of the sample then was composed of HFs' samples of size 2 (23.57%) or size 3 (66.96%). Additionally, most of the missing information (86.57%) concerns HFs of sample size 3 which is substantial considering that our empirical model can be executed only for HFs of at least size 3. However, we used a simple method of imputation to recover rational information on the explanatory variables with missing values. Descriptive statistics on the initial data, the data we use and some dependent variables used in our model are shown in tables A.2 and A.3 in appendix A.2. Figures A.3a to A.3d also show some statistics about some of the variables used as individual characteristics in the model, and differences between the initial data and the imputed data.

54. The data are baseline data from a panel sketch to be collected by the World Bank. The database initially contained information on 560 MCH workers from 250 health facilities. However, due to identification issues, we had to restrain the data to the groups with more than two workers represented in the sample.

55. See Table A.1 in Appendix to have a better understanding of the content of the test.

56. We would have preferred to use data on income. However, because people are usually reluctant to give their income, we cannot use this variable. This is due to the multiple missing values in the dataset. Consequently, the next most valuable information on individuals' income we can use is whether they had unpaid salaries and bonuses.

Method for Imputing Missing Data

We use a simple method of *job position mean imputation*. By *job position*, we mean that we had to do the imputations with regard to the work category or position of the MCH worker. We have four main categories or positions in the data: doctors, midwives, nurses and nursing auxiliaries. We make the imputations that way because the values of the individual characteristics variables we chose in the empirical work depend (for the most part) on the position or level of education of individuals. There is no missing information on the position occupied by the health workers in the data. We display in the Appendix A.2 figures A.3a to A.3d comparing the empirical density of some imputed variables before and after the job-position imputations. To show that the *job-position mean imputation* gives a distribution closer to the original one in comparison to a simple *mean imputation* (or *mode imputation* for job position variables) we add a third empirical density or histogram to the graphs.

For the correction method for unobserved individuals in the sample described in section 1.5.4, we need information on the real population of MCH workers in each health facility from which each sample is drawn. However, since data on the population of MCH workers is not clearly given in the data, we use some administrative information on each HF and a deductive approach to approximate it.[57]

1.7 Empirical Developments and Results

1.7.1 Outcome of Interest: a Proxy for the Level of Effort

We construct our proxy to the unobservable level of effort in two steps.

In the first step, we compute a proportion score of skills and knowledge for each MCH worker based on their responses to the first set of twenty-one questions included in the test questionnaire.[58] All these questions test their general knowledge on maternal and neonatal care. This score named *MCHWscore* is simply a raw percentage score of correct answers over all possible modalities of responses. Since each question has multiple responses, we compute the

57. In a given health facility, we know the population of health workers for several categories. Specifically, we have information on the population of doctors, midwives, nurses and nursing auxiliaries that may be involved in maternal care. Also, we know the populations of administrative staff, doctors involved in general or specialized medicine, engineers in biomedical analysis and imaging, and so on. We then use this information to make approximations on the real populations of only MCH workers in each targeted HF.

58. The whole questionnaire can be sectioned in two parts. The first asks general questions on ANC, birth attendance, C-section and neonatal care while the second focuses only on specific neonatal care for some MCH worker who received special training on resuscitation of the newborn. To be able to compare MCH workers within themselves, we need to build a more homogeneous instrument of measure and hence we ignore the second part of the questionnaire and focus only on its first one. In table A.1 in appendix A.2, there is a recap of topics and questions types addressed in the questionnaire.

percentage of modalities correctly mentioned by the MCH worker (spontaneously or not) over all the questions.[59] Some empirical studies in which authors build indexes use some more elaborated methods such as multiple components analysis (MCA) or principal component analysis (PCA). This usually helps them to give different weights to each variable or modality. However, although it seems that it could have improved our *MCHWscore* to account for the real weights maternal and child health specialists attribute to the response to one question versus another, we still limited ourselves to a *non-weighted average* of good responses instead. A simple reason for this choice is that whatever the potential composite index obtained using MCA or PCA would have been, it would have added an additional and unnecessary issue about the clinical validity of those weighting coefficients. Yet, we can always defend that some MCH worker who on average gives more good responses and details than another one is more likely to master their knowledge in maternal and child care.

The general statistics on the sample of MCH workers give an MCHWscore in the range of 13.3/100 to 83.2/100. The highest score is held by a midwife while the lowest one is held by a nursing auxiliary.[60] Furthermore, when we consider the whole data, we display the box plot in the following figure1.2.

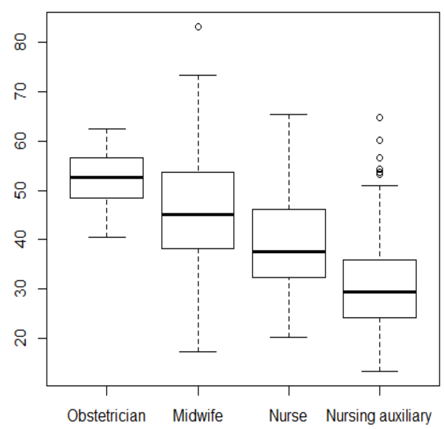


Figure 1.2 – Box plot of MCHWscore by HW position

59. For each modality of response, the data specifies if the interviewee gave the response spontaneously or if they had to take a few supplementary minutes to think of the response. However, in this work we only need to know if the MCH worker knows what to do or not. Nevertheless, we acknowledge that it would have been better to make an additional distinction between a worker who can spontaneously give a good response and another one who takes a lot of time to think of the response. In fact, if those two where to face a parturient with complications, the first could save the mother and her baby's lives; while the second, because they lack of spontaneity, could result in more complications or worse, death of the baby or the mother.

60. We were not surprised that the highest score in the sample is not held by a doctor. In fact, because there are only 5 doctors in the data versus 196 midwives, it is understandable. Additionally, most of the questions asked in the questionnaire addressed basic knowledge on maternal and child care including notions on cleanliness in maternal and child care. Also, in reality doctors do not take basic actions like preparing decontamination solution for instance. In most of the cases, they delegate those basic actions to midwives, nurses or nursing auxiliaries. Thus, this explains well why a midwife could have a better score than doctors in this small sample.

This simple plot shows that on average, the MCHWscores of nursing auxiliaries are lower than the MCHWscores of the nurses, and the latter are also lower than the MCHWscores of midwives and doctors. Yet, testing for significance in these differences using an ANOVA and a Kruskal and Wallis test, we find that there is no significant difference between doctors and midwives. [61] However, the difference in these scores are still significantly different between them and all the other groups (for instance the differences between doctors and nurses, midwives and nurses, midwives and nursing auxiliaries, etc.).

The descriptive results on the MCHWscore are somewhat consistent with what has been found in the literature. In fact, the relative safety of midwives in birth attendance over doctors has already been documented (Rifkin, 1997) suggesting that midwives are at least as good as obstetricians concerning birth attendance matters. [62]

The second step in the construction of the proxy variable for the level of effort is to define it as a concave function of the MCHWscore above. We choose this concave function to be the natural logarithm ($y_i = log(MCHWscore_i)$). The use of the *log* function instead of any other concave function (as $x^{\frac{1}{n}}, n \geq 1$ for instance) will also help us in the interpretation of the coefficients of the model 1.12.

An intuition behind the use of a concave function of the MCHWscore is given as following. Let M, m_h and m_l designate respectively the total number of modalities to be mentioned in the whole test and the number of modalities mentioned by two distinct individuals i_h and i_l. Let's assume that i_h has mentioned a higher number of modalities than i_l. Clearly, this means that $M > m_h > m_l$, $MCHWscore_h > MCHWscore_l$ and $y_h > y_l$. Now, let $\Delta m > 0$ designate a marginal increase in the number of modalities mentioned by both i_h and i_l. Taking $y_i = log(MCHWscore_i) = log(m_i * 100/M)$ means that we attribute a decreasing value to the impact of Δm on the final score y_i. Indeed, we have $\frac{\partial log(m_h*100/M)}{\partial m} = \frac{\Delta m}{m_h} < \frac{\Delta m}{m_l} = \frac{\partial log(m_l*100/M)}{\partial m}$.

Let's assume for instance that, as in figure A.1 below, two individuals i_h and i'_h mentioned 104 and $104 + \Delta m$ modalities respectively (which gives respectively $MCHWscore_{i_h} \simeq 60$ and $MCHWscore_{i'_h} \simeq 60 + \frac{\Delta m \times 100}{M}$). We chose $\Delta m = 17$ so that $\frac{\Delta m \times 100}{M} \simeq 10$. Let's also assume that two other individuals i_l and i'_l mentioned only 35 ($MCHWscore_{i_l} \simeq 20$) and $35 + \Delta m$ modalities respectively. Then, as shown in figure A.1, there is a greater perceived difference in the values of the proxy y between i_l and i'_l than there is between i_h and i'_h. For MCH workers with lower knowledge a mention of any additional modality shows more knowledge

61. We compute the (Kruskal and Wallis) rank sum test in addition to the ANOVA test to strengthen our interpretation because looking at figure A.2, some may agree or not to the hypothesis of normality of the *MCHWscore*. The p-values of Kruskal and Wallis test are available in Table A.4.

62. In her article, Rifkin (1997) states that because child birth is inherently dangerous, obstetricians were more able to injure mothers and their children than midwives. She also cited a study of 1986 which showed that perinatal mortality rates were higher for doctors than for midwives.

(in *real* value) than it does for MCH workers with higher knowledge. The questionnaire is not timed. Thus, while individuals with lower knowledge may try their best to give the maximum of what they know (maybe because they realize it's not enough), individuals with higher knowledge could lack mentioning absolutely all the modalities they know in reality (maybe by negligence or by giving speedy responses). A practical analogy with students would be as follows. Consider students who are taking an exam (not timed). When they know that they are not far from having 100%, usually they will not make as much effort to remember things they know and may have forgotten momentarily as those who know that they are not even sure of having a grade of 50%. In the end, there is usually a greater difference in the knowledge of two students who earned 40% and 50% during an exam than there is between two students with respectively 90% and 100% grades. The figure A.1 in appendix shows a representation of the proxy variable y in function of MCHWscore.

Empirical Interaction Matrix

Our knowledge a priori of the social structure inside the health facilities suggests focusing our work on group interactions. We take the whole set of MCH workers from each health facility as a group. In fact, in the current work, due to data constraints, we do not consider effects resulting from possible interactions of MCH workers with other health workers within or outside the HF.[63] Thus, in our model we do not allow for HFs overlapping or different services overlapping. In the model, the only members of a group are the MCH workers of the same sampled HF. Although we acknowledge that in the field this restricted pattern is not always observed, this constraint still makes sense. The reason is that in their duty, workers in a maternal service are expected to interact more with other workers in their service than they do with workers from other services or with workers from other health facilities in their area.[64]

1.7.2 Empirical Results

We use three sets of individual characteristics X_1, X_2 and X_3 as instruments to estimate three models (1), (2) and (3) whose results are detailed in Tables A.5 and A.6 in Appendix A.3.

63. In reality, in a given HF, apart from the case of small HFs (such as a dispensary, isolated dispensary, or isolated maternity) there may exist several other services different from the maternity service (ex. pediatrics, cardiology, trauma/psychiatry, etc.) and the doctors, nurses or nursing auxiliaries affiliated to those services may sometimes also interact with MCH workers. We assume that these effects are captured through the group fixed effect. We also do not consider the possible fact that an MCH worker can interact with another health worker (MCH worker or not) from another HF in the area. It means that we assume that a MCH worker interacts with members of their group only.

64. If they had to interact with health workers from other HFs, there is a good chance that these interactions would be in a personal cadre unless they meet at a professional conference or meeting for instance. Even in these cases, considering the short length of time of this kind of interaction in comparison to the daily or weekly interactions with peers from the same HF we consider that it is less likely for MCH workers to be influenced by those *outsiders* in their daily work. Therefore, our assumption stands reasonable.

The first model accounts for the level of education of MCH workers and their peers while the two others only account for the MCH worker's position in the health facility as an indicator of their ability.[65] We use the maximal level of education a MCH worker would have reached if they had achieved the academic cycle they claim to have attained as a measure of their education.[66]We compute general standard errors and within clusters (or groups) standard errors.

First, whatever the model (model (1), (2) or (3)), we find a strong evidence of negative endogenous peer effects not imputable to exclusion bias (see Caeyers and Fafchamps (2016) for details on exclusion bias in peer effects models). In the three models, the endogenous peer effect takes a value between $-.57$ and $-.66$, which testifies to the presence of very strong peer effects. Additionally the standard errors of these coefficients, computed within groups or not, show that these endogenous peer effects are significant at a level of 5% or 1% for the model (1) and 1% for the models (2) and (3). Though there is usually rare evidence of negative peer effect ($\beta < 0$) in the empirical literature on social interactions and peer effects, it is usually interpreted in the literature as the presence of *strategic substitutability.* This strategic substitutability could happen due to some workers *shirking* or *free riding* in the workplace. Usually in the workplace, shirking or free riding could happen, for example, when workers are not well monitored or when the relative sanctions or losses are not severe or serious enough. In developing regions like Benin, health workers may shirk or "free ride" for several reasons. In fact, in the context of Benin, there is an important lack of skilled personnel in the health sector which could demotivate or constrain HF's administrative authorities not to fire health workers easily. Additionally, in developing countries, as it is the case in Benin, health workers in the public sector are often paid fixed salaries independently of their performance at work. Similarly to the public sector in developing countries depicted in Chaudhury et al. (2006), characteristics like education levels, qualification, and experience are most determining in fixing wages.[67] Some previous literature which analyzes physicians' decision toward quantity and quality of healthcare have already highlighted the fact that in the presence of fixed salaries, physicians reduce their effort.[68] Moreover, statistics in some developing countries, often show

65. There are five possible positions for an MCH worker: (1)-midwives and doctors who are responsible for their colleagues (they supervise them), (2)-simple midwives and doctors, (3)-responsible nurses (who supervise their peers), (4)-simple nurses and (5)-nursing auxiliaries. Usually the position someone occupies in an HF is strongly correlated with their level of education.

66. In the data, individuals don't give the grade they attained in school or university. They only tell about their education level. In francophone education systems like Benin, there are usually four distinct levels: primary education (6 years), secondary education (cycle-1, 10 years), secondary education (cycle-2, 13 years) and university (20 years in average for doctors).

67. Those characteristics are most of the time predetermined, meaning that they do not change in the short term. However, workers increase their experience progressively and some of them may also increase their level of education or qualification, while still working, to increase their salaries. Yet, we do not explore these considerations in our model which is a static model.

68. For a quick review of some literature comparing fixed salary to fees for services system influence on physicians' productivity, see Fortin et al. (2008).

huge proportions of absence of health workers (35% on average), which also express that health workers shirk often through their absenteeism (see Chaudhury et al. (2006)). In response to shirking MCH workers peers, and to meet the healthcare needs of the populations served by the HF, some MCH workers have to compensate by increasing their own effort. However, this doesn't impede the existence of altruistic MCH workers who gain higher utility in providing high levels of effort regardless of (or with very little regard to) the working conditions or whatever their peers do. Then, likewise, when paired with these altruistic workers, some MCH workers may "free ride."

Second, we also find some evidence of contextual peer effects. The results show a negative and significant effect of tenure on effort at work. Specifically, we find that workers who are paired with some peers of more than 18 years of experience in the HF tend to display less effort at work compared to those whose peers are of less than 18 years of experience. It is as if workers with more than 18 years of tenure, through a certain mechanism, tend to encourage their peers to shirk or to display *social loafing*. However this effect has a very small magnitude in comparison to the endogenous peer effect. This result confirms the general finding that is often conveyed in working areas like public administration, teaching, and healthcare in developing countries. In fact, although it has not often been proved in the literature, it is a sort of general knowledge in the public sector of countries like Benin, that younger workers often copy the bad working habits of their more experienced peers. In our case here for instance, a bad habit would be a low level of effort at work. Our results also confirm a negative and significant effect of having an experience of more than eighteen years on the workers' effort at work.[69] However, this effect seems less strong in magnitude than the contextual effect of the variable *tenure* for highly weighted peers.

Another result we find interesting is the positive and very strong contextual effect of the presence, in an HF, of a responsible doctor or midwife on their peers' effort at work. This effect is close in magnitude to the endogenous effect and significant at the level of 5% in models (2) and (3). Considering that MCH workers holding an administrative responsibility in an HF often exert a certain authority, and thus a certain pressure, over their peers, this result confirms in a way the results of Kandel and Lazear (1992). Kandel and Lazear find that the level of effort can increase with peer pressure. However, in our case, peer pressure seems to come only from responsible midwives and doctors in the HF and not from responsible nurses.

Additionally, in model (3), the sensibility of a worker's effort or of their quality of healthcare to the unpaid salaries is negative and significant, as well as their sensibility to the number of dependent children a worker is taking care of in their household. It shows that not only

69. In all three models, the effect of the variable tenure, which takes value 1 when the worker has more than 18 years of experience and 0 otherwise, is negative. It is significant at the level of 10% in models ((1)) and ((2)), and at the level of 5% in model ((3)).

wage arrears demotivate MCH workers by inciting them to reduce their level of effort, but also, all else being equal, an increase in the wage arrears of an MCH worker's peers induces a decrease of the worker own effort. It seems to be similar to the effect that more experienced peers have in the workplace. This is also understandable in the sense that demotivated MCH workers can sometimes spoil the work atmosphere by rallying others to their cause. In fact, MCH workers with higher wage arrears are more likely to be demotivated (as shown with the negative individual effect of wage arrears). They are also more likely to have more years of working experience, given that unpaid salary should be correlated with work experience (in years). We don't focus on the value of the coefficient of elasticity of the wage arrears on effort because it depends on the unit of measure of effort.

Policy Implications

These results have a lot of policy implications for Benin and other countries with similar policies in their health sector or with health workers adopting a similar behavior to Benin's health workers. Although, they were focused on maternal and child-care workers, these results are still useful to address policies for all categories of health workers in Benin (even those who do not perform maternal care). The fact that we observe a strategic substitutability in effort instead of the strategic complementarity that would be more desirable in this case, raises the necessity of providing new policies to reduce the motivation of health workers to shirk. In fact, previous studies that already addressed the question of physicians productivity in different types of remuneration systems (see Fortin et al. (2008)) suggest that a change in the remuneration system can change health workers' productivity. A government like Benin government should adopt another remuneration policy like the *fee-for-service* (FFS) policy or it could also mix both the fixed wage and the FFS options to motivate health workers to increase their effort. Another more direct policy implication these results may suggest is to introduce some peer pressure in health workers environment. Actually, providing one responsible doctor or midwife to every HF could be a useful tool for that purpose.

1.8 Conclusion

This study analyzes the effect of maternal and child health workers chosen level of effort at work on the level of effort chosen by their colleagues. We use a spatial autoregressive (SAR) model to estimate this effect. Although the subject of peer effect in the workplace has been introduced in the economic literature by previous scientists, this paper makes a pertinent incursion in the subject by addressing the domain of health economics. The domain of health, particularly the domain of maternal and child health in developing countries is an important area of research mainly due to the worrying statistics on maternal and child deaths and general health in developing countries.

The results obtained in this paper open a window on understanding social interactions effects within maternal and child health (MCH) workers on their effort and the quality of their healthcare at work. We find a significant and strong negative endogenous peer effect within MCH workers. These results show that there is a mechanism of strategic substitutability between MCH workers of a same health facility (HF). The intuition behind this substitutability in the context of a country like Benin is multifold. First, it could be due to a phenomenon of shirking of MCH workers who put low effort in work-related activities and consequently provide a low quality of healthcare for the benefit of leisure or other activities. For instance, we find that the more children an MCH worker has, it tends to induce a reduction of their effort at work. In addition, the higher their unpaid wages are, the less they are motivated, and the more they shirk. Then, in reaction to these shirking habits of some MCH workers that can induce a disequilibrium of the workload within MCH workers, those who for some reason (altruism or better work ethic) cannot shirk may end up with a bigger workload and consequently will have to increase their effort at work to be able to fulfill the demand for healthcare of the HF (see Faton (2019)). However, we find that when an HF has a midwife or doctor posing as responsible for the clinical staff, their peers tend to give more effort compared to those who do not have any responsible midwives or doctor in their HF. Moreover, symmetrically to the first intuition, the strategic substitutability could be interpreted as follows. There could be altruistic or competent MCH workers who naturally provide high levels of effort at work and perform higher quality of healthcare whatever their peers do. In reaction to that, health workers who have more incentive to shirk (workers with more than eighteen years of tenure or more unpaid salary) or the ones with other personal constraints (like more children) may do so.

This work suggests that maternal and child health policy makers in developing countries, and more precisely in Benin, should modify the current policies by considering these facts. For instance, if each HF was endowed with one responsible midwife or a responsible doctor, it would help improve the resulting effort of MCH workers in the HF. In fact, responsible midwives or responsible doctors could help promote positive peer pressure in the HF. Then, this can lead to an increase in effort of their co-workers. In addition, to incite MCH workers to work more, or more adequately, we think that the application of a performance-based remuneration process could reduce their motivation to shirk. In that case, MCH workers would be more motivated to put more effort in their work and this could eventually induce less medical malpractice and other related maternal and child-care issues. Although our study could not directly compare results from a fixed wage payment scheme and a performance-based payment scheme, it can nonetheless suggest that a system of remuneration and promotion only based on years of experience is not always a good way to promote effort at work. Rather, it only helps workers over a certain threshold (18 years of experience in our case) to become less productive and to contaminate their colleagues and incite them to reduce their own level of effort. Actually, our results suggest that a young MCH worker (less than 18 years of experience) can be discouraged when they see that a more experienced co-worker (18 years of experience or more) who provide low levels of effort is promoted and have an increase in salary or status only because they are a long-standing member of the staff. At some point, these young MCH workers will eventually decide to reduce their effort at work because there is no gain for them (apart from a moral gain) in making a lot of effort at work.

1.9 Bibliography

Blomqvist Ake and Busby Colin. How to pay family doctors: Why "pay per patient" is better than fee for service. Commentary 365, C.D. HOWE Institute, 2012.

Coralio Ballester, Antoni Calvo-Armengol, and Yves Zenou. Who's who in networks. wanted: The key player. Econometrica, 74(5):1403–17, 2006.

Iwan Barankay. Rank incentives: Evidence from a randomized workplace experiment. http://repository.upenn.edu/bepp_papers/75, 2012.

J. Beugnot, B. Fortin, G. Lacroix, and M. C. Villeval. Gender and peer effects on performance in social networks. European Economic Review, 113:207–224, 2019.

L. E. Blume, W. A. Brock, S. N. Durlauf, and R. Jayaraman. Linear social interactions models. NBER, 19212, July 2013.

V. Boucher and B. Fortin. Some challenges in the empirics of the effects of networks. IDEAS Working Paper Series from RePEc, 2015.

V. Boucher, Y. Bramoullé, H. Djebbari, and B. Fortin. Do peers affect student achievement? evidence from canada using group size variation. Journal of Applied Econometrics, 29(1):91–109, 2014.

Y. Bramoullé and Rachel Kranton. Public goods in networks. Journal of Economic Theory, 135(1):478–94, 2007.

Y. Bramoullé, H. Djebbari, and B. Fortin. Identification of peer effects through social networks. Journal of Econometrics, 150:41–55, 2009.

Bet Caeyers and Marcel Fafchamps. Exclusion bias in the estimation of peer effects. NBER Working Paper, (22565), 2016.

Antoni Calvo-Armengol, Eleonora Patacchini, and Yves Zenou. Peer effects and social networks in education. The Review of Economic Studies, 76(4):1239–67, 2009.

M. Carlough and M. McCall. Skilled birth attendance: What does it mean and how can it be measured? a clinical skills assessment of maternal and child health workers in nepal. International Journal of Gynecology and Obstetrics, 89:200–208, 2005.

Nazmul Chaudhury, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers. Missing in action: Teacher and health worker absence in developing countries. Journal of Economic Perspectives, 20(1):91–116, 2006.

Estelle Codo and Bertille Agueh Onambele. Les caractéristiques des revenus des professionnels de santé et leur relation avec la fourniture des soins: Rapport bénin. http://www.who.int/alliance-hpsr/researchsynthesis/alliancehpsr_hwsbeninreport.pdf, Octobre 2009.

Emanuele Colonnelli, Edoardo Teso, and Mounu Prem. Patronage in the allocation of public sector jobs. mimeo, 2018. Version: job market paper.

Thomas Cornelissen, Christian Dustmann, and Uta Schönberg. Peer effects in the workplace. American Economic Review, 107(2):425–456, 2017.

L. Davezies, X. D'Haultfoeuille, and D. Fougère. Identification of peer effects using group size variation. IZA DP, (2324), 2006.

Peter M. DeMarzo, Dimitri Vayanos, and Jeffrey Zwiebel. Persuasion bias, social influence, and unidimensional opinions. The Quaterly Journal of Economics, 118:909–968, 2003.

Rokhaya Dieye and Bernard Fortin. Gender peer effects heterogeneity in obesity. CRREP Working Papers, 2017.

Rebecca Dizon-Ross, Pascaline Dupas, and Jonathan Robinson. Governance and the effectiveness of public health subsidies: Evidence from ghana, kenya and uganda. Working paper, Stanford University, 2017.

D. Eisenberg, E. Golberstein, and J. L. Whitlock. Peer effects on risky behaviors: New evidence from college roommate. Journal of Health Economics, 33:126–138, 2014.

Armin Falk and Andrea Ichino. Clean evidence on peer effects. Journal of Labor Economics, 24(1):39–57, 2006.

Elfried Faton. Workload share among peers in the health facility. mimeo, 2019.

Jeffrey A. Flory, Andreas Leibbrandt, and John A. List. The effects of wage contracts on workplace misbehaviors: Evidence from a call center natural field experiment. NBER Working Paper, (22342), June 2016.

Bernard Fortin and Myra Yazbeck. Peer effects, fast food consumption and adolescent weight gain. Journal of Health Economics, (42):125–138, 2015.

Bernard Fortin, Nicolas Jacquemet, and Bruce Shearer. Policy analysis in the health-services market: Accounting for quality and quantity. CIRPEE Working Papers, 2008.

Bernard Fortin, Nicolas Jacquemet, and Bruce Shearer. Labour supply, work effort and contract choice: Theory and evidence on physiscians. DIscussion Paper Series// Forschungsinstitut zur Zukunft der Arbeit, (5188), 2010.

David Gill and Victoria Prowse. A structural analysis of disappointment aversion in a real effort competition. American Economic Review, 102(1):469–503, 2012.

Benjamin Golub and Matthew Jackson. How homophily affects the speed of learning and best-responses dynamics. The Quaterly Journal of Economics, 127:1287–1338, 2012.

Benjamin Golub and Matthew O. Jackson. Naive learning in social networks and the wisdom of crowds. American Economic Journal: Microeconomics, 2(1):112–149, 2010.

J. J. Horton. Employer expectations, peer effects and productivity: Evidence from a series of field experiments. 2011.

S. A. Imberman, A. D. Kugler, and B. I. Sacerdote. Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. American Economic Review, 102(5):2048–2082, 2012.

Eugene Kandel and Edward P. Lazear. Peer pressure and partnerships. Journal of Political Economy, 100(4):801–817, 1992.

Harry H. Kelejian and Ingmar R. Prucha. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. Journal of Real Estate Finance and Economics, 17(1):97–121, 1998.

Harry H. Kelejian and Ingmar R. Prucha. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review, 40: 509–533, 1999.

Lung-fei Lee. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. Journal of Econometrics, 140:333–374, 2007.

Lung-fei Lee, Xiaodong Liu, and Xu Lin. Specification and estimation of social interaction models with network structures. The Econometrics Journal, 13:145–176, 2010a.

K. L. Leonard and M. C. Masatu. The use of direct clinician observation and vignettes for health services quality evaluation in developing countries. Social Sciences and Medecine, 61:1944–1951, 2005.

K. L. Leonard, M. C. Masatu, and A. Vialou. Getting doctors to do their best: The roles of ability and motivation in health care quality. Journal of Human Resources, 42(3):682–700, 2007.

Maureen Lewis. Governance and corruption in public health care systems. Center for Global Development, Working Paper, (78), 2006.

Xiaodong Liu and Lung-fei Lee. Gmm estimation of social interaction models with centrality. Journal of Econometrics, 159:99–115, 2010b.

Charles F. Manski. Identification of endogenous social effects: The reflexion problem. The Review of Economic Studies, 60(3):531–542, 1993a.

A. Mas and E. Moretti. Peers at work. American Economic Review, 99(1):112–145, 2009.

R. A. Moffit. Social Dynamics -, chapter 3: Policy Interventions, Low-Level Equilibria, and Social interactions, pages 45–82. 2001.

J. Neyman and E. L. Scott. Consistent estimates based on partially consistent observations. Econometrica, 16:1–32, 1948.

Eleonora Patacchini, Edoardo Rainone, and Yves Zenou. Heterogeneous peer effects in education. Journal of Economic Behavior and Organization, 134:190–227, 2017.

Danielle Rifkin. Midwifery: An international perspective–the need for universal legal recognition. Indiana Journal of Global legal Studies, 4(2):509–536, Spring 1997.

J. Rosaz, R. Slonim, and Marie-Claire Villeval. Quitting and peer effects at work. GATE WP, (1204), April 2012.

Carl Shapiro and Joseph E. Stigltitz. Equilibrium unemployment as a worker discipline device. The American Economic Review, June 1984.

Adam Wagstaff and Mariam Claeson. The millennuim development goals for health: Rising to the challenges. The International Bank for Reconstruction and Development / The World Bank, 2004.

WHO. Making pregnancy safer: Global Action for Skilled Attendants for Pregnant Women,. World Health Organization, 2002.

WHO. Pregnancy, Childbirth, Postpartum and Newborn Care: A guide for essential practice,. World Health Organization, second edition, 2006.

WHO. Trends in maternal mortality: 1990 to 2010,. World Health Organization, 2012.

Chapter 2

Workload Share among Peers in the Health Facility

Elfried Faton

2.1 Résumé

Dans cet article, j'utilise l'approche des jeux de négociations à la Nash pour établir un lien entre la part de travail de professionnels de santé maternelle et néonatale et leurs caractéristiques individuelles. Mes résultats révèlent que le niveau d'éducation, l'expérience et le nombre d'enfants d'un professionnel de santé déterminent leur pouvoir de négociation au travail, et par conséquent, leur productivité. De manière spécifique, *ceteris paribus*, les travailleurs ayant un profil d'éducation élevé ont tendance à détenir plus de pouvoir de négociation, ce qui leur donne la possibilité de maintenir une part de travail relativement peu élevée. À l'opposée, les travailleurs ayant plus d'enfants à charge ont moins de pouvoir de négociation à maintenir une charge de travail moins élevée.

Mots clés : pouvoir de négociation, professionnels de santé, pairs, productivité, part de travail.

Codes JEL : J24, J29

2.2 Abstract

In this paper, I use a simple Nash bargaining game approach, to establish how individual characteristics influence the equilibrium workload share of maternal and child care workers. My findings show that workers' education, experience and the number of their children determine their bargaining power in the workplace, and consequently, their productivity. More specifically, my results reveal that, *ceteris paribus*, highly educated workers tend to hold more bargaining power, giving them the capability to bargain for less workload. At the opposite, workers with more children are less capable of negotiating for a lower workload than their peers who have fewer children.

Keywords: bargaining power, health workers, peers, productivity, workload share.

JEL codes: J24, J29

2.3 Introduction

The scarcity of skilled birth attendants in developing countries induces a few adaptive changes in the way their health facilities (HF) are managed, in comparison with developed countries. Among these changes, nursing auxiliaries are sometimes used to deliver health interventions that are normally restricted to health workers with more official training as midwives and doctors (see WHO (2004)). These changes, in turn, affect the division of labor within health facilities (HF), and consequently workers and their co-workers' productivity. The sample data used in this paper show that 31.6% of all the nursing auxiliaries perform more birth deliveries than average in their HF, and 22.4% of them all perform more than the average overall HFs. In the paper, I focus on maternal and child healthcare (MCH) providers, and I use workers individual characteristics to explain how it affects their bargaining power inside the health facility.

There are two major contributions of this paper to the existing literature on workers' productivity when they interact with peers (see for example Falk and Ichino (2006), Mas and Moretti (2009), Fortin et al. (2010), Flory et al. (2016), Cornelissen et al. (2017), Beugnot et al. (2019)). First, this paper is the first to estimate MCH workers' bargaining power using their individual characteristics, and to explain how it affects their share of the total workload (demand for birth deliveries). Second, the paper uses peer productivity in a novel way, to describe some underlying mechanisms explaining the possible strategic substitutability observed in MCH workplace, as it is found in Faton and Fortin (2019).

Faton and Fortin (2019) use a non-cooperative game approach to explain how the productivity of MCH workers from Benin is influenced by their peers' own productivity. At the Nash equilibrium, they obtain a linear best-response function whose structure is similar to a spatial autoregressive (SAR) model structure.[1] They find the evidence of negative peer effect in effort among MCH workers, which they interpret as strategic substitutability. However, their model does not allow them to fully explain the mechanisms behind this substitutability. Their results suggest that in some health facilities, conditional on some of their individual characteristics (experience, position in the HF, etc.) some healthcare workers tend to shirk more in the presence of highly skilled and productive peers. Conversely, some workers tend to increase their effort in the presence of less skilled or shirking peers. In particular, the paper find that, in the presence

1. Linear models like SAR and linear-in-means models have been extensively used in the literature to estimate peer effects. For instance, Lee (2007), Lee et al. (2010a), Boucher et al. (2014), Dieye and Fortin (2017) among several other authors use such linear models to assess peer effects and social influence.

of a responsible (monitoring) peer, workers tend to increase effort and provide better quality of care.

In a different fashion, this paper proposes a cooperative Nash-bargaining game approach (Nash (1953)) to study the productivity of the same MCH workers studied by Faton and Fortin (2019). Hence, the best-response functions obtained are non-linear, and this approach provides additional arguments to unveil some underlying mechanisms in support of the results in their paper. In this paper, workers' bargaining power simply refers to their disagreement outcome ("threat point").[2] A low threat point means that the worker is able to accept a low outcome to stay on their job. Intuitively, all else being equal, such workers are more likely to accept a bigger workload, even though it decreases their utility, than workers with higher threat points. In the paper, I use a Nash-bargaining model to find which individual characteristics, affect workers' utility and disagreement outcome, and consequently their workload share. Then, I use a simple non-linear least square (NLS) method to estimate the parameters of the model obtained. In particular, given the form of the equilibrium outcome which is subject to hundreds of inequality constraints, I use an *Augmented Lagrangian Algorithm* to compute the estimates numerically (see Varadhan (2015)). The data I use is composed of a sample of 431 birth attendants (doctors, midwives, nurses and nursing auxiliaries) across more than 100 health facilities in Benin.[3]

My findings reveal that individual workload share tends to increase with their salary. However, interestingly, I find that workers who have higher levels of education are able to bargain for a lower workload share *ceteris paribus*. This result confirms the common knowledge that most nursing auxiliaries have less bargaining power than other healthcare professionals. At the opposite, workers who have more children tend to express less bargaining power. For 99% of the sample, the estimates show a negative partial dependence between individuals workload share and their experience in their HF.

The paper is organized as follows. In Section 2.4, I describe the model, its microfoundations, and I propose a few comparative statics. In Section 2.5 a brief description

2. In the Nash-bargaining game theory, it is also common to see that the bargaining power represents a particular parameter which affects the surplus utility $u_i - d_i$ of each individual in the model. However, when that parameter is normalized to 1 for everyone, then it is also usual to interpret an individual's disagreement outcome directly as their bargaining power (see Pollack (2005)).

3. Initially, the data is composed of a sample of 560 birth attendants, organized in groups (HF). However, because my model requires that their co-worker productivity be included, I remove all single groups. In addition, my model also requires the use of workers monthly salary which is not always provided. Therefore, I removed all groups in which fewer than two co-workers' wages are mentioned.

of the data is given, and the results are delivered in Section 2.6. Finally Section 2.7 concludes.

2.4 Model

This section describes the model, and is split into three parts. In the first part, I describe the workers' technology of production, that is how hours worked are transformed into a quantity (number) of births attended. The second part describes a representative worker's problem, and the third one develops the bargaining problem. Let i indexes a maternal and child health (MCH) worker at post in a given health facility. The health facility receives a total of Q maternal and child care patients. Then, each MCH worker i has to take care of a proportion $q_i = r_i Q$ patients, so that the demand for healthcare is entirely met. That is $\sum_i r_i = 1$. One of the main goals of my model is to explain the sharing rule under which MCH workers divide the demand for maternal care (Q) among themselves. To produce q_i units of maternal care to patients,[4] i has to make work-related effort during h_i hours. The proposed birth attendance technology is as follows.

2.4.1 Birth Attendance Technology

The worker's technology of production is described by the following Cobb-Douglas type equation[5]

$$q_i = h_i^{\gamma} \tag{2.1}$$

where $\gamma \in (0, 1)$ denotes the marginal return to the hours worked towards the production of birth attendance. I assume decreasing returns to scale. This means that over time, the level of output (birth attendance) produced by a worker increases at a decreasing rate. According to the data, a simple OLS estimation of the log-linear form of equation (2.1), I find an estimate of the parameter $\hat{\gamma} \approx .32$, significant at the level of 1%.[6] For simplicity, I consider only the case where γ is homogeneous over the whole

4. Here a unit of maternal care represents each patient to whom the worker provides healthcare. Therefore q_i is the total number of patients i has to provide care for.

5. In the short run, the worker's capital is held constant, normalized to 1. This capital here may include the endowment in capital, medical technology for maternal care, technical material and equipment, and other resources in their Health Facility (HF), necessary to produce birth deliveries.

6. See table B.1 in the appendix for details. Note that adding a constant to the linear form of equation (2.1), I still find an estimate $\hat{\gamma} \approx .33$ significant at 1%, in a 95% confidence interval $[.10, .56]$. However, the model predicts a very small value for the constant term; and the probability of rejection of the null hypothesis is very low ($p - value > .96$) for the estimate of that constant term.

population of MCH workers.[7]

2.4.2 MCH Worker's Problem

A representative MCH worker i is a consumer, whose utility u_i depends on their consumption goods c_i, and the hours worked to provide maternal and child care to their patients h_i. A worker's utility respects the usual properties of: *(i)* positive marginal utility of consumption ($\frac{\partial u_i}{\partial c_i} > 0$), and *(ii)* disutility of effort ($\frac{\partial u_i}{\partial h_i} < 0$). For more tractability, I propose a simple utility form:

$$u_i = u(c_i, h_i) = \beta ln(c_i + 1) - (1 - \beta) h_i^{\gamma} \qquad 0 < \beta < 1 \tag{2.2}$$

We can easily verify that, to allow for any consumption level $c_i \geq 0$ and any hours worked $h_i \geq 0$, the condition $0 < \beta < 1$ is required to meet both conditions *(i)* and *(ii)*.

A worker is subject to a 2-dimensional constraint. The first one is their budget constraint (BC), which depends solely on their monthly salary y_i in their associated HF.[8] Second, the maternal healthcare market equilibrium constraint (MEC) is defined as mentioned in the introductory paragraph on the model (demand for maternal and child healthcare (Q) = supply of healthcare by MCH workers ($\sum_i q_i = \sum_i h_i^{\gamma}$)). The consumption good is a *numeraire good*, and the (binding) constraints are given by:

$$\begin{aligned} BC: \quad & c_i = && y_i, \quad y_i \geq 0 \\ MEC: \quad \textstyle\sum_i q_i = \sum_i h_i^{\gamma} & = && Q, \quad h_i \geq 0 \end{aligned}$$

From the MCH worker's problem (2.2), all we can say is that worker i tries to work for the fewer hours possible in the set of their feasible outcomes, hence producing the smallest quantity q_i; and conditional on their peers production, they produce $Q - \sum_{j \neq i} q_j$. Depending on what their peers do, any outcome from $q_i = 0$ to $q_i = Q$ could be a solution. All we can say is that if co-workers' production is high, then the worker's resulting production is low; and vice versa. This does not give any information on

7. Given that the marginal return to the hours worked by a doctor may differ significantly from that of a nursing auxiliary, we could consider a case with heterogeneous γ_i such that the marginal return could vary from a worker to another, or from a category of workers to another one. In that case, the birth attendance technology would become: $q_i = h_i^{\gamma_i}$. Later however, given that the estimated model uses data on quantities q_i, a specification with heterogeneous marginal returns would not have changed the results obtained.

8. If data is available, it is also possible to consider the worker's non-labor income y_{nli} or any income from any other job they may have (for part-time workers). However, even if the researcher observes y_{nli}, adding it to y_i in the budget constraint does not change the theoretical results much.

the possible variables which may affect the workload sharing rule in real workplace situations. In the next section, I propose a refinement of the MCH worker's problem in which, all co-workers engage into a Nash-bargaining game which accounts for other significant workers' characteristics.

2.4.3 Workers' Bargaining Problem

Let n denote the number of MCH workers in a HF, and $N = \{1, 2, ..., n\}$ the set of all MCH workers. In this section, I assume that each worker i engages in a bargaining activity with their peers $j \in N \setminus \{i\}$ to determine their shares of the total workload r_i, r_j. A worker i has a disagreement point d_i, which is their outcome in case they fail to reach an agreement with their co-workers. In other words, it is the minimum outcome that worker i requires to keep engaging in the bargaining activity. Therefore, for each worker, the condition $d_i \leq u_i$ must hold for any agreement to be possible. In practice, the disagreement point may represent their outcome if they have to resign from their job in case they disagree with the sharing rule proposed to them. We can see d_i as their opportunity cost of working in the HF. This cost depends on the worker's potential outcome (expected income) outside the HF.

Let p denote the probability for a worker to find another job with *fair* remuneration according to their individual characteristics. With probability p they earn a potential income y^{pot}, and with probability $1 - p$, they stay unemployed with income y_{nli}. For simplicity, and without loss of generality, I take y_{nli} as null for all workers.[9] The vast literature on human capital and earnings, initiated by Mincer (1974), suggests that the characteristics influencing worker remuneration include principally their level of education and their experience. In the case of female workers, it is especially relevant to extend these characteristics to other variables like the number of children (Mincer and Polacheck (1974), Waldfogel (1997)).[10] In addition, each worker faces a job search cost JS_i which depends on a fixed job search cost JS_0 which is the same for everyone, and also on their individual characteristics. In sum, a worker i's disagreement outcome

9. Note that y_{nli} may take value zero, meaning that worker i expect no unemployment benefit or transfers (family) if they are unemployed. In the context of Benin studied in this paper, there is no unemployment benefits policy, therefore y_{nli} may only be interpreted as transfers from family or relatives. Accounting for y_{nli} in the model has no influence on the results of interest and their interpretation. Therefore I standardize its value as zero for simplicity.

10. In this paper, the model is applied to a sample of healthcare workers including more than 95% women. Thus the need to account for the number of children in the theoretical model as well. However, results omitting the number of children are also presented.

is modelled as follows:

$$d_i = \quad f(\boldsymbol{z_i}, \boldsymbol{\alpha}) \tag{2.3}$$

$$\begin{aligned} \text{s.t.} \qquad f(\boldsymbol{z_i}, \boldsymbol{\alpha}) = &\quad p\ ln(y_i^{pot}) - JS_i \qquad\qquad p \in [0,1] \\ \text{and} \qquad\qquad JS_i = &\quad JS_0 + \delta_s s_i + \delta_t t_i + \delta_{ch} ch_i \end{aligned}$$

The potential income y_i^{pot} represents the income that a worker i may expect given their education (schooling s_i), years of experience (t_i), and the number of children (ch_i) they have. Thus, following Mincer (1974) model of human capital (*log*) earnings and also accounting for the number of children, y_i^{pot} can be expressed as a log-linear form:

$$ln(y_i^{pot}) = \quad ln(y_0) + \gamma_s s_i + \gamma_t t_i + \gamma_{t2} t_i^2 + \gamma_{ch} ch_i \tag{2.4}$$

In general, a worker with very low education and experience should have lower opportunity cost than one with higher attributes. From the literature on human capital and earnings, I expect the parameters γ_s, often interpreted as a measure of return to schooling (see Heckman et al. (2003)) to be positive. However, the results obtained in the literature are ambivalent about the sign of the parameters γ_t and γ_{t2}. Nonetheless, these two parameters are expected to have different signs to ensure a concavity of the earnings in terms of experience. Concerning the job search equation JS_i, if $\delta_s > 0$, then *ceteris paribus*, *fair job* search cost increases with the level of education, and vice versa. Similarly $\delta_t > 0$ (respectively $\delta_{ch} > 0$) means that, *ceteris paribus*, *fair job* search cost increases with the level of experience (respectively the number of children), and vice versa. Intuitively, we can expect δ_{ch} to be clearly positive to express the higher job search cost for mothers with more children. However, the signs of δ_s and δ_t could be negative so that more qualified workers face lower job search costs. Nonetheless, considering the socio-economic climate in Benin, it is also possible that the more educated and experienced workers have more difficulty to find *fair* jobs. Let's recall here that a *fair job* is defined as a job with a *fair salary*, which means that the salary paid is in accordance with their education and experience, as indicated in the equation *à la* Mincer. $p = 1$ means that the socio-economic climate is such that workers always find a *fair* job, while $p < 1$ means that they may not. Overall, replacing the bottom equations in (2.3) and equation (2.4) into the second equation in (2.3), a worker's disagreement outcome can be estimated as follows:

$$f(\boldsymbol{z_i}, \boldsymbol{\alpha}) = \boldsymbol{z_i'\alpha} \tag{2.5}$$

where :

$$\boldsymbol{z}_i = (1,\ s_i,\ t_i,\ t_i^2,\ ch_i)'$$
$$\boldsymbol{\alpha} = (d_0,\ \lambda_s,\ \lambda_t,\ \lambda_{t2},\ \lambda_{ch})'$$

and such that:
$$\begin{aligned} d_0 &= p\ ln(y_0) - JS_0 \\ \lambda_s &= p\gamma_s - \delta_s \\ \lambda_t &= p\gamma_t - \delta_t \\ \lambda_{t2} &= p\gamma_{t2} \\ \lambda_{ch} &= p\gamma_{ch} - \delta_{ch} \end{aligned}$$

It is clear that the parameters $\gamma_s,\ \gamma_t, \gamma_{t2},\ \gamma_{ch}\ \delta_s,\ \delta_t,\ \delta_{ch}$ and p are not fully identified from equation (2.5). However, this does not represent an issue for the type of discussion intended in this paper. In addition, there is no constraint remaining for the signs of the parameters λ_s and λ_t. Any sign obtained from our estimations is valid and helpful to understand workers disagreement outcomes in the context studied. In particular, positive values of λ_s (respectively λ_t) interpret higher disagreement outcomes for highly educated (respectively experienced) workers; which means that more educated and experienced workers have in general higher income or lower job search cost. On the contrary, a negative λ_s (respectively λ_t) may interpret that the economic climate is such that more educated (respectively experienced) workers bear higher job search cost ($\delta_s > 0$, respectively $\delta_t > 0$) or that the probability of finding. This, in turn would make *ceteris paribus* highly educated workers, have lower disagreement points.

Given their disagreement points, each worker chooses a utility level u_i which is a solution of the following optimization problem:

$$\max_{u_1,\dots,u_n} \prod_{i=1}^{n}(u_i - d_i),\ \ s.t.\ (u_1, ..., u_n) \in U,\ \ (u_1, ..., u_n) > (d_1, ..., d_n) \tag{2.6}$$

Here, the constraint $u_i > d_i$ is strict because the sample concerns workers who are still working.[11] I assume for simplicity that all workers have the same bargaining capabilities (power). However, in a more sophisticated version described in section B.2

11. The strict constraint helps to avoid any issue arising from workers' indifference between staying on the job and resigning; as the equality constraint $u_i = d_i$ would allow.

in the appendix, I propose an alternative model allowing for heterogeneity in workers bargaining power. The next proposition rewrites the Nash-bargaining problem in terms of the workload sharing rule $(r_1, ..., r_n)$ for workers.

Proposition 2.1. *Nash-bargaining problem*
Given workers individual characteristics $(y_i,\ s_i,\ t_i,\ ch_i)$, the bargaining optimization problem (2.6) is equivalent to the following:

$$\max_{r_1,\ldots,r_n} \prod_{i=1}^{n} \left(\beta ln(y_i+1) - (1-\beta)r_iQ - d_0 - \lambda_s s_i - \lambda_t t_i - \lambda_{t2} t_i^2 - \lambda_{ch} ch_i\right) \tag{2.7}$$

$$\begin{aligned} s.t.\quad & (r_1, ..., r_n) \in [0,1],\ \textstyle\sum_i r_i = 1, \\ & \beta ln(y_i+1) - (1-\beta)r_iQ - d_0 - \lambda_s s_i - \lambda_t t_i - \lambda_{t2} t_i^2 - \lambda_{ch} ch_i > 0,\ \forall\ i \end{aligned}$$

Proof. Replacing equations (2.2), (2.5) and $q_i = r_iQ$ in (2.6), the equivalence is straightforward. It comes from the strict monotonicity of the objective function of the optimization problem (2.6) in $(u_1, ..., u_n)$. □

Solving problem (2.7), the following theorem 2.1 is obtained.

Theorem 2.1. *Workload sharing rule at equilibrium*
There exists a unique sharing rule $\boldsymbol{r^*} = (r_1^*, ..., r_n^*)$ solving problem (2.7), and satisfying the four axioms of a classical Nash-bargaining outcome. [12] This sharing rule verifies the

12. Nash (1953) gives a list of axioms specific to the solution point of a bargaining game. The four main axioms summarizing them are listed as follows (see Maschler et al. (2013), Barron (2013), Cho and Matsui (2013)):
(i) Efficiency: There exists no point in the feasible set S which weakly dominates the solution $\boldsymbol{r^*}$.
(ii) Symmetry: If the game is symmetric ($d_i = d_j\ \forall\ i,j$, and $(r_1, ..., r_i, ..., r_j, ..., r_n) \in S \Rightarrow (r_1, ..., r_j, ..., r_i, ..., r_n) \in S$), then $r_i^* = r_j^*\ \forall\ i,j$.
(iii) Invariance: Any positive affine translation of the utilities does not affect the outcome obtained in the bargaining process.
(iv) Independence of irrelevant alternatives (IIA): For every bargaining game restricted to S', $s.t.\ S \subseteq S'$ with the same disagreement points as the bargaining game with feasible set S, and a new solution point ($\boldsymbol{r'^*} \in S'$) also contained in S, the new solution in S' is the same as the solution in S ($\boldsymbol{r'^*} = \boldsymbol{r^*}$).

following equation:[13]

$$r_i^* = \frac{1}{Q(1-\beta)}\left(\beta ln(y_i+1) - d_0 - \lambda_s s_i - \lambda_t t_i - \lambda_{t2} t_i^2 - \lambda_{ch} ch_i\right) - \frac{1}{Q(1-\beta)}\left(\sum_{j\neq i}\left(\beta ln(y_j+1) - (1-\beta)r_j^* Q - d_0 - \lambda_s s_j - \lambda_t t_j - \lambda_{t2} t_j^2 - \lambda_{ch} ch_j\right)^{-1}\right)^{-1}$$
$$s.t.\ \beta ln(y_j+1) - (1-\beta)r_j^* Q - d_0 - \lambda_s s_j - \lambda_t t_j - \lambda_{t2} t_j^2 - \lambda_{ch} ch_j \geq 0,\ \forall\ j \tag{2.8}$$

Proof. See Appendix B.3. □

2.4.4 Comparative Statics

Following equation (2.8) giving the workload sharing rule at equilibrium in each HF, I summarize some comparative statics as follows. Given the equilibrium sharing rule, r^*_{i,x_i} represents the partial derivative of $r_i^*(x_i, ...)$ with respect to x_i. Let $\theta = 1/Q(1-\beta)$, and $B_j = \beta ln(y_j) - (1-\beta)r_j Q - d_0 - \lambda_s s_j - \lambda_t t_j - \lambda_{t2} t_j^2 - \lambda_{ch} ch_j,\ \forall\ j$. Then:

$$
\begin{aligned}
r^*_{i,y_i} &= \frac{\beta\theta}{1+y_i} && (>0)\\
r^*_{i,s_i} &= -\lambda_s\theta && (\text{opposite sign of } \lambda_s)\\
r^*_{i,t_i} &= -(\lambda_t + 2\lambda_{t2} t_i)\theta && (\lesseqgtr 0)\\
r^*_{i,ch_i} &= -\lambda_{ch}\theta && (\text{opposite sign of } \lambda_{ch})
\end{aligned}
$$

The wage elasticity of the labor supply (in terms of work share), is given by the quantity $\xi_{r_i,y_i} = \beta y_i / ((1-\beta) r_i Q(1+y_i))$. Therefore, there is a variability in the way workers would react to a change in their wage. Although, *ceteris paribus* the workload share of workers increase with their wages, the magnitude of this increase depends very much on the order of magnitude of the demand (Q) for maternal care in their HF. In particular, a worker's labor supply is elastic when the quantity $r_i Q(1+y_i)/y_i$ is below the threshold of $\beta/(1-\beta)$. For instance, if we compare two HFs composed of the same number of workers with identical vectors of workload shares in both groups, but different demand Q_1 and Q_2, those who face a higher demand in their HF would be less sensitive to changes in their wages than the others. This is quite intuitive in the sense that workers who are already doing almost their maximum at work cannot do more even though they are incentivized through an increase in wages. On the contrary, the least productive workers would react better to such incentives.

13. Equation (2.8) corresponds to the interior solution.

2.5 Data

I use the same database from the World Bank's baseline study for the RBF-financing in Benin Health sector, as in Faton and Fortin (2019). However, because there are several missing values for workers' salaries, I simply remove from the data all the health facilities where at least two co-workers salaries are not observed. This allows me to still describe the Nash-bargaining process for most health facilities. Given that the main parameters $\boldsymbol{\Lambda} = (\beta, \boldsymbol{\alpha}')' = (\beta, d_0,\ \lambda_s,\ \lambda_t,\ \lambda_{t2},\ \lambda_{ch})$' to be estimated should be identified in equation (2.8), all health facilities which sample is composed of a single individual, are also removed. After removing all irrelevant HFs in this sense, it remains 431 individuals in the sample, corresponding to 175 HFs. The data covers the period from late 2010 to 2011 and concerns mainly female MCH workers (95%). They are summarized in table 2.1 below.

2.6 NLS Estimation and Results

Given the non-linear form of model (2.7), I use a simple non-linear least square (NLS) method to estimate its parameters. The vector of parameters Λ in the sharing-rule is estimated through the following form:

$$\boldsymbol{q} = F(\boldsymbol{X}, \boldsymbol{\Lambda}) + \boldsymbol{\varepsilon} \tag{2.9}$$

where $\boldsymbol{q}$ is the vector of workload quantities q_i for all n workers in the data, $\boldsymbol{X} = (-ln(\mathbf{y}), \boldsymbol{i}, \boldsymbol{s}, \boldsymbol{t}, \boldsymbol{t}^2, \boldsymbol{ch})$ is the $n \times 6$ matrix of covariates. $F()$ is defined using the transformation $\boldsymbol{q}$=$\boldsymbol{r}^*Q$ in equation (2.7). $\tilde{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda}/(1-\beta)$, $\boldsymbol{i}$ is the n-dimensional unit vector, and $\boldsymbol{\varepsilon}$ is the $n \times 1$ vector of error terms ε_i. The estimated parameters are obtained by solving the minimization of the objective function:

$$\min_{\substack{\tilde{\boldsymbol{\Lambda}}, \\ \boldsymbol{q}+\boldsymbol{X}\tilde{\boldsymbol{\Lambda}} \leq \boldsymbol{0}}} \quad \boldsymbol{q}'\boldsymbol{q} - F(\boldsymbol{X}, \tilde{\boldsymbol{\Lambda}})'\boldsymbol{q} - \boldsymbol{q}'F(\boldsymbol{X}, \tilde{\boldsymbol{\Lambda}}) + F(\boldsymbol{X}, \tilde{\boldsymbol{\Lambda}})'F(\boldsymbol{X}, \tilde{\boldsymbol{\Lambda}}) \tag{2.10}$$

To solve this non-linear optimization problem with n linear inequality constraints, I use an Augmented Lagrangian (auglag) algorithm (Conn et al., 1997). [14] However, given the large number of inequality constraints (*dimension(q)* $:= 431 \times 1$) in (2.10), it prevents computations to find an optimal local minimizer for the non-linear objective function.

14. The algorithm already exists in several computation software. I use the one developed in R under the package *alabama* (see Varadhan (2015)).

Table 2.1 – Descriptive Statistics

Variable	Description	Mean	Std. Dev.	Min	Max
log of income ($ln(y_i)$)	Income in thousands XOF	10.72	.59	6.22	13.3
Education (s_i)	Minimum years of schooling required to reach current education level	11.36	2.7	6	20
Experience (t_i)	Number of years of experience	5.18	4.94	.08	34
Children (ch_i)	Number of children	2.27	1.59	0	9
Quantity (q_i)	Number of births attended over a six months period	57.68	54.28	0	343
Age	-	35.07	8.27	15	70
Worker category	**Proportion in sample**				
Midwife or doctor	30.63%				
Nurse	23.43%				
Nursing auxiliary	45.94%				
Female	95.82%				
Married (monogamy or polygamy)	61.95%				

Source: Author's calculations

This is due to the fact that, for such a rough (not smooth) objective function, starting from a vector of initial values of the parameters, may lead to suboptimal estimates.[15] Therefore, I perform an algorithm consisted of a set of 300 repeated *auglag* algorithms to find a mean estimator of the parameters obtained. The results are summarized in table 2.2.

The results confirm an estimate of the coefficient $\hat{\beta} \in (0, 1)$. This result confirms that workers labor supply is an increasing function of their wage. Therefore, the estimate of the elasticity ξ_{r_i,y_i} is approximately $19/q_i$.[16] Then, workers' labor supply q_i is inelastic

15. In fact, different vectors of initial values for the parameters in the computed Augmented Lagrangian algorithm may lead to slight differences in the optimum computed.

16. Note that wages are mostly expressed in tens and hundreds of thousands XOF. Thus the ratio $y_i/(1+y_i) \approx 1 \; \forall \; i$.

Table 2.2 – NLS estimation of the sharing rule

Estimates using BFGS method		
	Mean estimates of the coefficients nrep=303	Minimum among all repetitions
log of income ($\hat{\beta}$)	.95***	.95
	(0.00041)	
Constant ($\hat{d}_0$)	−7.98***	−8.34
	(0.148)	
Education ($\hat{\lambda}_s$)	.04***	.06
	(.002)	
Experience ($\hat{\lambda}_t$)	.17***	.23
	(0.003)	
Experience^2 ($\hat{\lambda}_{t2}$)	−.004***	−.006
	(0.00008)	
Children ($\hat{\lambda}_{ch}$)	−.35***	−.45
	(0.005)	

Standard Errors in brackets ()
Signif. level: *** 1‰,** 1%,* 5%

to changes in wages above the threshold of 19 birth attendance over a period of six months, and it is elastic below that threshold. According to the estimates in table 2.2, for 76.6% of workers in the sample, the wage elasticity of workers' supply (ξ_{r_i,y_i}) is positive and inelastic. In addition, given $pln(y_0) > 0$ and the negative estimate $\hat{d}_0$, workers face a positive fixed job-search cost $\hat{JS}_0 = pln(y_0) - \hat{d}_0$ irrelevant of their characteristics. As expected, the estimate $\hat{\lambda}_s$ is positive; and therefore, a worker's disagreement outcome increases with their level of education. In other words, the minimum outcome workers are ready to accept to stay on the job in a HF increases with their level of education. The intuition is straightforward. The more educated, the higher the potential wage expected outside of the HF ($\gamma_s > 0$) in case of disagreement. In turn, the relative workload share of workers decreases with their level of education. On the contrary, compared to co-workers without children, the more children workers have, the lower their potential salary, and their disagreement outcome ($\hat{\lambda}_{ch} < 0$); and consequently, the higher their relative workload share. Given that workers in the data are mostly women, this result confirms the literature stating that mothers' earnings decrease in comparison with women without children (see Mincer and Polacheck (1974),Waldfogel

(1997),Grogger (2009)). The disagreement outcome of workers is a concave function of their experience ($\hat{\lambda}_t > 0,\ \hat{\lambda}_{t2} < 0$). In consequence, up to a certain threshold, their workers disagreement outcome increases with their experience. However, after a certain threshold, workers lower their disagreement outcome (respectively increase their workload share). This also results from the concavity of the log of the potential wage in terms of experience. In fact, numerous empirical papers in the literature often find a concave relationship between the log of workers wage and their work experience (see Lagakos et al. (2018)). In particular, the threshold for an increase in workload share in terms of experience ($r^*_{i,t_i} > 0$), is estimated at $\tilde{t}_i = -\hat{\lambda}_t / 2\hat{\lambda}_{t2} \approx 19.26$ years.[17] Figure 2.1 shows the variations of the estimated workload share in terms of workers tenure when the remaining variables and the demand Q are fixed for all HFs.

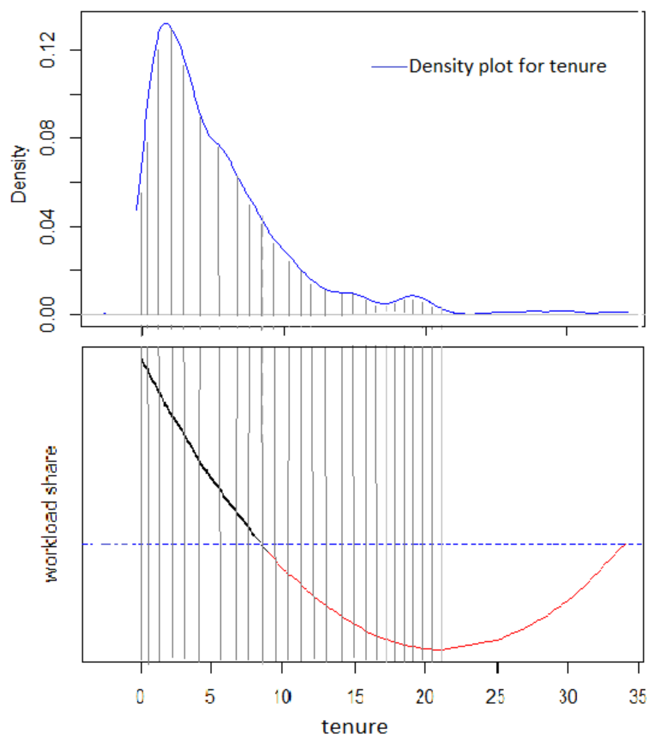


Figure 2.1 – Variations (partial) of the estimated workload share $\hat{r_i^*}$ in terms of tenure t_i

The lower plot represents the partial variations of the opposite of the estimated disagreement point in terms of workers tenure. The upper plot in figure 2.1 represents the density plot of the variable tenure in the data. It shows that only a very small proportion (around 1%) of individuals in the data has an experience above the 21.25 years threshold. The hatched areas in both parts of the figure represent workers with

17. This quantity is computed from the rough (not rounded) values of the estimates. Using the rounded values in table 2.2, the threshold should be approximately 21.25 years.

tenure below 21.25 years. They represent around 99% of the data. Therefore, the results confirm in majority the findings in Faton and Fortin (2019) who find that workers with more tenure tend to provide less effort.

2.7 Conclusion

This paper explains how individual characteristics influence the workload sharing rule of maternal care workers within a health facility. A major innovation in the paper is the direct application of a Nash bargaining model to the division of labor, in terms of the number of patients treated over a given period of time, among maternal and child care workers. The empirical results presented in the paper are specific to the case of Benin, a sub-Saharan country. In the model, a worker's decision to provide healthcare for a certain share of the total demand addressed to their HF depends on their preference over consumption and labor (opposed to leisure) and some of their individual characteristics. The model describes characteristics influencing workers' disagreement outcome, which is an infimum of the set of outcomes that workers require to stay on their job.

The results show that the workload share of a worker depends significantly on their wage, and some other characteristics like their education, experience and number of children. All else being equal, higher wages are often associated with a higher workload share. Moreover, compared to those who do not have any children, workers with children have a lower disagreement outcome, and therefore, *ceteris paribus*, they have to produce for a greater part of the workload at equilibrium. At the opposite, more educated workers have more leverage, and thus, are able to bargain for a lower share of the workload in comparison with less educated workers with identical characteristics. The wage elasticity of workers' supply is positive and inelastic for most (76.6%) workers.

2.8 Bibliography

Emmanual N. Barron. Game Theory: An Introduction. John Wiley and Sons, Inc., second edition, 2013.

J. Beugnot, B. Fortin, G. Lacroix, and M. C. Villeval. Gender and peer effects on performance in social networks. European Economic Review, 113:207–224, 2019.

V. Boucher, Y. Bramoullé, H. Djebbari, and B. Fortin. Do peers affect student achievement? evidence from canada using group size variation. Journal of Applied Econometrics, 29(1):91–109, 2014.

Arun Chandrasekhar, Horacio Larreguy, and Juan Pablo Xandri. Testing models of social learning on networks: Evidence from a framed field experiment. mimeo, 2012.

In-Koo Cho and Akihiko Matsui. Search theory, competitive equilibrium, and the nash-bargaining solution. Journal of Economic Theory, 148:1659–1688, 2013.

A. R. Conn, Nick Gould, and Ph. L. Toint. A globally convergent lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds. Mathematics of Computation, 66(217):261–288, 1997.

Thomas Cornelissen, Christian Dustmann, and Uta Schönberg. Peer effects in the workplace. American Economic Review, 107(2):425–456, 2017.

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

Rokhaya Dieye and Bernard Fortin. Gender peer effects heterogeneity in obesity. CRREP Working Papers, 2017.

Armin Falk and Andrea Ichino. Clean evidence on peer effects. Journal of Labor Economics, 24(1):39–57, 2006.

Elfried Faton and Bernard Fortin. Peer effects in effort within maternal and child care workers in benin. mimeo, 2019.

Jeffrey A. Flory, Andreas Leibbrandt, and John A. List. The effects of wage contracts on workplace misbehaviors: Evidence from a call center natural field experiment. NBER Working Paper, (22342), June 2016.

Bernard Fortin, Nicolas Jacquemet, and Bruce Shearer. Labour supply, work effort and contract choice: Theory and evidence on physiscians. DIscussion Paper Series// Forschungsinstitut zur Zukunft der Arbeit, (5188), 2010.

Benjamin Golub and Matthew O. Jackson. Naive learning in social networks and the wisdom of crowds. American Economic Journal: Microeconomics, 2(1):112–149, 2010.

Jeffrey Grogger. Welfare reform, returns to experience, and wages: Using reservation wages to account for sample selection bias. The Review of Economics and Statistics, 91(3):490–502, 2009.

James J. Heckman, Lance J. Lochner, and Petra E. Todd. Fifty years of mincer earnings resgression. NBER Working Paper, (9732), 2003.

Matthew O. Jackson. Social and Economics Networks -, chapter 8, pages 223–256. Princeton University Press, 2008.

David Lagakos, Benjamin Moll, Tommaso Porzio, Nancy Qian, and Todd Schoellman. Life cycle wage growth across countries. Journal of Political Economy, 126(2), 2018.

Lung-fei Lee. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. Journal of Econometrics, 140:333–374, 2007.

Lung-fei Lee, Xiaodong Liu, and Xu Lin. Specification and estimation of social interaction models with network structures. The Econometrics Journal, 13:145–176, 2010a.

A. Mas and E. Moretti. Peers at work. American Economic Review, 99(1):112–145, 2009.

Michael Maschler, Eilon Solan, and Shmuel Zamir. Game Theory. Cambirdge University Press, 2013. Translated from Hebrew by Ziv Hellman.

Jacob A. Mincer. Schooling, Experience, and Earnings. New York: NBER Press, 1974.

Jacob A. Mincer and Solomon Polacheck. Economics of the Family: Marriage, Children, and Human Capital, chapter Family Investments in Human Capital: Earnings of Women. University of Chicago Press, 1974.

John Nash. Two-person cooperative games. Econometrica, 21(1):128–140, 1953.

R. A. Pollack. Bargaining power in marriage: Earnings, wage rates and household production. NBER Working Paper, (11239), 2005.

Habib Some. Modelling and Estimating Models of Physician Labour Supply and Productivity. PhD thesis, Université Laval, 2016.

Ravi Varadhan. alabama: Constrained Nonlinear Optimization, 2015. URL https://CRAN.R-project.org/package=alabama. R package version 2015.3-1.

Jane Waldfogel. The effect of children on women's wages. American Sociological Review, 62(2):209–217, 1997.

WHO. Making pregnancy safer: the critical role of the skilled attendant: A joint statement by WHO, ICM and FIGO, 2004.

WHO. Trends in maternal mortality: 1990 to 2010,. World Health Organization, 2012.

Chapter 3

Listening with Cognitive Biases within a Network

Elfried Faton

3.1 Résumé

Dans cet article, j'aborde la question de la formation d'opinion au sein d'un ensemble d'individus reliés par une structure de réseau, lorsque ceux-ci ont des biais cognitifs. En particulier, j'étudie le comportement des individus lorsque leurs croyances sont extrêmes, ou se rapprochent davantage des extrêmes. Deux biais cognitifs principaux sont explorés dans le modèle : le biais de confirmation et le biais de supériorité relative des extrémistes. D'après mes résultats, même à partir d'un réseau *a priori* fortement connecté et apériodique, les croyances ne convergent pas nécessairement vers un consensus. Spécifiquement, je précise quelques caractéristiques de réseaux et de croyances initiales qui influencent la probabilité de consensus ; et je donne quelques prédictions sur la borne supérieure de la probabilité de consensus en groupe. De plus, je propose une nouvelle explication aux éventuels mécanismes qui sous-tendent des enjeux sociaux tels que le radicalisme politique, certains comportements extrémistes, ainsi que la non-convergence d'opinions dans un réseau d'individus.

Mots clés : formation des croyances, biais cognitifs, consensus, extrémisme, apprentissage, réseaux sociaux.

Codes JEL : D83, D85, D91, Z13

0. I thank my thesis co-advisor, Vincent Boucher, for extensive counsel and guidance throughout this project. I am also grateful to Roland Pongou and Gilles Koumou for their comments. I thank participants at the 49th congress of the *Société Canadienne de Science Économique* (SCSE), participants at the 2018 Africa meeting of the Econometric Society, and participants at *Jeudi-midi* seminars (Université Laval), for their useful comments and suggestions. However, I am solely responsible for any remaining errors.

3.2 Abstract

In this paper, I give a new theoretical insight on opinion formation among agents, subject to some cognitive biases within a social network. In particular, I study the behavior of agents near and toward extreme beliefs. Two cognitive biases are considered in the model: a confirmation bias and an extremist relative superiority bias. I find that, in an *a priori* strongly connected and aperiodic network, beliefs do not necessarily converge to a consensus. Specifically, I identify some characteristics of a network and initial beliefs influencing the probability of consensus, and I predict an upper bound of the probability of consensus in groups. Also, in the paper I propose a new understanding of the plausible mechanisms behind social issues like political radicalism, various extreme behaviors and the non-convergence of opinions within a network.

Keywords: belief formation, cognitive bias, consensus, extremism, learning, social networks.

JEL codes: D83, D85, D91, Z13

3.3 Introduction

The expansion of extreme or radical beliefs is a question of great interest in our societies often subject to ceaseless news on terrorism, religious extremism, political radicalism, and other extreme behaviors. While analyzing the process of political radicalism (or de-radicalism) for individuals who take part into terrorist plans or attacks, several experts psychologists reject the theory stating that most of these individuals have psychological health issues or that they are psychopaths (Mullins and Dolnik (2010), Webber and Kruglanski (2017)). Specifically, psychologists acknowledge the influence of networks in explaining radicalism (Mullins and Dolnik (2010)). In various domains like crime, employment, education, health and others, several recent papers have shown the undeniable role played by networks and social interactions on behaviors (see Drago and Galbiati (2012), Boucher et al. (2014), Chandrasekhar et al. (2015), Fortin and Yazbeck (2015), Faton and Fortin (2017)). Likewise, political radicalism is not an exception (Wintrobe, 2006). Consider the following example.

Example 3.1. A population of inmates within a prison is such that a few prisoners are radicalized, and hold extreme political, religious or social beliefs. Let's say we know the listening structure inside the prison. That is, we know exactly who listens to whom, while interacting. Assume again that prisoners are semi-complex thinkers, subject to some cognitive biases which influence their listening behavior; so that they can increase or decrease their attention to co-inmates.
Under what conditions (on the network), may some prisoners become (de)radicalized (or not)? What network structures are more (or less) favorable to consensus at extreme beliefs? These are examples of questions answered in this paper.

In this paper, I explore the formation of extreme (or radical) beliefs among agents within a network. Specifically, I study the influence of some cognitive biases on the process of belief formation in the network, and the underlying mechanisms through which extreme beliefs can prevail over moderate ones, over time. First, I introduce an intuitive approach for beliefs updating in a network, where individuals are subject to some cognitive biases, which can affect the strength of ties in the network over time. Second, I discuss properties for the convergence of beliefs and for reaching consensus in the network. In particular, I address cases where beliefs converge to some extreme value, interpreting extremism. I find that, accounting for cognitive biases can modify the path and the speed to consensus. Especially in a strongly connected network, I find that they can even affect the very existence of a consensus; which is not expected at first. In that respect, I argue about a possible polarization and fragmentation of beliefs in a strongly connected network.

More specifically, I consider two main sources of cognitive bias in the paper: a *confirmation bias*, and an *extremist relative superiority bias*. The first refers to this natural inclination that people often have, of listening more to those whose beliefs are closer to theirs (see Lord et al. (1979), Nickerson (1998), Rabin and Schrag (1999), McFadden and Lusk (2015), Charness and Dave (2017)). It can also, to some extent, be identified as homophily in beliefs (see Golub and Jackson (2012), Boucher (2015) on homophily). The second is based on the results of authors like Toner et al. (2013) and Brandt et al. (2015), which motivate the idea that people tend to listen more to themselves, and consequently less to others, as their beliefs become more extreme. Following these authors, I use the term *extremist relative superiority* bias (*ERS bias* hereafter), and shortened as extremist superiority bias in the remaining of the paper. Thus, combining these two sources allows my model to capture some dynamics usually observed in opinion formation in networks, which have not yet been expressly considered in the literature on belief formation. Therefore, unlike the naive setting of DeGroot (1974) in which people's listening attention to peers stays constant over time, accounting for cognitive biases results in a more intuitive framework in which people can update their attention to these peers over time.

In the paper, I propose a simple model in which, at first, agents have a prior belief on a matter of interest; let's say their position on a social or political matter. Consider for instance the case of detainees who have entered a prison with a prior belief about the pertinence of using violence to achieve a crime.

Second, while in prison, these agents interact with other prisoners. Regarding the example 3.1 above, prisoners interact with their peers in their cells, at meal time, or at any other interaction moments within the prison. Then, these agents form a prior opinion on who they should listen more, or less (or not) to. This part of the process is subjective to the agent, and does not depend yet on the beliefs of the others. In other words, a prisoner may decide *a priori* to listen (or not) to someone, based on their physical attribute (e.g., race, physical shape, tattoos), or because of their tastes and preferences. All these degrees of listening attention are then quantified by weights, and normalized in a row-stochastic square matrix of all individuals in the network. That matrix represents the structure of the network *a priori*, and is called *a priori* listening matrix throughout the paper. In the paper, the *a priori* matrix is considered given.

Third, at each period, agents communicate their beliefs to others in the network and simultaneously update their beliefs and the listening attention weights they assign to their peers and themselves. To update their beliefs, agents use a simple average-based process. At each period, they take a weighted average of all beliefs in the preceding period, using the values of the listening weights they assign to their peers, in that previous period. At the same time, the listening matrix is also updated at each period, through a process detailed later in the paper, and which considers agents cognitive biases. Then, these steps are repeated over time.

The average-based beliefs updating process used in this paper is a general model of opinion formation, rooted in precursor studies like DeGroot (1974) and DeMarzo et al. (2003). Several prior studies on learning in social networks, have shown the adequacy of an average-based updating rule. Among these studies further detailed in section 3.4.2, DeMarzo et al. (2003) and Golub and Jackson (2012) explain a few good properties of this rule. Moreover, experimental studies like Chandrasekhar et al. (2015), also show its best fitness to their data, over a Bayesian learning setting. Henceforth, although more refined than the simple DeGroot (1974) setting, my model still holds its good properties of simplicity and tractability, and the nice intuition behind the average-based rule. Moreover, the introduction of cognitive biases in the process does add some interesting twist into the analysis. In fact, although there is no new link formation in the model, the sole presence of cognitive biases, is likely to alter the structure of the network.

The main contribution of this article to the literature is twofold. First, the paper contributes to the already existing literature on opinion formation, by considering both an extremist superiority bias and a confirmation bias in a social network analysis. In fact, even outside the political spectrum as in Toner et al. (2013), Brandt et al. (2015), we usually observe that *ceteris paribus*, individuals with radical ideologies tend to be more self-confident in their views than people with more moderate beliefs. Hence, my interest in considering the extremist superiority bias in a study on extreme opinions formation. The second contribution of the paper is to introduce an intuitive theory which justifies the severance of ties in networks, as it is often observed in practice. The model developed in this paper lays down some theoretical groundwork on how ties can be severed, over time. As a result, I present a new refinement on the conclusions on the convergence of beliefs and consensus in a strongly connected and aperiodic network.

Besides, I use some simulated data on small groups, to predict the probability of consensus, given the *a priori* degree of polarization in the network, and its connectivity. Moreover, I implement some comparative statics presenting the incidence of cognitive biases on the existence and the value of a limiting belief (after an infinitely long time) for *a priori* networks. In particular, I present convergence results in a network, in the case where the level of extremism *a priori* of one agent varies.

I show that the rate of convergence of beliefs to a consensus (if any), and to the consensus value (if any), vary according to the initial beliefs of individuals, and the degree of polarization in the network. Specifically, I find that a stronger polarization in beliefs *ceteris paribus* leads to slower convergence time and vice versa. In the worse cases, it can even impede the existence of a limiting belief (under aperiodicity in the initial network), and consensus, even for an *a priori* strongly connected network.

Furthermore, in my general model, if convergence is obtained, I identify a sufficient condition for consensus. That is, over time there exists a moment from when everyone in the network who initially listens to people other than themselves, monotonously decrease their self-confidence. This is also a key result of this paper. That is, consensus among individuals who hold different beliefs, requires that at some point everyone who is not strongly attached to their (own) beliefs only, should start to "let go," and let themselves be convinced by others.

In addition, in the presence of confirmation bias only (i.e., self-confidence is null), I find that consensus is not always guaranteed, even in a complete network like a *group.* [1] In fact, when the initial beliefs are drawn very close to the extreme values, and form two opposite clusters, consensus is more difficult to reach. Beliefs, in that case are more likely to stay polarized, although the clusters may become significantly closer. However, not only are these predictions affected by the size of the network, but also introducing a non-partisan moderator in the network changes some convergence results. [2] While exploring convergence to extreme beliefs, besides discussing group polarization, I derive some conditions for consensual extremism.

The rest of the paper is organized as follows. Section 3.4 explains the foundations of my model and how beliefs are formed throughout time. In section 3.5, I discuss convergence

1. In a group, everyone is directly influenced by everyone else. It is the most strongly connected and aperiodic form that a network of individuals can be.

2. A non-partisan moderator, as defined in this paper, is an individual who consistently listens to peers with distinct beliefs, in a manner which guarantees that both distinct peers (or groups of peers) keep listening to them (moderator) over time, and do not disconnect from the rest of the network.

and consensus properties and results of the model and explore some practical cases of convergence to extreme beliefs in a given network. In that section, I also propose a model to predict the probability of consensus in a network, and give some results for small groups. Then, in Section 3.6, I discuss some general predictions of the model, and Section 3.8 concludes.

3.4 A Simple Model of Belief Formation

Consider an individual i who has to form their belief on a certain matter for which a true state of the world exists. i receives a signal about the true state, yet its value is unknown to them. This signal allows i forming an *a priori* belief on the true state. Then, i and a finite number $n-1$ of other individuals, who have also received a signal on the true state and formed their *a priori* belief, are put together to create a network. Then, based on their own subjectivity (or lack of subjectivity) about their peers observable characteristics, each individual forms another *a priori* on how much they should listen to each of their peers.[3] In the remainder of the paper, the term *listening weight* represents a quantity expressing how much an individual listens to a given peer, relatively to their other peers. Both *a priori* beliefs and listening weights are formed at the beginning of the time horizon, arbitrarily fixed at time $t = 0$. To illustrate this, consider the following matter on which some people are to form their beliefs:

What do you believe is the probability that Trump runs for re-election and wins?

Then, assume the signal they receive about the true state of the world is an approximation of the percentage of all U.S. news articles, which predict that Trump will win during their future presidential elections. Therefore, each individual i could form their *a priori* belief on Trump's potential re-election. Note here that individuals own characteristics and subjectivity can also influence them in the formation of these *a priori* beliefs.[4] Next, all the individuals are put in contact with one another and they get to know some of their peers' characteristics. For instance, they can notice or learn about one another's gender, occupation, social status, skin color, political affiliation, or the

3. Recall here that a network shows how much everyone listens to their peers.

4. For instance, someone who is strongly against Trump's politics, may not be as much influenced by these statistics as any random individual. Similarly, someone who prefers Trump as president, may predict a win for Trump, even if the signal, here the journals' statistics, were not in his favor. In fact, because their own opinion represents a more reliable information for them, they may possibly hold a belief that is, ultimately, independent of the signal.

TV channels they listen more to; to cite a few potential observable characteristics.[5] Then, because of their subjectivity about their peers' observable characteristics each of them will assign an *a priori* listening weight to the beliefs of each of these peers.[6] Therefore, they may decide not to listen to someone (weight equals zero), to listen to them to a certain extent (weight higher than zero but still less than one), or to listen to only one individual (weight equals 1).[7]

At time $t = 0$, an individual$i$'s *a priori* belief is noted b_i^0, and the weight related to the listening attention they assign to a peer j is noted w_{ij}^0. As a general rule, I note b_i^t the belief of individual i at time t, and w_{ij}^t the listening weight that i assigns to a peer j, at time t. At time t, if i listens to j, then $w_{ij}^t > 0$.[8] Following DeMarzo et al. (2003), w_{ij}^t also represents the *direct influence* of individual j on individual i at time t.

I focus on what happens from the arbitrary period $t = 0$ and onward. After forming their *a priori* beliefs and listening weights, everyone communicates their belief to everyone in the network. I denote $\mathbf{b}_0$ the vector (of size $n \times 1$) of all individuals *a priori* beliefs. Similarly, the vector of individuals' beliefs at time t is denoted $\mathbf{b}_t$. I assume $\mathbf{b}_0$ given. At each period, while individuals communicate their beliefs to one another through discussions; they update their beliefs.[9] The belief updating is a repetitive process that happens over time, after interactions with their peers; until their belief converges to a fixed value, or almost identical to that of the peers they listen to.[10] In

5. Let me precise here that, these characteristics should stay constant during the time horizon we consider. This is to avoid a more complex scenario where listening weights are likely to change according to the changes in individuals observable characteristics, on top of the potential changes due to the cognitive biases to be accounted for.

6. For instance, Algan et al. (2019) study friendship formation among Science Po students and find that homophily in political opinions, gender, academic background (admission category), high school major and family income explain friendship ties.

7. For instance, a fervent follower of Trump may decide not to listen to someone who listens to CNN. An individual who has racial prejudice may decide not to listen to someone because of their physical traits. Another person with gender bias may decide to listen less to someone of a certain gender than the other. In the example 3.1 mentioned in the introduction, let's say some prisoners are petty criminals who desire to win the favor of more "successful" or notorious criminals. Then, some of them (not necessarily all) may decide to listen more to these infamous criminals than they do any other prisoner.

8. Using some notations from graph theory, let $\mathcal{N}_t = (V(G_t), E(G_t))$ the associated network represented by a digraph G_t, where $V(G_t) = \{1, 2, ..., n\}$ denotes the set of all individuals (also called vertices or nodes), and $E(G_t) = \left\{(i, j) \in V(G_t) \times V(G_t)/w_{ij}^t > 0\right\}$ the set of all directed links (edges) in the network. Then, in a network $\mathcal{N}_t$, each positive listening weight ($w_{ij}^t > 0$) is represented by a directed link from i to j in the associated digraph G_t.

9. Through this framework individuals implicitly try to guess the true state of the world, as close as possible. However, they can never know for sure if their guess is right. Therefore, if the beliefs of some peers they listen to are different from their own, they may infer that their current belief is somehow wrong and keep updating their belief based on the new information received.

10. Note that the condition to update their belief is that a peer to whom they assign a strictly positive weight to, differs from theirs. This also means that, at a given period t, they do not care about the beliefs of peers who do not have any direct influence on them ($w_{ij}^t = 0$).

particular, here, the formation of new beliefs (updating) solely depends on the listening weights individuals assign to peers and everyone's belief in the preceding period. The precise rule of beliefs updating is average-based, as described in section 3.4.2. All *a priori* beliefs b_i^0 and all *a priori* listening weights w_{ij}^0 are given. At each period, the listening weights are private information. Yet, following the successive interactions and discussions, individuals may decide to listen more (or less) to some of their peers. In the model, their adjustment of the listening weights is based on a *confirmation bias* and another bias called *extremist relative superiority bias* in this paper.[11] Section 3.4.3 describes the exact updating rule for the listening weights.

The *confirmation bias* is accounted for through the similarity that "attracts" people to one another, which is measured in terms of the distance (absolute value) between their beliefs. The smaller this distance, the more these peers are listened to.[12] As for the *ERS bias*, I use the distance to the middle belief to measure how extreme individuals' beliefs are.[13] Next, I define various levels of extreme beliefs, and precise a definition domain for all beliefs in my model.

3.4.1 Beliefs domain and extreme beliefs

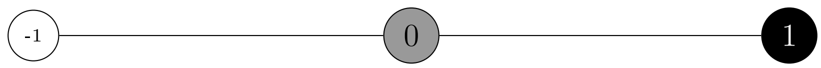
Without any loss of generality, I assume beliefs are quantifiable with values in $[-1, 1]$. The value -1 may for example represent extreme-left beliefs; or in politics, it could also be identified to a left-wing belief. Whereas 1 represents extreme-right beliefs or a right-wing belief in politics. The scalars 1 and -1 are chosen for simplicity. For any matter for which people have to hold views and beliefs, the set of beliefs can be easily generalized to $[-1, 1]$. Then, $\mathbf{b}_t = (b_1^t, ..., b_n^t)'$ the vector of all individuals' beliefs in the network at time t is defined on $[-1, 1]^{n\times 1}$. In the example above where people have to tell what they think is the probability that Trump wins if he were to run again for president in 2020, their answers should fall within a range of 0 to 1. However, knowing there always exists a bijection of $[0, 1]$ into $[-1, 1]$, it is possible to rescale their answers

11. Toner et al. (2013) propose the term *belief superiority* to refer to the fact that, for a given issue (political, religious, etc.), an individual would be more confident in his belief as more correct comparatively to others' beliefs. In their paper, they conduct an experiment, and show that people with extreme political views (conservatives or liberals) tend more to believe that their belief is the right one, than moderates or non-partisans. This general attitude of extremists relative belief superiority has been replicated in several other studies in various domains including environment(Toner Raimi and Leary (2014)) and religion (Hopkin et al. (2014)).

12. In the prison network above, let there be a criminal i who, in a given period t, listens equally to criminals j and k ($w_{ij}^t = w_{ik}^t$). Then, i receives information about the beliefs of j and k at period t. If $|b_i^t - b_j^t| > |b_i^t - b_k^t|$, then $w_{ij}^{t+1} < w_{ik}^{t+1}$. In other words, in the next period $t+1$, i will listen more to k than they do j. Note that, this does not necessarily mean that $w_{ik}^{t+1} > w_{ik}^t$.

13. The middle belief refers to the belief which is equidistant from both extremes.

such that 0 is now the extreme-left belief.[14] The following voting scale represents well the space of beliefs limited by two extremes.



For more clarity, I define a concept extensively used in remaining of the paper, and referring to different levels of extreme belief as follows:

Definition 3.1. *ϵ-extreme belief*
Let $\epsilon \in (0, \frac{1}{2})$ a scalar. An individual i is considered to have ϵ-extreme belief at time t if their belief b_i^t is such that $|b_i^t| \geq 1 - \epsilon$.
In the case where $|b_i^t| \geq 1 - \epsilon \ \forall \ \epsilon \in (0, \frac{1}{2})$, that is $b_i^t \in \{-1, 1\}$, beliefs are said to be (simply) extreme.

In other words, an ϵ-extreme believer (or someone with ϵ-extreme beliefs) is someone who is closer to an extreme belief than the middle belief 0. This definition will be more useful later. According to this, any belief valued in the interval $[-\frac{1}{2}, \frac{1}{2}]$ is never an ϵ-extreme belief. In the remaining of the paper, any individual with such belief is called a moderate believer, or as someone with moderate beliefs.[15] Then, $V_{\epsilon,t}^{-} = \{i \in V(G_t)/b_i^t \leq -1 + \epsilon\}$ and $V_{\epsilon,t}^{+} = \{i \in V(G_t)/b_i^t \geq 1 - \epsilon\}$ denote respectively, the subsets of all left ϵ-extreme believers and right ϵ-extreme believers of $V(G_t)$. In the case $\epsilon \in [0, \frac{1}{2})$, the notations $V_{\epsilon,t}^{-}$ and $V_{\epsilon,t}^{+}$ also include respectively, all left extreme believers and right extreme believers.

3.4.2 Beliefs Updating Process

To update their belief at a period $t + 1$, individuals use the information they receive from their peers (beliefs) at period t and their own subjectivity. It is a dynamic, yet deterministic process. Beliefs in period 1 only depend on beliefs and listening weights in period 0; beliefs in period 2 only depend on quantities in period 1 and so on.

14. In the specific case of any multidimensional issue, where individuals' answers are each represented by a k-dimensional vector of values in $\mathbb{R}^k$, the model requires that, at first, all beliefs are mapped to the segment $[-1, 1]$ of real numbers.

15. Here, let us make a quick connection with the median voter theorem. Assume all necessary rules for the median voter theorem hold. Let L and R two political parties, such that L advocates for a left ϵ-extreme belief and R for a right moderate one. If the median voter holds the middle belief 0, then the moderate party wins the election.

Let $\mathbf{W}_t = \left(w_{ij}^t\right)$ denote an $n \times n$ matrix composed of all the listening weights that each individual assigns to everyone in the network at time t, including themselves. Each line i of $\mathbf{W}_t$ represents, at time t, all the listening weight that i assigns to each individual in the network, and each column j represents all the weights that each individual in the network assigns to j. The weights are nonnegative and defined such that the sum of all elements on each line equals 1. Put another way, $\mathbf{W}_t$ is row-stochastic. This condition is useful because it allows each value composing $\mathbf{W}_t$ to be comparable to any other one. In other words, for four individuals i, j, k, and l, $w_{ij}^t > w_{kl}^t$ means that i listens more to j than k does to l. In addition, as shown in appendix C.1.3, the existence of a consensus in the model, as defined later in the paper, requires that $\mathbf{W}_t$ is row-stochastic. Thus, $\mathbf{W}_t \in [0,1]^{n \times n}$, and the matrix sequence $(\mathbf{W}_t)_t = (\mathbf{W}_0, \mathbf{W}_1, ..., \mathbf{W}_t, ...)$ is set to follow a Markov process as described in equation 3.4.

As aforementioned, the vector of *a priori* beliefs and the matrix of *a priori* listening weights are given. Therefore, given $\mathbf{b}_0$ and $\mathbf{W}_0$ two matrices of *a priori* beliefs and weights, for any period $t \geq 0$, the vector of beliefs at time $t+1$ is defined as:

$$\mathbf{b}_{t+1} = \mathbf{W}_t \mathbf{b}_t \quad \forall t \geq 0 \tag{3.1}$$

This model implies that, knowing the beliefs of everyone in the network at a given time t, everyone forms their updated belief by taking a weighted average of all beliefs, using the weights they formed at period t. A main advantage of this average-based rule is that it is tractable. Using equation 3.1 iteratively, $\mathbf{b}_{t+1}$ is reduced to an expression composed only of all the intermediate matrices $\mathbf{W}_m$ $(m = 0, 1, ..., t)$ and the *a priori* belief vector $\mathbf{b}_0$ as following:

$$\mathbf{b}_{t+1} = \mathbf{W}_t \mathbf{W}_{t-1} ... \mathbf{W}_0 \mathbf{b}_0 = \prod_{m=0}^{t} \mathbf{W}_{t-m} \mathbf{b}_0 \tag{3.2}$$

The particular case of the model 3.2, in which each matrix $\mathbf{W}_m$ of the sequence is identical to $\mathbf{W}_0$, equates the naive learning model of DeGroot (1974):

$$\mathbf{b}_{t+1} = \underbrace{\mathbf{W}_0 ... \mathbf{W}_0}_{t \ times} \mathbf{b}_0 = \mathbf{W}_0^t \mathbf{b}_0 \tag{3.3}$$

Several papers in economics advocate for the use of such a simple average-based rule. DeMarzo et al. (2003) argue that the average-based rule is consistent with their boundedly rational theory, which focuses on individuals subject to persuasion bias. Acemoglu et al. (2013) on the other hand, under the label *inhomogeneous stochastic gossip process*, use this rule to model beliefs evolution in a society composed of what they call *stubborn* and *regular* agents. In addition, in the case of a pure coordination game, Golub and Jackson (2012) interpret this rule as a myopic best response updating rule. Furthermore, several papers use the naive DeGroot specification to model learning in experimental networks (Mueller-Frank and Neri (2013), Chandrasekhar et al. (2015)). These studies usually promote this model over its rather sophisticated, Bayesian alternative. As a matter of fact, while making a comparison with a Bayesian learning model, Chandrasekhar et al. (2015) use an experiment to show that agents learning behavior in networks are more robustly described by the DeGroot model (also called naive learning model).

A key argument in favor of the average-based rule is that agents do not usually make such complex calculations as in the Bayesian setting. The DeGroot model appears to be intuitive enough and fits best the attitudes of individuals during these experiments. Another main advantage of the DeGroot model resides in its simplicity. In that model, knowing only the initial beliefs and listening weights, beliefs can be easily determined at any period, and the conditions for convergence of beliefs and for reaching a consensus are simple to derive. The model is thus tractable enough, and allows finding clear conditions for the wisdom of crowds (Golub and Jackson (2010)).

However, while taking advantage of the good properties of an average-based rule, it is also important to account for individuals cognitive biases. Provided that they can affect the attention individuals assign to their peers and themselves over time, they may impose a particular dynamic to $\mathbf{W}_t$, usually ignored in the naive model.[16] To complete the definition of the model, in the next section, I describe the matrices updating process for all periods consecutive to the initial time $t = 0$.

3.4.3 Updating the listening matrix

At each period $t \geq 0$ and for any pair of individuals i and j in the network, we have:

Assumption 3.1. *Cognitive biases*

16. Note that, in the case that I allow for a non-empty subset of uninformed agents in the model 3.2 above, it would also become, to some extent, a generalization of the *generalized DeGroot* (GDG) model of Banerjee et al. (2016). However, to keep a certain level of simplicity in the current paper, I do not explicitly model the possible presence of uninformed agents.

(i) Confirmation bias: all else being equal, in the next period $t+1$, *people listen more to those whose beliefs were closer to their own. This is equivalent to:* $\forall\, t \geq 0,\ \forall\, i,\ j \in V(G_t),\ \dfrac{\partial w_{ij}^{t+1}}{\partial |b_i^t - b_j^t|} < 0.$

(ii) Extremist relative belief superiority: all else being equal, in the next period $t+1$, *people whose beliefs at period* t *are closer to the extremes, listen more to themselves than those who are farther. Put another way, ceteris paribus, people's self-confidence increases as their beliefs are closer to the extremes. This means:* $\forall\, t \geq 0,\ \forall\, i \in V(G_t),\ \dfrac{\partial w_{ii}^{t+1}}{\partial |b_i^t|} > 0$

Based on these assumptions, I propose a functional variation of each component of $\mathbf{W}_t$ as following:

$$w_{ij}^t = \begin{cases} \dfrac{w_{ij}^{t-1} f(|b_i^{t-1} - b_j^{t-1}|)}{\sum_{k=1,k\neq i}^{n} w_{ik}^{t-1} f(|b_i^{t-1} - b_k^{t-1}|) + w_{ii}^{t-1} g(|b_i^{t-1}|)} & \text{if } \ i \neq j \\[2ex] \dfrac{w_{ii}^{t-1} g(|b_i^{t-1}|)}{\sum_{k=1,k\neq i}^{n} w_{ik}^{t-1} f(|b_i^{t-1} - b_k^{t-1}|) + w_{ii}^{t-1} g(|b_i^{t-1}|)} & \text{if } i = j \end{cases} \tag{3.4}$$

where the functions $f : x \mapsto f(x)$ and $g : x \mapsto g(x)$, are such that:

1. $f()$ is continuous and strictly decreasing over $[0, 2]$ such that $f(0)$ is positive (can tend to $+\infty$, but not necessarily) and $f(2) \to 0$ ($f(2) \neq 0$).
   The objective here is to make sure that, even if at $t-1$ two individuals receive the same attention, because of homophily, the one whose belief is closer will receive more attention next. Thus the decreasing $f()$. $f(2)$ should tend to zero to magnify the effect of homophily. This allows individuals to lower drastically the attention they give someone, when they find that their belief is at the opposite extreme of their own belief which is also extreme. I impose the condition $f(2) \neq 0$ to ensure that an existing link is not systematically broken, only because individuals realize that their peer's belief is opposed to their own. In theory, the updated network structure as defined here should remain the same and only the values assigned to its links (weights) could change over time. However, links can break, asymptotically, after their associated weight becomes too small. Besides, I acknowledge that intuitively if a listening weight becomes too small, we might as well assume its associated link is broken.[17] For simulations purpose, I use a simple form of f as follows: $f : x \mapsto -x + 2 + \xi$, with $\xi > 0$ a very small real number (close to 0).

17. Consider a network of individuals. Let's say that *a priori*, an individual i assigns 90% of their attention to another individual j. But then right from the start individual i realizes that individual j is 100% for Trump's

2. $g()$ is continuous and strictly increasing over $[0, 1]$ such that $g(0) \to 0$ ($g(0) \neq 0$), and $g(1)$ is positive (could also tend to $+\infty$). $g()$ is actually the function that carries the properties in assumption 3.1*(ii)*. Consider two individuals, who initially have the same amount of self-confidence, meaning that they listen to themselves the same way. Then, in the following period, the one whose belief is closer to an extreme will, have more self-confidence than the other. And here also, someone who had a positive amount of self-confidence at first should keep listening to themselves even slightly. However, asymptotically w_{ii}^t can tend to zero.[18] It is possible to define f and g such that $f(2) = g(0)$ and $f(0) = g(1)$, however it is not necessary. In my simulations, I use $g : x \mapsto 2x + \xi$.

From equation 3.4, the sequence $(\mathbf{W}_t)_t$ is continuous and each matrix composing the sequence is row-stochastic and nonnegative. Now that the model is fully defined, let us focus on another main goal of the paper. This paper aims at analyzing various networks situations where all individuals' beliefs converge to a unique one. When this happens in a network, we say that people have reached a consensus. In the DeGroot (1974) model, for any vector $\mathbf{b}_0$ of *a priori* beliefs, consensus is obtained starting from a period when the matrix product $\mathbf{W}_0^t$ becomes ergodic. Conditions for that are detailed in the paper of Golub and Jackson (2010). However, in my model defined by equations 3.2 and 3.4, these conditions do not guarantee consensus. In the next section 3.5, I give more details about convergence and consensus properties for my model.

3.5 Reaching Consensus

A consensus is reached in a network when everyone's beliefs becomes the same. In this paper I'm particularly interested in networks which favor the shift of individual beliefs to extreme beliefs and those where extreme believers manage to keep or lose their extreme beliefs. In fact, not all network structures are favorable for the expansion of

re-election while i is 100% against. Usually in reality, i will not instantly decide to stop listening to j forever. Indeed, i's attention to j may drastically be affected (lowered). However, for most people, the process that leads them to stop listening to someone is not sudden in general. Ultimately, if the gap between i and j never tends to close itself, they can reach a point where their attention towards one another becomes so small that it seems nearly non-existent. And surely, when that happens the real network structure at that time should change such that the link between i and j will be missing. In absolute terms, the link remains. However, intuitively, the direct influence of j on i is "too small" for j to still have any real effect on i. Thus, the pattern displayed by the network then is more consistent with a network structure that is different from the initial one. This issue will become clearer later in the simulations.

18. Referring to the same example, let say prisoners i and k self-confidence at time t amounts to 10% each, and their beliefs are respectively quantified $b_i^t = -1$ and $b_k^t = .5$. Then i clearly holds a more extreme beliefs than k. Consequently, at $t+1$, i's self-confidence will become greater than j's. Note here that I'm comparing w_{ii}^{t+1} and w_{kk}^{t+1}, not w_{ii}^{t+1} with w_{ii}^t or w_{kk}^t. This means that depending on the other parameters in the network, and according to this example, we can have $w_{kk}^{t+1} < w_{ii}^{t+1} < .1$.

extreme beliefs. In this section, I discuss some basic characteristics of a network *a priori* and a vector of beliefs *a priori*, that are required for beliefs to possibly reach a consensus. Before defining these conditions in my model, here are some useful definitions on concepts related to networks' structures that I use.

Definition 3.2. *Walk*
In a network, a walk from an individual i, to another one j is a sequence of individuals j_1, j_2, to j_k such that i listens to j_1, j_1 listens to j_2, $\dots$, j_{k-1} listens to j_k, and j_k listens to j. The length of such a walk is $k+1$.

Note that in the case i listens to j, it is a walk of length 1.

Definition 3.3. *Path*
A path from i to j is a walk consisted of a sequence of distinct individuals.

Definition 3.4. *Cycle*
A cycle is a path starting and ending with the same individual.

Definition 3.5. *Connected network*
A network is connected if for any pair (i, j) of individuals, a path from i to j or from j to i exists.

Consider a listening matrix $\mathbf{W}_t$ represented by a graph G_t. It represents a connected network if for any pair (i, j) of individuals, either there exist m_1 individuals $k_1, ..., k_{m_1}$ ($\in V(G_t)\backslash\{i\}$) such that w_{ik_1}, $w_{k_1k_2}, ...,$ $w_{k_{m_1}j}$ are all positive, or there exist m_2 individuals $k_1, ..., k_{m_2}$ ($\in V(G_t)\backslash\{j\}$),such that w_{jk_1}, $w_{k_1k_2}, ...,$ w_{k_mi} are all positive (see an example in figure 3.1d). In the case $w^t_{ij} > 0$ ($m_1 = 1$), I say that "i listens directly to j". Otherwise, if w_{ik_1}, $w_{k_1k_2}, ...,$ $w_{k_{m_1}j}$ $\in]0, 1[$, I use the expression "i listens remotely to j". This also means that i is *influenced* by j. The notion of *peer influence* is further defined in section 3.5.1.

Definition 3.6. *Weakly connected network*
A network is weakly connected if for any individual i, there exists an individual j ($\neq i$) such that a path from i to j exist.

In this case, there is no isolated individual i with $w^t_{ii} = 1$; and there may exist two individuals who do not listen directly, nor remotely to themselves. That is to say, they do not influence each other. An example in shown in figure 3.1c. Note that, the corresponding undirected graph of a weakly connected network $\mathcal{N}_t$ is also connected (see figure 3.1f).

Definition 3.7. *Strongly connected network (Gross and Yellen, 2006)*
A network is strongly connected if for any pair (i, j) *of individuals there is a walk from* i *to* j *and from* j *to* i*.*

In other words, a strongly connected network refers to a network where, for any pair (i, j) of individuals, i listens directly or remotely to j, and j also listens directly or remotely to i. That is, everyone in the network is somehow influenced by everyone else (see figure 3.1e).

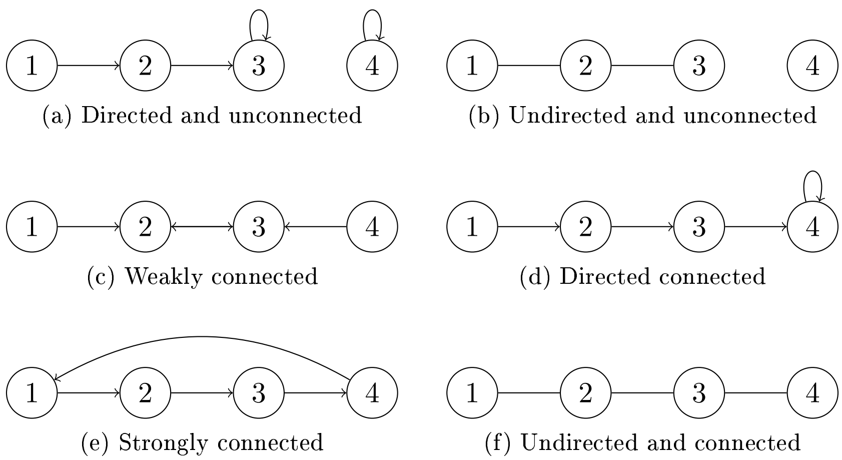


Figure 3.1 – Levels of connectedness in a network

In summary, for directed networks, a clear distinction exists between a weakly connected network, a connected network and a strongly connected network. If a directed network is strongly connected, then it is connected. If it is connected, then it is weakly connected. But the opposites are not true. However, for the undirected version of such networks' graph, a distinction of the level of connectedness no longer exists. An undirected network is either connected or not. To illustrate this, I represent the undirected version of the networks in figures 3.1c, 3.1d and 3.1e in figure 3.1f. Figure 3.1b however, represents the undirected version of the network in figure 3.1a. All these definitions on various levels of connectedness within a directed network are more useful later in the paper.

Definition 3.8. *Convergence of beliefs versus consensus*
Beliefs updated through equation (3.1) converge if there exists a vector $\bar{\boldsymbol{b}} \in [-1, 1]^{n\times 1}$

such that, given $\boldsymbol{b}_0$ *a vector of initial beliefs and* $(\boldsymbol{W}_0, \boldsymbol{W}_1, ..., \boldsymbol{W}_t, ...)$ *a time-varying sequence of listening matrices:*

$$\lim_{t \to +\infty} \boldsymbol{W}_t \boldsymbol{b}_t = \bar{\boldsymbol{b}} \tag{3.5}$$

In particular, if there exists a scalar b *such that* $\bar{\boldsymbol{b}} = b\boldsymbol{i}$*, where* $\boldsymbol{i}$ *is the* $n \times 1$ *unit vector; then consensus is reached, and* $\bar{\boldsymbol{b}}$ *is a consensus vector.*

In plain words, convergence refers to the fact that there exists a moment, from which individual beliefs tend to become constant and reach a steady state. Whereas consensus imposes that, all beliefs tend to become and stay the same starting from some moment in the time horizon.

Example 3.2. *Consider the following unconnected two-individuals trivial example, using the identity matrix:*

$$\mathbf{W}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \mathbf{b}_0 = \begin{bmatrix} -.5 \\ .5 \end{bmatrix}$$

The belief sequence generated by such *a priori* network (clearly) converges to the *a priori* beliefs vector$\mathbf{b}_0$. However, for any *a priori* beliefs vector $\mathbf{b}_0$ different from a consensus vector, as in the example $\mathbf{b}_0 = (-.5, .5)'$ above, consensus is never reached.

3.5.1 Convergence of Beliefs

There is a clear analogy between my model and a Markov chain, although their interpretations may differ. In fact, as it is the case for the existence of a *unique* stationary distribution of a finite-state discrete and irreducible Markov chain, there is a necessary condition in my model, without which beliefs cannot converge.[19] That requirement, defined below, is *aperiodicity* in the structure of the network. Thus, in my model, if we start with any *a priori* network $\mathcal{N}_0$, a necessary condition required for the convergence of beliefs, is the aperiodicity of $\mathbf{W}_0$. In fact, if $\mathbf{W}_0$ is not aperiodic, all subsequent matrices $\mathbf{W}_t$ are probably also not aperiodic. First, consider the following definition of

19. Note that the irreducibility of a Markov chain is somewhat equivalent to the strong connectedness condition for a network.

an aperiodic matrix and network.

Definition 3.9. *Aperiodic Network (Golub and Jackson, 2010)*
A matrix $\boldsymbol{W}$ *is aperiodic if the greater common divisor (gcd) of the length of its cycles is* 1. *Then, its associated network is aperiodic.*

To give an intuition to why aperiodicity is required for beliefs to converge, I use the concept of *influence* as introduced by DeMarzo et al. (2003) and those of *influence set* and *influence graph* mentioned in definition C.1 in the appendix.

A network that is not aperiodic is periodic, and the *gcd* of the length of all its cycles is greater than one. An intuitive way to understand periodicity of a network is that, for that type of network, some patterns repeat themselves *periodically* and indefinitely in the influence graphs (see figure 3.2). In a periodic network, one can notice that, at each period, at least one proper subset of individuals in $V(G_t)$ has an influence set which is always disjoint from the influence set of the other individuals. And in the end, it sometimes looks like some proper subsets of individuals only permute their beliefs, so that, at time goes to infinity, we can never guess who influences who in the network. This is the main reason preventing convergence of beliefs to a fixed value. In such case, it is also impossible to identify the long-term influence or the social influence of individuals in the network. In example 3.3 below, $\mathbf{W}_{per}$ shows a simple four individuals network case with four periods.
Overall, if a network $\mathcal{N}$ is aperiodic, the influence graphs do not follow a specific periodicity, but rather tend to show only one pattern (one period) after a little while (see figure 3.3 [20]). As a result, when the network stays aperiodic over time, beliefs converge and the long-term influence of each individual is computable.

Example 3.3. *Let's consider a network* $\mathcal{N}_{per}$ *defined by the following matrix* $\boldsymbol{W}_{per}$ *and represented in figure 3.2a.*

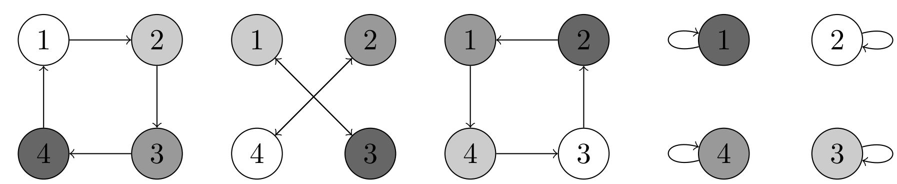
$$\boldsymbol{W}_{per} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \qquad \boldsymbol{W}_{aper} = \begin{bmatrix} a & b & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

20. As shown in figure 3.3, starting from period 3, the general structure of the influence graph stays the same. However, the values of each component may continue to change for a while, before its complete stabilization. Further in this section, I also precise some other specifications where a network like $\mathcal{N}_{aper}$ which should normally converge in the Degroot case, does not always converge in my model.

In each of the digraphs 3.2a, 3.2b, 3.2c and 3.2d, the edges indicate who (origin) is influenced by whom (destination) at the period specified. The color tone of each node indicates the belief they hold at the period mentioned. Four distinct beliefs are illustrated initially; and the figure 3.2 shows that, the network starting from $\mathcal{N}_{per}$ is periodic, and beliefs always transit successively through four values.

In this specific network, beliefs never converge. Each individual always comes back to their initial belief after four periods and never stabilize their beliefs. They keep forever permuting their beliefs between one another.

By adding an edge into $\mathcal{N}_{per}$ such that the *gcd* of all cycles becomes 1, the network obtained can be aperiodic. Usually, having a loop in a connected network is sufficient for aperiodicity. Thus, consider adding a loop at one node, as in figure 3.3. We obtain a network structure represented by the matrix form $\mathbf{W}_{aper}$ above ($a + b = 1$). The influence graphs on figure 3.3, show that there is no periodicity in the patterns displayed. After period 3, everyone is influenced by everyone else. For my model, if the *a priori* network structure is as $\mathcal{N}_{aper}$, it is very *probable* that beliefs in the network converge; and if so, consensus is reached.[21]



(a) Influence graph G_{per} of $\mathcal{N}_{per}$ after $t = 4p$, $p \in \mathbb{N}$

(b) Influence graph of $\mathcal{N}_{per}$ after $t = 4p + 1$, $p \in \mathbb{N}$

(c) Influence graph of $\mathcal{N}_{per}$ at $t = 4p + 2$, $p \in \mathbb{N}$

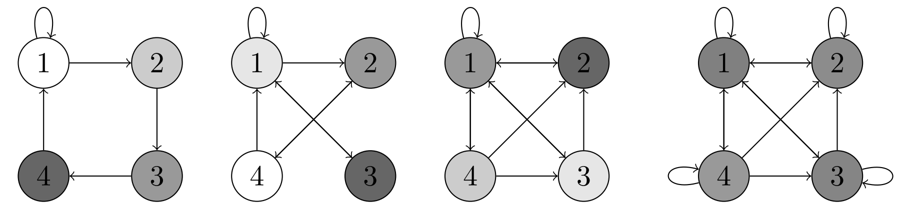
(d) Influence graph of $\mathcal{N}_{per}$ at $t = 4p+3$, $p \in \mathbb{N}$

Figure 3.2 – Periodic evolution of beliefs and influence graphs for $\mathcal{N}_{per}$

A sufficient condition for convergence in a model with non-homogeneous matrix sequence would be that *each* matrix of the sequence is aperiodic. However, because in my model, the network itself could converge to one with a different structure, a necessary condition is that the listening matrices sequence should be aperiodic *enough*. By aperiodic enough, I mean that the network should stay aperiodic as long as beliefs have not yet converged to a fixed value. Otherwise, if the matrix sequence converges faster than beliefs to a periodic matrix, the latter may never converge.

Figure C.4 in appendix shows beliefs paths for individuals in a network $\mathcal{N}_{aper}$, using

21. For the simple Degroot case, with a network as $\mathcal{N}_{aper}$ beliefs converge surely, and there is consensus.



(a) Influence graph G_{aper} of $\mathcal{N}_{aper}$ at period $t = 0$

(b) Influence graph of $\mathcal{N}_{aper}$ at $t = 1$

(c) Influence graph of $\mathcal{N}_{aper}$ at $t = 2$

(d) Influence graph of $\mathcal{N}_{aper}$ from $t = 3$ and on

Figure 3.3 – Aperiodic evolution of beliefs and influence graphs for $\mathcal{N}_{aper}$

two sets of values for the couple (a, b) in $\mathbf{W}_{aper}$ and a unique *a priori* belief vector $(1, \frac{1}{2}, -1, -\frac{7}{10})'$. In figure C.4a, $w_{11}^0 = \frac{3}{4}$ and given $\mathbf{b}_0$, the network is aperiodic "enough", thus beliefs converge. While, in figure C.4c, $w_{11}^0 = \frac{1}{100}$ and beliefs do not converge.

Note that the *a priori* belief vector always plays a role in the components of the listening matrices over time. More specifically, the form specified in equation 3.4, imposes that self-listening weights should always converge, either to 0 or 1. Meanwhile, clearly, first, if all self-listening weights converge to 1 (and $w_{ii}^0 \neq 1, \forall\, i$), although convergence is guaranteed, consensus is impossible. Such a behavior creates a fragmented network where everyone's belief likely converges to a belief different from everyone else. Second, if some self-listening weights converge to 0, a basic assumption required to draw *sufficient* conditions for convergence in such a model, provided by previous authors (see Blondel et al. (2005)) is violated. Lastly, also because of the "perpetual" interdependency between the sequences $(\mathbf{W}_t)_t$ and $(\mathbf{b}_t)_t$, conditions like the absolute infinite flow property (see Touri and Nedic (2012), Rezaienia et al. (2017)), cannot be used to determine a general characterization of convergence (or almost sure convergence) in my model. As a consequence of proposition 3.1 below and from the results on consensus time depicted in Section 3.7, depending on *a priori* beliefs, it is possible for beliefs not to converge, even starting from a group structure; which is usually the structure with higher connectivity properties. Consequently, the main (necessary) condition for convergence of beliefs, that I mention in this paper, is that $\mathbf{W}_0$ should be aperiodic.

3.5.2 Convergence without Consensus in a strongly connected network

Consider the following key proposition.

Proposition 3.1. *In general, for a strongly connected a priori network $\boldsymbol{W}_0$, even if $\boldsymbol{W}_0$ is aperiodic and beliefs converge to a belief vector $\bar{\boldsymbol{b}}$, $\bar{\boldsymbol{b}}$ is not necessarily a consensus vector.* [22]

Proof. As shown using the two-individual identity matrix example above, proposition 3.1 is trivial for unconnected networks. Intuitively, in an unconnected network, there exists at least two individuals who never listen to themselves; neither directly, nor remotely. In my model, if an individual does not *a priori* listen to someone else at all, then, they will never listen to them forever. So, none of them is influenced by the other. Then obviously, unless by pure luck, beliefs in such a network will not converge to a consensus.

To prove proposition 3.1, we only need to find a strongly connected network and an *a priori* belief vector such that beliefs converge numerically, but there is no consensus. In fact, if it is true for a strongly connected network, then the result holds for weak connectedness too. [23]

Consider the following strongly connected three individuals *a priori* network $\mathcal{N}_{0,s}$ represented by the digraph $G_{0,s}$ in figure 3.4a, and associated with the listening matrix:

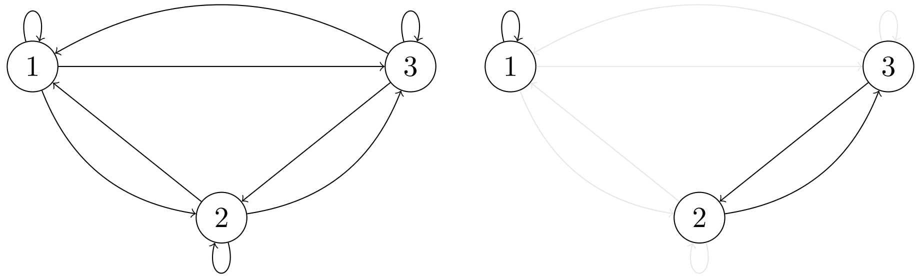
$$\mathbf{W}_{0,s} = \begin{bmatrix} \frac{17}{20} & \frac{1}{10} & \frac{1}{20} \\ \frac{1}{5} & \frac{7}{10} & \frac{1}{10} \\ \frac{1}{10} & \frac{1}{10} & \frac{4}{5} \end{bmatrix}$$

The values on (or near) each arrow of the digraph are the listening weights. For instance, for $\mathbf{b}_0 = (1, -.5, -1)'$ beliefs converge numerically to $\bar{\mathbf{b}}_s = (.65, -.51, -.51)'$, which is not a consensus vector. [24] On the contrary, $\bar{\mathbf{b}}_s$ reflects polarization instead. Here, a polarization of beliefs over time (or polarization of a network), refers to a situation in

22. The strong connectedness is a stronger assumption than any other level of connectedness (weakly connected, or connected). Therefore, the proposition is also true in those cases. The framework designed in a recent study by Algan et al. (2019) gives a real-life example of networks (not necessarily strongly connected or aperiodic) in which opinions do not necessarily converge to a consensus. In their study, they find for instance that even if a pair of students become friends, if there were a very large difference in their pre-Science Po political opinion (*a priori* belief) their opinions are likely to diverge anyhow.

23. Given all the extensive results for consensus using the DeGroot model (DeGroot (1974), DeMarzo et al. (2003), Golub and Jackson (2010)), it is unnecessary to show that, for a strongly connected network such that beliefs converge, they will usually converge to a consensus. This is what should have been expected at first.

24. Recall that for all simulations, I use $f(x) = -x + 2 + \xi$ and $g(x) = 2x + \xi$. The parameter ξ is fixed at 10^{-2}.



(a) Network digraph $G_{0,s}$ of $\mathbf{W}_{0,s}$

(b) Network digraph $G_{t,s}$ of $\mathbf{W}_{t,s}$ as $t \rightarrow +\infty$

Figure 3.4 – Evolution of Network digraph of $\mathbf{W}_{0,s}$ with $\mathbf{b}_0 = (1, -.5, -1)'$

which beliefs converge, but not to a same value.[25] I give more details on polarization later in the paper. Figure 3.4b above, shows what the limiting network $\mathcal{N}_{t,s}$ would look like, for $\mathbf{b}_0 = (1, -.5, -1)'$, as $t \rightarrow \infty$ (asymptotically).

Note that, for different values of initial beliefs, the same system could have converged to a consensus. For instance, for $\mathbf{b}_0 = (1, -.3, -.5)'$, beliefs converge to the consensus vector $(.39, .39, .39)'$.[26] □

Cognitive biases as a key impediment to consensus

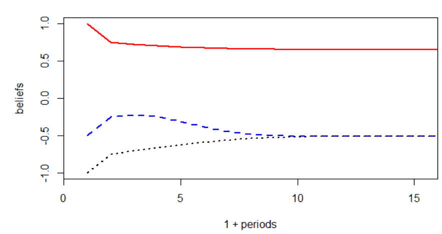
Up to now, I have established that, even if everyone in a network is influenced by everyone else (strong connectedness), it is still possible that they will never converge to a consensus. This result makes a clear contrast with the DeGroot model. In fact, for the DeGroot model, Golub and Jackson (2010) proved in their paper that for a strongly connected and aperiodic network like $\mathcal{N}_{0,s}$ above, consensus is reached. In my model, the source preventing consensus in a network starting with an *a priori* strongly connected structure, comes from the two main cognitive biases mentioned above. In fact, each bias triggers a particular behavioral mechanism. Consider the example used in the preceding proof (see figure 3.5). When peers' beliefs become further apart (see red and blue lines in figure 3.5a), it corresponds to a behavior such that individuals tend to reduce their attention to those peers (figures 3.5b and 3.5c), immediate result

25. Specifically, when beliefs are polarized, it is as if there are several sub-network's "consensus", reached in smaller components of a network. Strictly speaking, there cannot be two consensus in a same network, thus the need to refer to the terms polarization or fragmentation. In fact, in a network $\mathcal{N}_t$, it is possible that everyone's belief in several components $\mathcal{N}_{t,1}, ..., \mathcal{N}_{t,l} \subset \mathcal{N}_t$ converges to a similar belief; yet the "consensus" belief in each of these components is different from that in the other components. In the example used here in proposition 3.1, the two sub-"consensus" values are .65 in $\mathcal{N}_{t,1} = \{\{1\}, \{(1,1)\}\}$ and $(-.51, -.51)'$ in $\mathcal{N}_{t,2} = \left\{\{2,3\}, \left\{(i,j) \in \{2,3\} / w_{ij}^t > 0\right\}\right\}$.

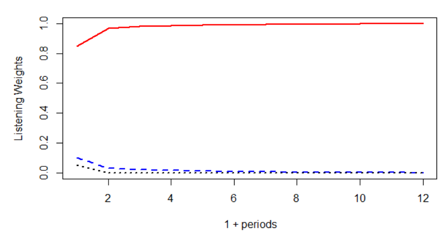
26. Consult figure C.3 in appendix for numerical simulations.

of the confirmation bias. Besides, because they lower their attention to those peers, it has an impact on the speed at which the beliefs could get closer. Moreover, the *ERS bias* also tend to slower the speed of convergence toward a hypothetical consensus, the more people hold more extreme views. This, subsequently, results in a snowball effect, where listening weights to others are more and more reduced, and beliefs tend to stay forever apart.

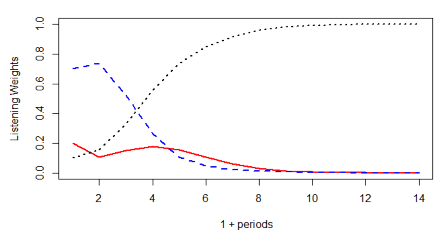
Figures 3.5b to 3.5d depict the underlying mechanism through which 1 tends to (asymptotically) "disconnect" from the rest of the network as shown in the digraph in figure 3.4b. In figure 3.5b, individual 1 self-confidence (w^t_{11}), represented by the plain red line, increases when they see, at period 0, that peers 2 and 3 beliefs and theirs are different. Furthermore, it converges to 1, and reciprocally individual 1 decreases their attention towards individuals 2 (blue dashed line) and 3 (black dotted line). In figure 3.5c, at first, individual 2 decreases their attention to 1 (plain line), but seems to increase it again from period 1 to period 3. However, because the gap between 1 and 2's beliefs seems to increase slightly (figure 3.5a), 2's attention to 1 finally decreases again and converge to 0; whereas their attention to 3 keeps increasing. As time goes to infinity, apart from individual 1 themselves, no one in the network listens to 1. Similarly individual 1 tend to listen less and less to anyone else.



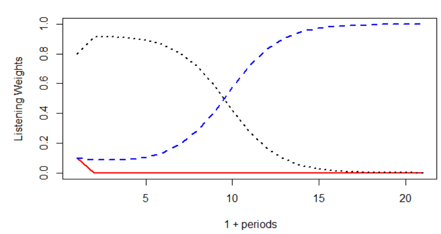
(a) $\mathbf{b}'_0 = (1, -\frac{1}{2}, -1)$, $\bar{\mathbf{b}} = (.65, -.51, -.51)$



(b) Indiv. 1: $\mathbf{w}^1_{0,s} = (\frac{17}{20}, \frac{1}{10}, \frac{1}{20})$, $\mathbf{w}^{\bar{1}}_{0,s} = (1, 0, 0)$



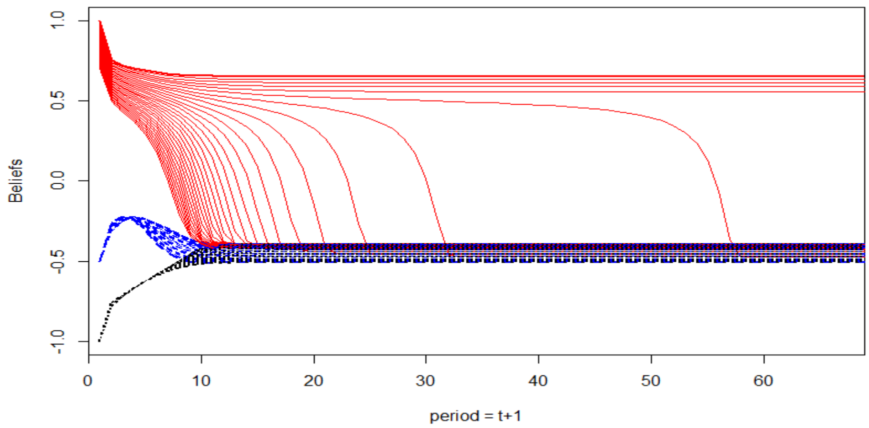
(c) Indiv. 2: $\mathbf{w}^2_{0,s} = (\frac{1}{5}, \frac{7}{10}, \frac{1}{10})$, $\mathbf{w}^{\bar{2}}_{0,s} = (0, 0, 1)$



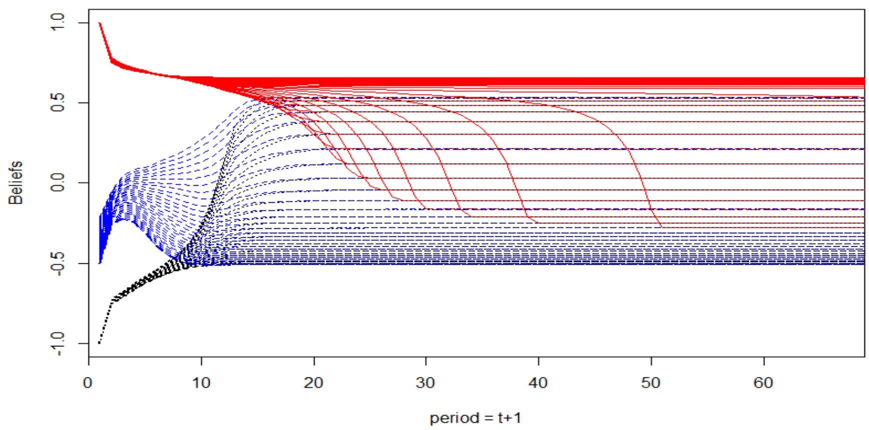
(d) Indiv. 3: $\mathbf{w}^3_{0,s} = (\frac{1}{10}, \frac{1}{10}, \frac{4}{5})$, $\mathbf{w}^{\bar{3}}_{0,s} = (0, 1, 0)$

Figure 3.5 – No Consensus with *a priori* strongly connected network $\mathbf{W}_0 = \mathbf{W}_{0,s}$

Figure 3.6a below shows a sequence of belief paths for the three individuals to their steady state, as individual 1's *a priori* belief takes different ϵ-extreme and extreme



(a) $b1^0 \in [\frac{7}{10}, 1]$, $b_2^0 = -\frac{1}{2}$, $b_3^0 = -1$: Pronounced *ERS bias*



(b) $b1^0 = 1$, $b_2^0 \in [-\frac{1}{2}, -\frac{1}{5}]$, $b_3^0 = -1$: Pronounced Confirmation Bias

Figure 3.6 – Effect of the ERS and confirmation biases on convergence rates and consensus

values in $[\frac{7}{10}, 1]$. The *a priori* network is still $\mathcal{N}_{0,s}$, and associated listening weight paths to their convergence values are represented in figures C.6a, C.6b and C.6c. Although both ERS and confirmation biases contribute to each path, this figure allows us to understand more on the effect of *ERS bias*. In contrast, in figures C.6d, C.6e and C.6f instead, the listening weights paths represented are related to a sequence of b_2^0 taking values in $[-\frac{1}{2}, \frac{1}{5}]$, as shown in figure 3.6b.

Over time, individuals whose beliefs *a priori* are at both extremes of the beliefs spectrum, tend to listen (at most) to one of their peers (see figure C.6a, C.6c, C.6d and C.6f). This is a consequence of their confirmation bias which makes them side, if possible, with their closest peer (in belief). The only time they consistently listen to more than one peer is when those peers have the same beliefs over time. In that case, among those peers, the one they listen more to (*a priori*), remains the one they listen more to among those peers over time.

In general, as shown in figures 3.6b and 3.6a, not only do both bias influence the consensus time, but also they influence its value. Consider an individual whose belief *a priori* is farther from their peers' (see individual 1). The closer these peers' beliefs are among themselves, and the more they listen to themselves (black versus red lines in figure C.6b and blue versus red lines in figure C.6c), the higher the gap between the consensus belief and individual 1's *a priori* belief. However, in comparison, as shown in figures C.6e and C.6f, we see that individual 2 is the one who determines which one of the two extreme believers will make more concessions for consensus. Individual 2 is a mediator. Specifically, at equal beliefs distance, a peer of individual 2 whom they listen more to (here individual 1 *a priori*) is likely to make fewer concessions for consensus.

From figure 3.6b and C.6e, individual 2, who is less subject to *ERS bias*, is the one who determines the existence of a consensus.

The importance of a "good" mediator for consensus

As shown in figure C.6b and C.6e in the appendix, the self-attention 2 assigns to themselves (blue dashed lines) always decreases for all *a priori* ϵ-extreme beliefs values for individual 1 ($\epsilon = .3$). This is because, individual 2 already has a consensual attitude towards their peer 3. However, a whole network's consensus does not exist for all these values of b_1^0. In fact, the mechanism through which their belief becomes closer to individual 1's, is through the level and the rate at which their attention to 1 reaches its steady state (i.e., becomes almost constant). The plain red lines in figure C.6b show that, the highest one seems to become constant and non-zero around $t = 8$, the third highest, around $t = 11$, and so on. For the lowest lines, although they seem to reach zero quickly, it is not exactly the case. In reality they slowly tend to zero or very small values close to zero, which we cannot see due to the scale of the figure. We can make the same analysis with figure 3.6b. In other words, the slower individual 2 decreases their attention to 1, the higher the chances for consensus, and the faster the speed toward consensus (figure 3.6a). In addition, when individual 2 consistently listens to both individuals 1 and 3, consensus is guaranteed. Moreover, the higher 2's attention to 1 gets, the closer the consensus value is to 1's *a priori* belief. In this network, individual 2 is a mediator between the two opposite extreme parties (1 and 3). If they "act wisely" and do not take sides for one of the two (convergence value of w_{21} and w_{23} differ from zero), consensus is guaranteed. This is what I call a "good" mediator in this paper; someone who consistently listens to both immediate sides around them. Individuals 1 and 3 are not mediators because, at each time, their beliefs are always at one end, or the other of the beliefs distribution in the network.

In the remaining of the paper, I mostly focus on conditions for consensus, when convergence is guaranteed. I also discuss the matter of polarization in a network (as shown in the preceding example).

3.5.3 A Characterization of Consensus for Strongly Connected *a Priori* Networks

Proposition 3.2. *Let $\mathcal{N}_0$ a strongly connected and aperiodic a priori network, and let $\boldsymbol{b}_0$ a vector of beliefs a priori such that beliefs converge to a vector $\bar{\boldsymbol{b}}$.*
Then, (i) implies (ii):

(i) $\bar{\boldsymbol{b}}$ is a consensus vector $b.\boldsymbol{i}$ such that $|b| \neq 1$;

(ii) Every individual's self-confidence converges to zero in the network:

$$\lim_{t \to +\infty} w_{ii}^t = 0, \ \forall \ i$$

Proof. See appendix C.1.2. □

The numerical simulations in the appendix (see figures C.3, C.5) show that when all beliefs converge to a consensus value ($\notin \{1, -1\}$), all self-listening /self-confidence weights converge to zero. On the opposite, we can see in figures 3.5b,3.5c, and 3.5d that, when beliefs converge, but at least one individual's belief converges to a unique value different from their peers (polarization), then at least one individual's self-confidence converges to 1. In addition, when beliefs converge to an extreme consensus (1 or -1), some self-confidence weights also converge to 1.

Proposition 3.2 means that if there exists at least someone in the network who does not adopt a conciliatory attitude by consistently reducing the weight they assign to themselves as of a certain moment, a non-extreme consensus never exists. Either the network reaches a consensus and it is an extreme belief, or it becomes polarized.

3.5.4 Absolute Consensus

In this paper, absolute consensus in a network means that a consensus exists, whatever the *a priori* belief vector $\mathbf{b}_0$. Here are a few trivial examples of networks with absolute consensus property.

Constant *a priori* network (identical lines in W_0)

Claim 3.1. *Let* $\boldsymbol{u}_{sto}$ *a* $1 \times n$ *stochastic vector, and a network's a priori trust matrix given by* $\boldsymbol{W}_0 = \boldsymbol{i}\boldsymbol{u}'_{sto}$*. Then, for all a priori belief vector* $\boldsymbol{b}_0$*, the network converges to a consensus at period 1.*[27]

Proof. Given any vector $\mathbf{b}_0$, let $(\mathbf{M}_1, ..., \mathbf{M}_t, ...)$ a random sequence of stochastic matrices. We have:
$\mathbf{W}_0\mathbf{b}_0 = \boldsymbol{i}\mathbf{u}'_{sto}\mathbf{b}_0 = \bar{b}_0\boldsymbol{i}$, where $\bar{b}_0 = \sum_{j=1}^{n} w_{ij}^0 b_j^0$. Then:
$\mathbf{M}_1\mathbf{W}_0\mathbf{b}_0 = \bar{b}_0\mathbf{M}_1\boldsymbol{i} = \bar{b}_0\boldsymbol{i}$.
Now, suppose $\prod_{m=0}^{t-1}\mathbf{M}_{t-m}\mathbf{W}_0\mathbf{b}_0 = \bar{b}_0\boldsymbol{i}$. It follows that $\prod_{m=0}^{t}\mathbf{M}_{t+1-m}\mathbf{W}_0\mathbf{b}_0 = \bar{b}_0\boldsymbol{i}$.
Therefore for any vector $\mathbf{b}_0$ of *a priori* beliefs, $\lim_{t\to+\infty}\prod_{m=0}^{t}\mathbf{W}_{t-m}\mathbf{b}_0 = \bar{b}_0\boldsymbol{i}$. □

In a constant network, both cognitive biases have no direct effect on the consensus time.[28]

A constant network represents a network in which everyone assigns the same weight to the same individuals. Assume for instance that the individual characteristics on which individuals base their *a priori* listening weights are public knowledge and everyone has the same listening preferences over these characteristics. Therefore, everyone gives the same weight to the same individuals.

Although it is less common in real-world networks, that everyone assigns to their peers (including to themselves), the same listening weight as everyone else does, it may happen for a subset of peers in some networks. Albeit in such networks, consensus may not be reached as fast as in a constant network; still, the more there exist such peers in the network who receive the same attention from everyone, the lower the time to consensus (*ceteris paribus*). Consider the following real-life example where a subset of members of a network may receive the same weight from everyone else.

Consider for instance prison gangs. Skarbek (2014) gives a detailed description of their internal organization. Before an inmate joins a gang, they usually have to do a "background check" on him. In case they find out that he was in protective custody, all gang members tag him as a low-quality recruits, that is, a potential traitor. If recruited despite that bad reputation, then "[...]he becomes like garbage" to everyone in the

27. Note here that the consensus time is quite precise here. This is due to $b_i^t = b_j^t,\ \forall\ t \geq 1, \forall\ i,\ j$.
28. Figure C.5 in the appendix gives an example of network composed of three individuals.

gang, and everyone else assigns a very low listening attention (almost uniformly) to him (Skarbek, 2014).

Similarly, the leader of a gang might receive the same listening attention from his subordinates. However, as the top member of the gang, he might assign a higher weight (or the total weight) to himself than others do.

A Stochastic "Rooted Tree"

A stochastic rooted tree denotes any weakly connected directed network which possesses a unique cycle, which is a self-loop. Such a network has a structure similar to that of a rooted tree, oriented toward the root, with a unique self-loop related to the root (top) of the tree. [29] Let's note them D for "dictator" (or L for leader) as represented in figures 3.7 and 3.8 below. In that type of network, two individuals i and j, $(i, j \neq D)$ cannot influence each other. The relationship between two peers is always strictly hierarchical, and no strongly connected component of the digraph exists. In fact, any network with this property guaranties consensus, because the existence of a cognitive bias does not influence the final outcome. A cognitive bias only influences the listening weights at each time, but not the consensus value, neither the consensus time. Intuitively, notwithstanding the updated values of the listening matrix, the network can never "get stuck" below D. In fact, during the process, some inward oriented links may break; but, for everyone in the network, at least one link always remains in the end.

A first real-life example of stochastic rooted-tree network is for instance the previously mentioned prison gang. In fact, still according to Skarbek (2014), gangs are very reputation-based networks, with a strong hierarchical organization. Therefore, it is often observed, that based on the reputation they have established, some inmates might receive a higher consideration from their peers than others. Such gangs usually have a clear hierarchy, and members often know who they shouldn't dare ignore. In consequence, among the peers they are connected to, gang members often know pretty well who they (should) the most listen to; leading to a clear hierarchy within all members. Depending on the gang, the leader may be picked, or imposed (if too influential).

Such a hierarchical gang can be represented by a graph like in figure 3.7. And as in that figure, everyone needs not to be directly influenced by the leader. In fact, they may

29. Actually, a rooted tree is an oriented tree, and has no loop *per se*. Therefore, the matrix representing it should contain a line filled with zeros, which is not stochastic. However, the graph representation of a directed network with only one cycle, associated with a stochastic matrix, is very similar to a rooted tree, oriented towards the root. The only difference is that there is a loop at the root.

not even know him. There are immediate subordinates of the leader who also influence a subset of members, each. And the subordinates, also have their own subordinates, and so on. On the organization level, gangs use pressure instruments to make their members abide to their rules, and consequently maintain a tight grip over its intrinsic structure. That is their "constitution", which usually contains very clear sanctions in case someone exits the gang, which is considered a huge betrayal (Skarbek, 2014). Those rules guaranty, in a certain way, that individuals put almost zero weights on themselves. And that is the main reason why they often reach consensus over their extreme actions in such gangs. In fact, although at first, some of its members were only petty criminals with very moderate views on the participation in a crime or the use of violence, over time they may adopt more extreme views, and consequently participate in more dangerous crimes. And that is what my model also predicts.

Consider for instance individual 15 in figure 3.7. Assume he is *a priori* a petty criminal whose belief on the use of violence in a crime is *a priori* 0. At first, because of his subjectivity, he connects with 4 (Mr. i), 8 (Mr. j), and 11 (Mr. k) without knowing their true beliefs. Let's say he is impressed by i's appearance and seemingly dignity, by j's apparent authority, and by k's apparent experience of the prison systems. These are a few reasons why he decides to connect with them and to enter the gang. He may think that with those attributes they might make his life as an inmate easier and more bearable. My model predicts that, although he is not aware of it initially, if the leader has an extreme view on violence, he will someday acquire the same belief and will have to act accordingly. Meanwhile, the only freedom he has is that over time he may listen more to someone between i, j, or k. Due to his biased behavior (cognitive biases), he may even break ties with one or two of them if he finds out that they have very different views. However, in the end, he agrees with what the leader believes, although he never met with him. The only way for him to escape that fact is to add another self-loop for 15. In that case, the network's structure changes and is no longer a stochastic rooted tree. The new structure implies that he listens to himself and does not necessarily agree to abide by his peers beliefs over time. He may exit the gang and become an isolated inmate in the prison. However, as mentioned previously, that is too dangerous a behavior, which might lead to an increase of abuse against him or even his death.

Another plausible example of stochastic rooted tree is an army network. An army usually has a clear hierarchical organization with several components, so that a hierarchy also remains in each component. In my view, an army is one of the examples for which a stochastic rooted-tree network may have a positive outcome to society instead of the gangs depicted earlier. In fact, in the case of a battalion for instance, in which officers listen to their superiors, the hierarchy here is a canal leading to absolute consensus in the unit, whatever the initial beliefs of the officers. More specifically, the network structure of an army (or of any of its components) is more like an anti-arborescence, where there is exactly one path from any individual to the root, as shown in figure 3.8. Such a network has pretty trivial behavior because its associated listening matrix is a permutation matrix. Therefore, cognitive biases are ignored and my model reflects the same results as for the naive model.

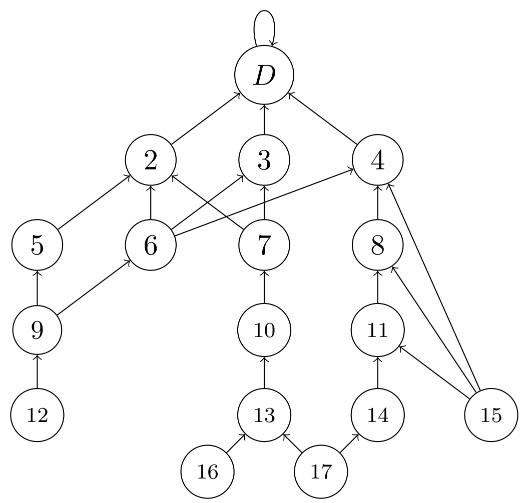


Figure 3.7 – Stochastic rooted tree

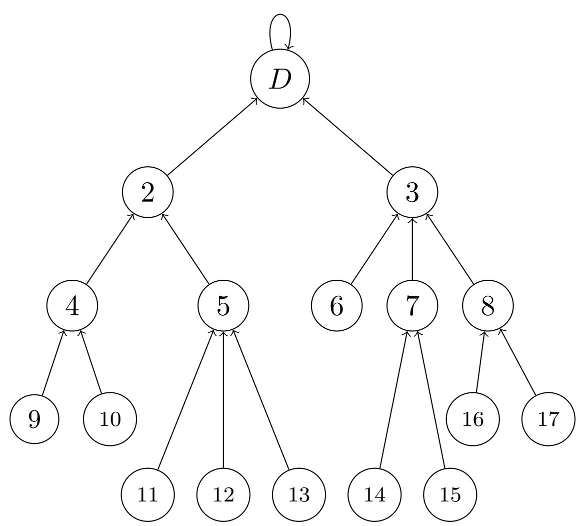


Figure 3.8 – Stochastic rooted tree: "anti-arborescence"-like structure

The interaction matrix associated with a stochastic rooted-tree network, may be represented by a stochastic lower triangular (or upper triangular) matrix $\mathbf{W}_{rt}$ with all its diagonal elements equal to 0 apart from only one equal to 1 as below:

or

$$\mathbf{W}_{rt} = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ 1 & 0 & \dots & \dots & 0 \\ w_{31} & w_{32} & 0 & \dots & \vdots \\ \vdots & \dots & \dots & \ddots & \vdots \\ w_{n1} & \dots & \dots & w_{n\ n-1} & 0 \end{bmatrix} \qquad \mathbf{W}_{rt} = \begin{bmatrix} 0 & w_{12} & \dots & \dots & w_{1n} \\ 0 & \ddots & \dots & \dots & \vdots \\ \vdots & \dots & 0 & w_{n-3\ n-1} & w_{n-3\ n} \\ \vdots & \dots & \dots & 0 & 1 \\ 0 & \dots & \dots & \dots & 1 \end{bmatrix}$$

If an *a priori* listening matrix is lower (respectively upper) triangular with the properties of the matrix $\mathbf{W}_{rt}$, then all subsequent matrices obtained have the same properties for all *a priori* beliefs vector (see proposition C.1 in the appendix). This, combined with next lemma 3.1 prove that starting from all sizes of *a priori* networks with a stochastic rooted-tree structure, consensus is bound to happen. Specifically, the consensus time is at most $n-1$. Intuitively, it should equal the diameter of its associated digraph minus one. Moreover, the consensus belief obtained is always the belief of the dictator D (or leader L).

Lemma 3.1. *Product of $\boldsymbol{W}_{rt}$-type matrices.*
Let $(\boldsymbol{A}_t)_t = \left(a_{ij}^t\right)_t$ an infinite sequence of $n \times n$ stochastic lower triangular matrices (upper triangular), with $a_{11}^0 = 1,\ a_{ii}^0 = 0\ \forall\ i \neq 1$ (respectively$a_{nn}^0 = 1,\ a_{ii}^0 = 0\ \forall i \neq n$). Let $(\boldsymbol{P}_{lk})_{k\geq 1}$ (respectively $(\boldsymbol{P}_{uk})_{k\geq 1}$) a random sequence of lower (upper) triangular matrices extracted from the sequence $(\boldsymbol{A}_t)_t$, $\Pi_{l,r} = \prod_{m=1}^{r} \boldsymbol{P}_{l\ r+1-m},\ r \geq 1,\ \Pi_{u,r} = \prod_{m=1}^{r} \boldsymbol{P}_{u\ r+1-m},\ r \geq 1,\ \boldsymbol{i}_l = (1\ 0 \dots 0)'$ and $\boldsymbol{i}_u = (0 \dots 0\ 1)'$ two $n \times 1$ unit vectors. Then:

$$\Pi_{l,n-1} = \boldsymbol{i}\boldsymbol{i}_l'$$
$$\Pi_{u,n-1} = \boldsymbol{i}\boldsymbol{i}_u'$$

Proof. See appendix C.1.2 □

Another hierarchical network

Consider the connected network in figure 3.9, associated with the listening matrix $\mathbf{W}_h$ below, such that $w_{ii}^0 \leq \frac{1}{n-i+1}$, and $w_{ij}^0 = \frac{1-w_{ii}^0}{n-i}$, $\forall\, j > i,\ \forall\, i \in \{1, 2, \ldots, n-1\}$.

$$\mathbf{W}_h = \begin{bmatrix} w_{11} & w_{12} & w_{13} & \ldots & \ldots & w_{1n} \\ 0 & w_{22} & w_{23} & \ldots & \ldots & w_{2n} \\ 0 & 0 & w_{33} & w_{34} & \ldots & w_{3n} \\ \vdots & \ldots & \ldots & \ldots & \vdots & \\ 0 & 0 & \ldots & \ldots & w_{n-1\ n-1} & w_{n-1\ n} \\ 0 & 0 & \ldots & \ldots & \ldots & 1 \end{bmatrix}$$

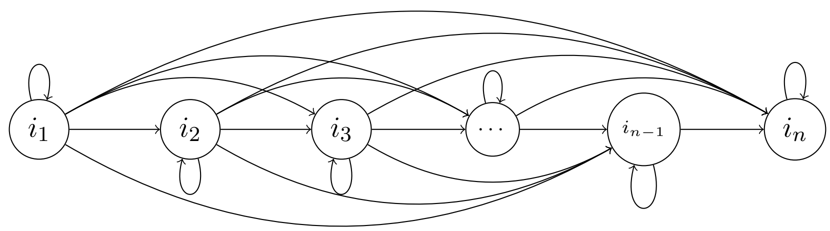


Figure 3.9 – Hierarchy Network with $n-1$ leaders and $n-1$ followers

This network is like a line (or a chain) of individuals all facing the same side. Everyone in the network, except the first individual i_1, leads (i.e., is listened to) everyone behind them; and everyone in the chain, except the last individual i_n, listens to everyone before them. In such a network, consensus always exists. Similarly to the stochastic rooted-tree example, the consensus value, in this network is always the *a priori* belief of the most influential leader, who is individual i_n in this case. i_n is clearly the most influential individual in the network; however, over time everyone does not necessarily put more weight on i_n. These results are also valid for a similar network, represented by a triangular matrix, such that everyone listens to the person before them, and maybe to some others immediately after them, and so on, without necessarily listening to the ones who are too far ahead in the line. The main idea is that the weight they put on themselves is lower or equal the weight on their peers (who are all equal). Evidently, if some individuals were to put too much weight on themselves *a priori*, it could jeopardize the existence of a consensus.

Consider for instance the following scenario. A long line of people walking at night in a deep forest. Assume, everyone put the same weight on people in front of them, and that, that weight is greater or equal to the one they put on themselves. Assume they are travelers, trying to escape their original land. They may be refugees escaping a war or an attack, or people trying to immigrate illegally, but through a dangerous zone. They know the statistics on the success rate when people take such a route to escape. That is part of the signal they received. Assume then that they hear from a distance an explosion, or a sound similar to gunfire. Yet the leader of the group, who is far ahead in the line, gives a reassuring statement like -"Do not worry, they are very far from us"- while still moving forward. What will the others (right behind the leader) do? My model predicts that, in absence of any additional information, the ones closer to the leader will keep going forward, likewise those behind them will do the same, and the whole group will keep moving forward. And this happens, although they all heard the same sound. On the contrary, if the leader says -"They are coming after us"- or -"This way is not safe"-, most certainly they will all instantly run back without thinking twice. More, the last person in line will probably not wait to know what the leader has said. The simple fact that some people before them are running back constitutes a more than enough way of communicating their belief.

Several other real-life examples which could also fit with the results on this particular network. It could be a waiting line outside a closure sale store, or a line of cars stuck in traffic, etc. All the results on absolute consensus to the leader's belief, mentioned in Sections 3.5.4 and 3.5.4 hold, no matter how extreme or moderate their belief. In next Section 3.5.5, I focus on convergence and consensus to extreme and ϵ-extreme beliefs.

In general, to derive conditions for consensus in an average-based beliefs updating process, authors often analyze the ergodic property of the sequence of backward product of matrices. Aside from ergodicity, there may use other ways like the paracontracting properties of each matrix composing the sequence $(\mathbf{W}_t)_t$, as introduced by Nelson and Neumann (1987). Let's say one can verify that the *a priori* listening matrix is paracontracting and respects additional properties such that the resulting whole sequence is also paracontracting. Then, under that condition, the results of Nelson and Neumann (1987) imply for our model that beliefs converge to a consensus.[30] However, due to the way cognitive biases affect the whole matrices updating in equation 3.4, it prevents finding an analytical form of the sequential w_{ij}^ts, using most forms of the *a priori* matrices $\mathbf{W}_0$. Therefore, I limit my analysis to the estimation of a probability of consensus as in previous Section 3.7. Nonetheless, next Section gives some additional theoretical results for extreme consensus.

30. See section C.1.4 in the appendix.

3.5.5 Convergence to Extreme Beliefs

Consider a network $\mathcal{N}$, in which some individuals hold *a priori* extreme or ϵ-extreme beliefs. The graph in figure 3.10 is an example of weakly connected network in which individuals 1 and 2 have left extreme or $\epsilon-$extreme beliefs (lighter colors) while 4 and 5 have both right extreme or ϵ-extreme beliefs (darker colors). In figure 3.10, the subsets of individuals $\{1,2\}$ and $\{4,5\}$ represent what is defined as ϵ-extreme poles next.

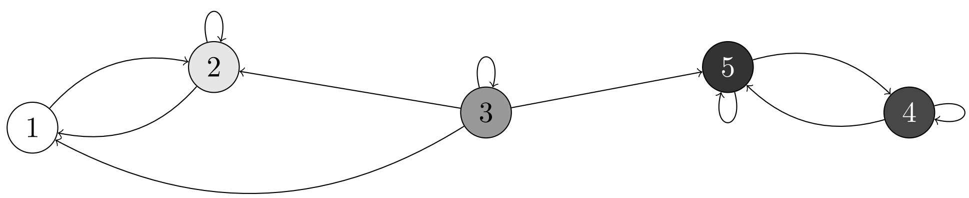


Figure 3.10 – A weakly connected network with two ϵ-extreme poles

Definition 3.10. *ϵ-extreme pole*
Let $\epsilon \in [0, \frac{1}{2})$ a scalar. Consider the following subsets of ϵ-extreme believers:
$\mathcal{V}^-_{\epsilon,t} = \left\{ i \in V^-_{\epsilon,t} / \ \sum_{k \in \mathcal{V}^-_{\epsilon,t}} w^t_{ik} = 1 \right\}$, *and* $\mathcal{V}^+_{\epsilon,t} = \left\{ i \in V^+_{\epsilon,t} / \ \sum_{k \in \mathcal{V}^+_{\epsilon,t}} w^t_{ik} = 1 \right\}$
$\mathcal{V}^-_{\epsilon,t}$ and $\mathcal{V}^+_{\epsilon,t}$ represent respectively, the left and right ϵ-extreme poles of the network $\mathcal{N}$ at time t. At time $t = 0$, if $|b^0_i| = 1 \ \forall i \in \mathcal{V}_{\epsilon,0} = \mathcal{V}^-_{\epsilon,0} \cup \mathcal{V}^+_{\epsilon,0}$, then $\mathcal{V}^-_{\epsilon,0}$ is the left extreme pole of $\mathcal{N}$ and $\mathcal{V}^+_{\epsilon,0}$ its right one. I denote $n^-_{\epsilon,t}$ (respectively $n^+_{\epsilon,t}$) the cardinal of $\mathcal{V}^-_{\epsilon,t}$ (respectively $\mathcal{V}^+_{\epsilon,t}$), and we have $n^-_{\epsilon,t} \leq n$ (resp. $n^+_{\epsilon,t} \leq n$) $\forall\ \epsilon,\ \forall\ t$.

From this definition, someone in a pole can be influenced only by the members of the same pole, and by no one outside of it. Then, there is no edge from any member of an ϵ-extreme pole pointing at non-members of their pole in the network's digraph.

Proposition 3.3. *Consider a network $\mathcal{N}_0$, connected at time $t = 0$. Then $\mathcal{N}_0$ does not have both a non-empty right ϵ-extreme pole and a non-empty left ϵ-extreme pole.*

Proof. See Appendix C.1.2 □

The proof is straightforward. In fact, from the definition of each pole, there is no path from a member of a left ϵ-extreme pole to a right ϵ-extreme pole (and vice versa); which contradicts the fact that the network is connected. In the previous figure 3.10, if any edge is added from the vertices in $\{1,2\}$ to $\{3,4,5\}$ or from $\{4,5\}$ to $\{1,2,3\}$, the network loses one of its ϵ-extreme poles and becomes connected.

Proposition3.3 is equivalent to the converse that is, if there is both a left and a right ϵ-extreme pole in $\mathcal{N}_0$, then $\mathcal{N}_0$ is not connected.

Proposition 3.4. *A necessary condition for consensus to an ϵ-extreme belief Consider a network in which beliefs converge to a consensus b. Let $\epsilon \in [0, \frac{1}{2})$. Then:*

$$|b| \geq 1 - \epsilon \iff \mathcal{V}_{\epsilon,0} = \mathcal{V}^-_{\epsilon,0} \neq \emptyset, \text{ or } \mathcal{V}_{\epsilon,0} = \mathcal{V}^+_{\epsilon,0} \neq \emptyset$$

Proof. See Appendix C.1.2 □

The simple fact that there exists some individuals who hold extreme views in a network is not sufficient for convergence to extreme beliefs. From proposition 3.4, a consensus is ϵ-extreme if and only if a non-empty right or left ϵ-extreme pole exists in the *a priori* network. Particularly, for $\epsilon = 0$ and p$|b| = 1$roposition 3.4 implies that, to obtain a consensus at a right (left) extreme, the network must necessarily have a right (left) extreme pole.

Example 3.4. Consider the 3-agent network $\mathcal{N}_{0,a}$ represented in figure 3.11, containing an extreme pole $\{3\}$ ($|b^0_3| = 1$). Its associated *a priori* listening matrix is defined as:

$$\mathbf{W}_{0,a} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}$$

Beliefs always converge to a consensus in $\mathcal{N}_{0,a}$. As shown in all the figures A.3a,A.3b, A.3c, and A.3d, the consensus value is always the *a priori* belief of individual 3.

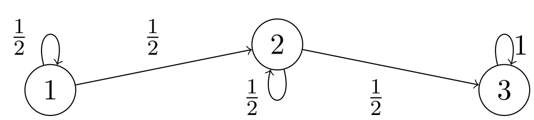
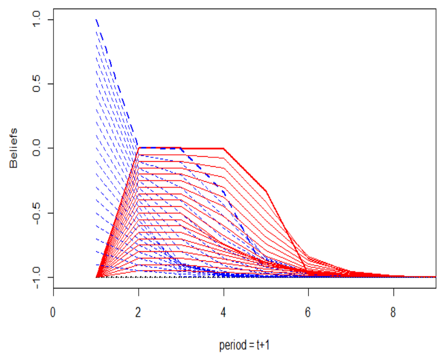
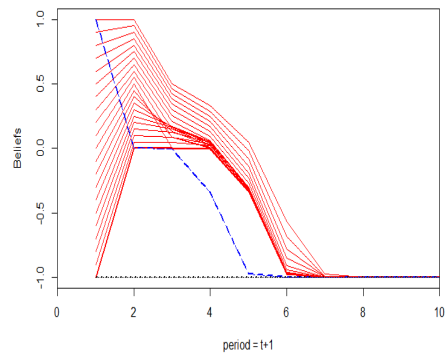


Figure 3.11 – Digraph for network $\mathcal{N}_{0,a}$ listening matrix $\mathbf{W}_{0,a}$

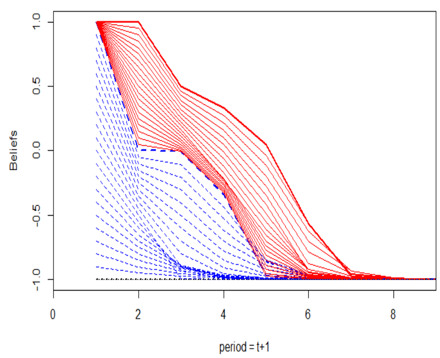
Figure A.3d represents a case where $\{2,3\}$ is an extreme pole. In the case that b^0_3 is no longer extreme or ϵ-extreme, as shown in figures 3.12e and 3.12f, beliefs still converge to a consensus. However, it is no longer extreme because b^0_3 is not. In these cases, clearly, some individuals have *a priori* extreme beliefs in the network. However,



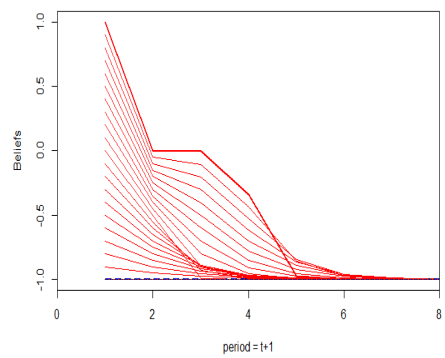
(a) $b_1^0 = b_3^0 = -1,\ b_2^0 \in [-1, 1]$



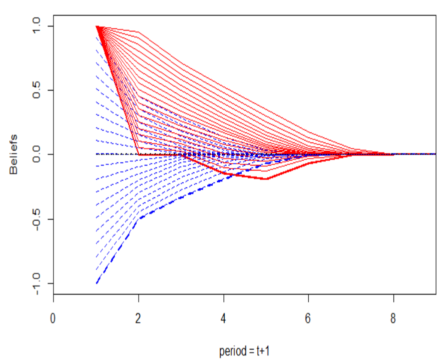
(b) $b_1^0 \in [-1, 1], b_2 = 1, b_3^0 = -1$



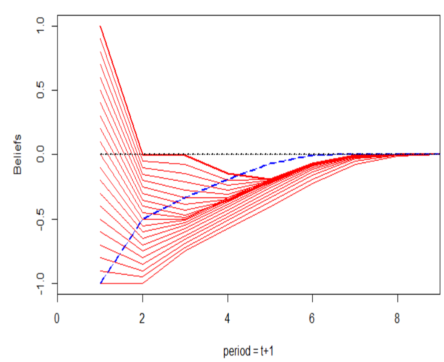
(c) $b_1^0 = 1, b_2^0 \in [-1, 1], b_3^0 = -1$



(d) $b_1^0 \in [-1, 1], b_2^0 = b_3^0 = -1$



(e) $b_1^0 = 1, b_2^0 \in [-1, 1], b_3^0 = 0$



(f) $b_1^0 \in [-1, 1], b_2^0 = -1, b_3^0 = 0$

Figure 3.12 – Convergence to extreme and non-extreme beliefs: $\mathbf{W}_0 = \mathbf{W}_{0,a}$

because no extreme or ϵ-extreme poles exist, consensus is neither extreme nor ϵ-extreme.

3.6 Discussions

As a consequence of proposition 3.4, in any network that is a complete (strongly connected) symmetric digraph[31] people can never converge to an extreme or ϵ-extreme belief, unless they all had extreme or ϵ-extreme beliefs from the start. Therefore, my results suggest that, if the goal is to prevent the convergence to extreme beliefs in a society, to do so the best way is through integration. By integration, I mean people can be connected with others, they may have a desire to listen to, but such that those individuals do not hold extreme views. When a society can make arrangements so that the individuals tagged as potential extremists are connected to less extreme individuals, then that society can lower the chances that the level of extremism of those individuals will stay the same. However, to decrease their level of extremism, it is better to put them in contact with individuals with very moderate or opposite beliefs. The worst that can happen is that the extremists voluntarily drop out of the network, due to their cognitive biases, and become isolated. The presence of a good mediator may also help with maintaining extreme believers for a longer time, in the network.

A network that has a rooted-tree structure (see figure 3.7, 3.8) or a structure like in figure 3.9, on the other hand, has more potential to drive social extremism when the most influential individuals in that network constitute an ϵ-extreme pole of that network. Specifically, if there exists only one cycle of size one in a weakly connected network, then all beliefs converge surely to the belief of the dictator. This is a very interesting result which gives a potential explanation to the underlying reason why we usually observe such a structure in the most influential extremist groups or in organizations such as the army. The model also shows that, to drive dictatorship in a network, a dictator doesn't need to be directly connected to their "subjects". Rather, all they have to do, is to ensure that a well-defined and unbroken hierarchy constantly remains in the network. Then, all they have to do is to become the main influence canal of a few most influential individuals in the network (even only one person can be enough, depending on their influence on others). However, to break the *chain of command* in such a network, the best action can just come through adding a strategical and powerful link between two individuals (or more), such that it creates additional cycles in the

31. In the literature on network topology, a complete symmetric digraph is a directed graph in which there is a directed edge from each vertex to every other one (Chartrand and Lesniak, 2005). Basically, it's a group with a zero-diagonal matrix.

network. Preferably these individuals should have opposite or very different views with the leader. Additional cycles means that information is likely to go back-and-forth, and bigger cycles means that more individuals are exchanging information. By adding more cycles in a network (through new connections between individuals), it may hinder or at least reduce the unthinking follower bandwagon effect, and sometimes increase wisdom. On a positive note, people can make smarter choices and decisions. Sometimes, fragmentation may ensue, which is not necessarily a negative outcome for a society. For instance, consider the case of countries governed by dictatorial regimes where the people's rights are constantly abused. In such countries fragmentation starting in the political sphere and driven by key influential individuals with better policy choices may have a positive and desired outcome. However, this example extrapolates my results.

3.7 Consensus Time

So far, I have established that in the presence of cognitive biases, convergence and consensus not only depend on the *a priori* network structure, but they also depend on the initial beliefs. If consensus is always reached in a network, it means that there exist a finite integer T such that the time to consensus T_C^{δ} is smaller than T, starting from any *a priori* belief vector $\mathbf{b}_0$. However, accounting for cognitive biases, makes consensus not always evident, even in the case of a strongly connected and aperiodic network $\mathcal{N}_0$; as in the example given in the proof of the proposition 3.1.

Note that, usually the time to consensus is measurable only up to a consensus distance δ. That is, if it exists, consensus is often obtained asymptotically. Therefore, consensus time in this paper refers to the superior bound, over all possible beliefs vectors, of the shortest time needed for the distance to consensus to fall below δ. For high precision, in all my simulations in this paper, I fix $\delta = 10^{-6}$. However, there are some special networks for which the consensus time mentioned is actually the time it takes for all beliefs to become equal. See Section 3.5.4 above for examples.

In a simple DeGroot-style model, the time to consensus depends strongly on the second largest (in modulus) eigenvalue of the network matrix. The spectrum of a graph gives a lot of information on its behavior. The larger the second eigenvalue, the longer the consensus time. Specifically, if $\lambda_2(\mathbf{W}_0)$ denotes the second-largest eigenvalue of $\mathbf{W}_0$, Golub and Jackson (2010) and Golub and Jackson (2012) express the consensus time in a DeGroot setting, as proportional to $\frac{log(n)}{log\left(\frac{1}{|\lambda_2(\mathbf{W}_0)|}\right)}$.

Intuitively, the link between consensus time and the second-largest eigenvalue of a graph, resides in the fact that, the later embodies some information about the connectivity of the graph. In fact, the larger the difference between the modulus of the first and the second-largest eigenvalue is, the better its connectivity (Brouwer and Haemers, 2011). In other words, the smaller the second-largest eigenvalue is, the lower is the probability that the network will become disconnected if some links are removed. And this is exactly the issue encountered in the presence of cognitive bias, leading to fewer cases of absolute consensus among strongly connected and aperiodic networks.[32] Therefore, instead of the consensus time proposed in Golub and Jackson (2012), I explore the probability of consensus as follows.

$$Pr(T_C^{\delta} < +\infty | \mathbf{W}_0, \mathbf{b}_0, \mathbf{B}) \approx F(|\lambda_2(\mathbf{W}_0)|, n, \mathbf{b}_0, \mathbf{B}) \tag{3.6}$$

$\mathbf{B}$ is a vector of parameters associated with the weights updating process. It captures the intensity of the cognitive biases used in $f()$ and $g()$ in assumption 3.1. In a general case where $f(x) = -ax + 2a + \epsilon,\ a > 0$, and $g(x) = bx + \epsilon,\ b > 0$, $\mathbf{B} = (a, b, \epsilon)$. In the examples used in the simulations, $\mathbf{B} = (-1, 2, .01)$.

F is a continuous and differentiable function with its first partial derivative satisfying $F_x < 0$. This means that, all else being equal, beliefs in a network with a relatively high modulus of $\lambda_2(\mathbf{W}_0)$ are less likely to converge to a consensus.

Complete networks like groups are the one with the highest possible connectivity. Therefore, next, I propose a model to predict the probability of consensus in such networks.

Predicted Probability of Consensus in Small *a Priori* Groups

Given that beliefs are more likely to converge to a consensus in groups, the predicted probability of consensus associated with each parameter considered is an upper bound of the probability of consensus starting from any network. For fast computation purpose, I focus on small groups of sizes 3, knowing that the consensus time also increases with the network size (all else being equal). In consequence, the probability of consensus for groups of sizes 3 is also an upper bound of the probability of consensus for any bigger network (*ceteris paribus*). Moreover, to avoid having too many trivial cases of consensus for groups of individuals with closer or similar *a priori* beliefs, I focus on

32. I use the term absolute consensus to indicate that a consensus is reached whatever the initial values in the *a priori* belief vector. Section 3.5.4 gives a few examples of networks leading to absolute consensus, even in the presence of cognitive biases.

groups in which at least two individuals have *a priori* beliefs that are opposite and ϵ-extreme (I fix $\epsilon = .05$). In that respect, the variance of beliefs will be great enough to ensure at least a few cases of polarization in beliefs.

For this purpose, I generate 100 *a priori* beliefs vectors and I compute their variance, and skewness.[33] Next, for each belief vector, I generate data on 1000 random *a priori* groups of sizes 3. For each group, I compute its eigenvalues and capture the second-largest one.[34] And for each matrix, and each belief vector, I propose an algorithm which allows me to compute numerically the consensus time.[35] I fixed the maximum time at $T = 10^4$ periods and the minimal distance between beliefs is 10^{-6}. Then, I retrieve the data on the 10^5 consensus time obtained, their associated *a priori* matrices eigenvalues, and the variance and skewness of their associated *a priori* belief vectors.

I then estimate a *probit* model on the dependent variable y:[36]

$$y = \begin{cases} 1 \text{ if } y^* < 10^4 \\ 0 \text{ if } y^* \geq 10^4 \end{cases}$$

where

$$y^* = \alpha + \mathbf{X}\beta + \eta \tag{3.7}$$

$\mathbf{X} = (log(\Lambda_2), \Sigma_2, |\kappa|)$. $\mathbf{y}^*$ is the vector containing all 10^5 possible values of consensus time (T_C^δ) computed. Λ_2 is the vector of all second-largest eigenvalues (in modulus) $\lambda_2(\mathbf{W}_0)$. Σ_2 is a vector containing all the 100 possible values of the variance (σ^2) of beliefs, repeated 1000 times for each matrix. Likewise, κ contains all the skewness values. Using the absolute value of κ in the model (3.7) is enough to know if the beliefs *a priori* are skewed or not. The side they lean towards is of less importance.

I estimate the parameters of the probit model, and the results are summarized in table C.1 in the appendix. They confirm that, even in the case of cognitive biases, the consensus time still depends strongly on the second-largest eigenvalue of the *a priori* matrix. The coefficients of the characteristics of the *a priori* belief vector (variance and skewness) are also significant. Last, I estimate the predicted value of the probability

33. Intuitively, I expect that if beliefs are very dispersed or polarized *a priori*, this might also influence the consensus time. I use the absolute value of the skewness as a measure of symmetry in beliefs. It tells us how much the beliefs are leaning together toward an extreme or the other. In fact, I expect that there should be a difference in consensus time for two distinct *a priori* belief vectors with similar variance (dispersion) but different skewness. In particular, if two individuals whose beliefs already lean toward each other exist in the network, beliefs should converge faster to a consensus.

34. In stochastic networks, the largest value is always 1. Therefore, the second-largest eigenvalue (in modulus) should be at most 1 (or -1).

35. All algorithms used for simulation purpose in this paper are available upon request.

36. For small groups of sizes 3, I take arbitrarily the period 10000 as an "approximation" of infinity.

of reaching a consensus for various values of $\lambda_2(\mathbf{W}_0), \sigma^2, \kappa$. Results are summarized in table C.2 in appendix C.3.

As expected, the smaller $\lambda_2(\mathbf{W}_0)$, the more likely consensus is to happen.[37] Similarly, when the beliefs of individuals in the network are less dispersed at first (smaller variance), consensus is more probable. Also, consensus is more probable in a network in which beliefs are skewed.

In all future figures representing beliefs and listening weights in this paper, the plain red line is related to individual 1, the dashed blue line, individual 2, and the dotted black one to the third individual.

3.8 Concluding Remarks

In this paper, I analyze the effects that some cognitive biases have on individuals learning behavior in a network. Generally, people's learning behavior usually depends on the networks' structure. In particular, I find that, in the presence of a confirmation bias and an extremist superiority bias, beliefs in strongly connected and aperiodic networks may not converge to a consensus. I propose an upper bound of the probability of consensus in groups. Specifically, I find that in groups for which the associated listening matrix has a smaller second largest eigenvalue, individuals' beliefs have more chances to converge to a consensus. In other words, the higher the network's connectivity is, the higher the chances of consensus. Similarly, in a network where individuals' *a priori* beliefs are more polarized, there are lower chances of convergence of beliefs to a consensus.

The underlying mechanism through which consensus is obtained is explained, and the role of individuals' self-confidence is also identified. In fact, I find that the existence of a consensus always requires that there exist a moment, over time, from when individuals' self-confidence monotonously (strictly) decreases. Put another way, the individual behavior leading to consensus in a network, is that there exists a moment from which everyone refrains from clinging to their own beliefs. It's this attitude of "letting go" which increases the chances of consensus. This is a concept we already know from organizational behavior, and the model help in understanding some plausible foundations. Individuals with a strong self-confidence need to find at least one peer whose beliefs are not too far from theirs, to become able to start that process of lowering the weight

37. Recall that the values displayed in table C.2 in appendix are upper bounds of the probability of consensus. This means that, although consensus is bound to happen when $\lambda_2(\mathbf{W}_0) = .05$, $\sigma^2 = .5$, and $\kappa = 0$, for bigger networks the probability of consensus may be smaller than 1.

they put on their own beliefs. Without such peers, there is an increased chance that beliefs stay (or become) polarized and fragmented. Specifically, in the paper, I discuss the necessity of a good mediator in a network, to promote consensus. The results show that even for networks where individuals put no weight on themselves, the absence of a closer peer (with not identical views), a mediator, may help maintain polarization in beliefs. This case is more evident when all *a priori* beliefs in the network are extreme ($b_i^0 \in \{-1, 1\}$), with at least two values different. In general, in most networks, the presence of a non-partisan mediator, increases the chances of consensus.

In the paper, I also define a few networks' structures always leading to consensus. Specifically, I address three cases. In the first case, for any *a priori* beliefs, the cognitive biases have no effect on consensus time, yet the path to consensus may slightly change. However in the second and third cases, cognitive biases affect both the path and time to consensus. In such networks in particular, the presence of an extreme (respectively an ϵ-extreme pole), guarantees that consensus is obtained at an extreme (respectively ϵ-extreme) value. I show that, in general, consensus at extreme (respectively ϵ-extreme) values always require the existence of an extreme (respectively ϵ-extreme) pole, as defined in the paper.

A potential limitation of this paper involves its lack of application to concrete data. A following step would be the implementation of an experimental framework through which data on different *a priori* network structures and beliefs vectors can be gathered to test the main results obtained.

3.9 Bibliography

Daron Acemoglu, Giacomo Como, Fabio Fagnani, and Asuman Ozdaglar. Opinion fluctuations and disagreement in social networks. Mathematics of Operations Research, 38(1):1–27, 2013.

Yann Algan, Nicoló Dalvit, Quoc-Anh Do, Alexis Le Chapelain, and Yves Zenou. Friendship networks and political opinions: A natural experiment among future french politicians. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3397092, May 2019.

Abhijit Banerjee, Emilie Breza, Arun G. Chandrasekhar, and Markus Mobius. Naive learning with uninformed agents. mimeo, 2016.

Vincent D. Blondel, Julien M. Hendrickx, Alex Olshevsky, and John N. Tsitsiklis. Convergence in multiagent coordination, consensus, and flocking. In 44th IEEE conference on Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC '05. IEEE, 2005.

Vincent Boucher. Structural homophily. International Economic Review, 56(1), 2015.

Vincent Boucher, Yann Bramoullé, Habiba Djebbari, and Bernard Fortin. Do peers affect student achievement? evidence from canada using group size variation. Journal of Applied Econometrics, 29:91–109, 2014.

Mark J. Brandt, Anthony M. Evans, and Jarret T. Crawford. The unthinking or confident extremist? political extremists are more likely than moderates to reject experimenter-generated anchors. Psychological Science, 26(2):189–202, 2015.

Andries E. Brouwer and Willem H. Haemers. Spectra of graphs. Springer, 2011. Monograph.

Arun G. Chandrasekhar, Horacio Larreguy, and Juan Pablo Xandri. Testing models of social learning on networks: Evidence from a lab experiment in the field. NBER Working Paper, (21468), 2015.

Gary Charness and Chetan Dave. Confirmation bias with motivated beliefs. Games and Economic Behavior, 104:1–23, 2017.

Gary Chartrand and Linda Lesniak. Graphs and Digraphs. CRC Press, fourth edition, 2005.

S. Chatterjee and E. Seneta. Towards consensus: Some convergence theorems on repeated averaging. Journal of Applied Probability, (14):89–97, 1977.

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

Peter M. DeMarzo, Dimitri Vayanos, and Jeffrey Zwiebel. Persuasion bias, social influence, and unidimensional opinions. The Quaterly Journal of Economics, 118:909–968, 2003.

Franscesco Drago and Roberto Galbiati. Indirect effects of a policy altering criminal behavior: Evidence from italian prison experiment. American Economic Journal: Applied Economics, 4(2):199–218, 2012.

L. Elsner, I. Koltracht, and M. Neumann. On the convergence of asynchronous paracontractions with application to tomographic reconstruction from incomplete data. Linear Algebra and its Applications, (130):65–82, 1990.

Elfried Faton and Bernard Fortin. Peer effects in effort within maternal and child care workers in benin. mimeo, 2017.

Bernard Fortin and Mira Yazbeck. Peer effects, fast food consumption and adolescent weight gain. Journal of Health Economics, (42):125–38, July 2015.

Benjamin Golub and Matthew Jackson. How homophily affects the speed of learning and best-responses dynamics. The Quaterly Journal of Economics, 127:1287–1338, 2012.

Benjamin Golub and Matthew O. Jackson. Naive learning in social networks and the wisdom of crowds. American Economic Journal: Microeconomics, 2(1):112–149, 2010.

Jonathan Gross and Jay Yellen. Graph Theory and its Applications. CRC Press, second edition, 2006.

Cameron R. Hopkin, Rick H. Hoyle, and Kaitlin Toner. Intellectual humility and reactions to opinions about religious beliefs. Journal of Psychology and Theology, 42 (1):50–61, 2014.

Matthew O. Jackson. Social and Economics Networks -, chapter 8, pages 223–256. Princeton University Press, 2008.

Charles G. Lord, Lee Ross, and Mark R. Lepper. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. Journal of Personality and Social Psychology, 37(11):2098–2109, 1979.

Brandon R. McFadden and Jayson L. Lusk. Cognitive biases in the assimilation of scientific information on global warming and genetically modified food. Food Policy, 54:35–43, 2015.

Manuel Mueller-Frank and Claudia Neri. Social learning in networks: Theory and experiments. mimeo, 2013.

Sam Mullins and Adam Dolnik. An exploratory, dynamic application of social network analysis for modelling the development of islamic terror-cells in the west. Behavioral Sciences of Terrorism and Political Agression, 2(1):3–29, 2010. DOI:10.1080/19434470903319441.

Stuart Nelson and Michael Neumann. Generalizations of the projection method with application to sor theory for hermitian positive semidefinite linear systems. Numerische Mathematik, (51):123–141, 1987.

Michael Neumann and Hans Schneider. The convergence of general products of matrices and the weak ergodicity of markov chains. Linear Algebra and its Applications, 287: 3017–314, 1999.

Raymond S. Nickerson. Confirmation bias: A ubiquitous phenomenon in many guises. Review of General Psychology, 2(2):175–220, 1998.

Tara O. The Collapse of North Korea: Challenges, Planning and Geopolitics of Unification. Palgrave Macmillan; 1st ed. 2090 edition, 2015.

Matthew Rabin and Joel L. Schrag. First impressions matter: A model of confirmatory bias. The Quaterly Journal of Economics, 114(1):37–82, 1999.

Pouya Rezaienia, Bahman Gharesifard, Tamas Linder, and Behrouz Touri. Push-sum on random graphs. mimeo, 2017.

Eugene Seneta. Coefficients of ergodicity: Structure and applications. Advances in Applied Probability, 11(3):576–590, September 1979.

David Skarbek. The Social Order of the Underworld: How Prison Gangs Govern the American Penal System, chapter 5. Oxford Scholarship Online, 2014.

Kaitlin Toner, Mark R. Leary, Michael W. Asher, and Katrina P. Jongman-Sereno. Feeling superior is a bipartisan issue: Extremity (not direction) of political views predicts perceived belief superiority. Psychological Science, 24(12):2454–2462, 2013.

Kaitlin Toner Raimi and Mark R. Leary. Belief superiority in the environmental domain: Attitude extremity and reactions to fracking. Journal of Environmental Psychology, 40:76–85, 2014.

B. Touri and A. Nedic. On backward product of stochastic matrices. Automatica, (48): 1477–1488, August 2012.

David Webber and Arie W Kruglanski. The social psychological makings of a terrorist. Current Opinion in Psychology, (19):131–134, 2017.

Ronald Wintrobe. Rational Extremism. Cambridge University Press, 2006.

Conclusion

My thesis has provided some new insights on the role played by social interactions on labor outcomes like effort and productivity, as well as individuals' beliefs. Specifically, its main contribution to the literature on the economics of networks is twofold. First, on the empirical side, in a context of fixed salaries irrelevant of workers' performance, Chapters 1 and 2 investigate how the results on peer influence and peer effects depart from the standard ones obtained so far in the literature. Some details about the mechanisms at play are also given. Second, Chapter 3, which fits more into a theoretical framework, explains how the beliefs of individuals subject to some cognitive biases, and affected by their peers' beliefs, vary in a dynamic setting. In the chapter, I present some properties for reaching consensus over time.

Even more specifically, Chapter 1 explains how the level of effort provided by maternal and child healthcare (MCH) workers from a developing country, Benin, is influenced by their co-workers outcome and characteristics. In the chapter, we estimate a Spatial Autoregressive (SAR) model on MCH workers' performance, and we find evidence of strategic substitutability among them. In other words, some workers tend to free ride or shirk when their peers show higher levels of effort and productivity. We assert that their payment scheme which is fixed and independent of their performance plays a major role in this result. In addition, the findings indicate that working with peers who have an administrative mandate to monitor them, tend to induce a positive effect on workers' productivity. Furthermore, workers who have more children, and respectively those who have more experience within the health facility, tend to make less effort than their peers with fewer children, respectively those with less experience. On another note, we find that higher levels of unpaid wages induce a lower level of effort and quality of healthcare produced by MCH workers.

Chapter 3 on another hand develops a theoretical framework explaining how cognitive biases influence the beliefs updating process for individuals connected under a particular network. In fact, the chapter reveals that cognitive biases are a key impediment to consensus. Specifically, I find that for some networks which still have the particularity of being strongly connected and aperiodic initially, accounting for individuals confirmation bias and extremist relative superiority bias into their beliefs updating process may nor lead to consensus. My analyses indicate that, in those specific cases, cognitive biases cause the severance of links, which, in turn, prevent consensus.

Lastly, Chapter 2 complements the results in Chapter 1, and proposes a Nash-bargaining approach that explains the division of labor among MCH workers from Benin. The results show that workers bargaining power has an important role in their workload share; and this bargaining power depends on some of their individual characteristics, like their level of education, experience and the number of their children. Particularly, workers with less education, like most nursing auxiliaries, have less bargaining power when it comes to choosing their workload share within the total demand for healthcare addressed to their health facility. In addition, workers who have more children tend to have less bargaining power in comparison with their peers of same characteristics and fewer children. Moreover, workers experience tend to be positively associated with their bargaining power in their first nineteen working years or so.

Appendix A

Appendix to Chapter 1

A.1 Proof of proposition 1.1

First, let's consider the following proposition.

Proposition A.1. *Consider a game of effort with rational players. For any player i who receive utility (or payoff) $u_i(e_i, \boldsymbol{e_{-i}})$ defined as in equation (1.2) for exerting effort e_i, there exists a real-valued vector $(\underline{E}_i, \overline{E}_i)$, such that:*

(i) Player i always chooses their strategy in $[\underline{E}_i, \overline{E}_i]$.

(ii) In particular, $\underline{E}_i$ and $\overline{E}_i$ are images of $\boldsymbol{e_{-i}}$ through a real valued function defined on $\mathbb{R}^{n-1}$.

Proof. A rational player always tries to maximize their payoff. If the player decides not to provide any effort ($e_i = 0$, also equals doing nothing), then he receives utility 0. If by doing nothing a worker is guaranteed to receive at least a payoff $u_i(e_i, \boldsymbol{e_{-i}}) \geq 0$ (whatever the other players do), then they will choose to exert effort $e_i \in \mathbb{R}^*$ if and only if it gives them at least 0. For any player i, given scalars and matrices $f_g, \mathbf{X}, \eta_i, \mathbf{W}_i, \upsilon, \tau, \beta$, if $M_i(\boldsymbol{e_{-i}})$ designates the quantity: $f_g + \mathbf{x}_i\upsilon + \eta_i + W_i\mathbf{x}\tau + \beta W_i\boldsymbol{e}$, then $M_i(\boldsymbol{e_{-i}})$ is the absolute maximum payoff of i. Consequently, a rational player i always choses strategy e_i such that $0 \leq u_i(e_i, \boldsymbol{e_{-i}}) \leq \frac{M_i(\boldsymbol{e_{-i}})^2}{2}$. Finally, given the variations of u_i, $\min(0, 2M_i(\boldsymbol{e_{-i}})) \leq e_i \leq \max(0, 2M_i(\boldsymbol{e_{-i}}))$. Hence, $\underline{E}_i$ and $\overline{E}_i$ exist, with $\underline{E}_i = \min(0, 2M_i(\boldsymbol{e_{-i}}))$ and $\overline{E}_i = max(0, 2M_i(\boldsymbol{e_{-i}}))$. □

Existence of the equilibrium (1.6): According to proposition A.1, the set of strategies adopted by any individual i, is equivalent to an interval $[\underline{E}_i, \overline{E}_i] \subset \mathbb{R}$, which is a closed and bounded subset of $\mathbb{R}$; and thus compact, according to Borel-Lebesgue (or Heine-Borel) theorem. Consequently, according to the theorem of the maximum, the set $\Theta(\boldsymbol{e_{-i}}) = \arg\text{Max}_{\ e_i \in [\underline{E}_i, \overline{E}_i]}\, u_i(e_i, \boldsymbol{e_{-i}})$ is non-empty and has a compact image. In addition, the topological space product, $\prod_{i=1}^{n} [\underline{E}_i, \overline{E}_i]$, is compact and also convex. Therefore, from Kakutani's fixed-point theorem, a Nash equilibrium of problem 1.5 exists.

Uniqueness: Using equation (1.2) in the optimization problem (1.5), the F.O.C. gives the unique form (1.6) of the interior solution. Moreover, the condition $|\beta| < \dfrac{1}{\|\mathbf{W}\|}$ guarantees that $(\mathbf{I} - \beta\mathbf{W})$ is invertible.

A.2 Descriptive statistics and Data content

Table A.1 – Topics for the skills and knowledge test questions

Topics	Questions answered
Focused ANC	1- Elements of focused ANC (Q4) 2- Pregnant woman to pay more attention to during focused ANC (Q5) 3- Labour signs (Q6)
Record keeping	Record during labour delivery (Q8)
Labour, Delivery and immediate Postpartum care	1- Monitoring labour delivery (Q7) 2- Third stage of labour (Q9) 3- What to do in case of non hemorrhagique placental retention (Q13)
Newborn care	1-Immediate newborn care during previous assisted delivery (Q10) 2- Signs and symptoms of infection (Q14) 3- First steps in case of infection (Q15) 4- Care if $birthweight < 1.5kg$ (Q16) 5- Care if $1.5kg < birthweight < 2.5kg$ (Q17)
Bleeding	1- If pregnant woman bleeds: Signs to be careful to (Q11) 2- PPH: Signs to be careful to (Q12)
Post-abortion care	1- List complications due to unsafe abortion (Q18) 2- Actions to take if woman suffer from complications due to unsafe or uncompleted abortion (Q19) 3- Informations to give to patient treated for complications due to unsafe or uncompleted abortion (Q20)
Standard precautions and cleanliness	1-Actions to prevent infections in the HF (Q22) 2- How to prepare decontamination solution (Q23) 3- How to preserve decontamination solution (Q24) 4- How to maintain the contaminated material (Q25)
Rape	Actions if woman is victim of rape (Q21)

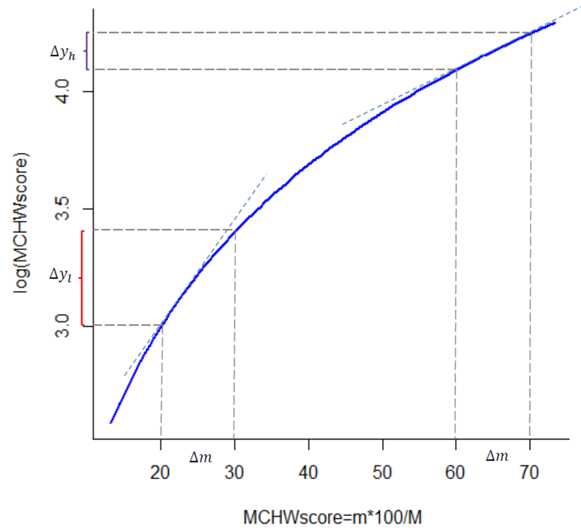


Figure A.1 – Representing the proxy variable in function of MCHWscore

Table A.2 – Descriptive statistics

Variable	Mean	Median/ Mode	Stand. Dev.	Range
MCHWs level				
Gender (female=1)	0.95	(1) *Fem.*	0.21	$\{0,1\}$
Age	35.31	33	8.26	$[15,70]$
Marital status (married=1)	0.62	(1) Married	0.48	$\{0,1\}$
Number of children*	2.27	2	1.49	$[0,9]$
Tenure (in years) *	5.22	4	5.10	$[0.08,34]$
Practical experience:				
-Birth attendance last month	10.13	7	11.76	$[0,82]$
-Birth attendance during last 6 months*	57.52	47	49.73	$[0,343]$
Education level:	2.47	(2)*Sec. cycle*1	0.68	$\{1,2,3,4\}$
		Proportions:		
(1) Primary school only		6.43%		
(2) Some secondary school-cycle 1		44.29%		
(3) More secondary school-cycle 2		44.46%		
(4) Some university level		4.82%		

* These statistics are taken on the whole imputed data. However, because we did some mean imputations, it doesn't change these statistics much. The median also was still the same.

Table A.3 – Additional descriptive statistics

HF level			
Status of the HF		Proportions	
	public	92.82%	
	semi-public	1.08%	
	private (for-profit)	2.87%	
	NGO	1.08%	
	Religious/confessional	2.15%	
Number of groups...	233		
... of size 3 or more	128		
	Average	Stand. Dev.	Values Range
Population size (for 233)	5.16	3.61	$[1, 25] \cap \mathbb{N}$
Sample size (for 233)	2.67	0.69	$\{1, 2, 3, 4, 5\}$
Population size (for 128)	6.17	3.78	$[3, 25] \cap \mathbb{N}$
Sample size (for 128)	3.06	0.33	$\{3, 4, 5\}$

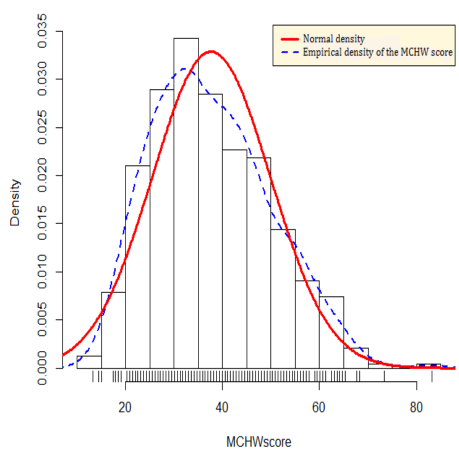
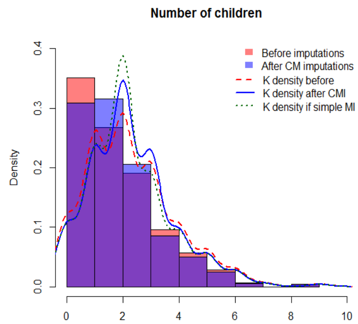


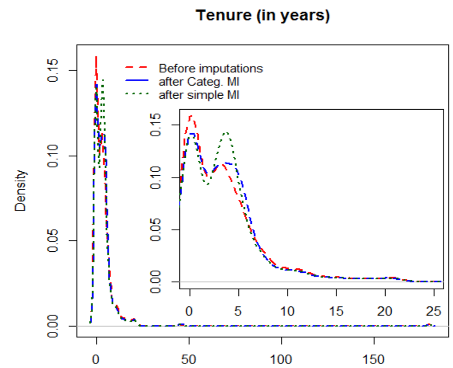
Figure A.2 – Empirical density of MCHW score vs normal density

Table A.4 – p-values of Kruskal-Wallis test on the MCHW score between professional categories

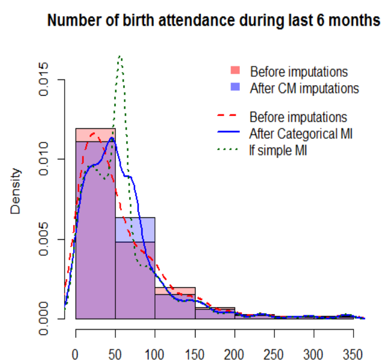
Category	Obstetricians	Midwives	Nurses	Nursing Auxiliaries
Obstetricians	-	0.1633	0.01705	transitivity
Midwives	-	-	$2.035e-05$	transitivity
Nurses	-	-	-	$1.904e-13$
Nursing Auxiliaries	-	-	-	-



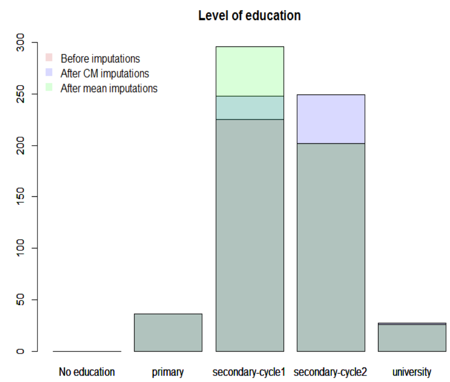
(a) Number of children



(b) Number of years of experience in the HF



(c) Number of births attended during last 6 months



(d) Level of education

Figure A.3 – Density histograms and empirical density lines of some independent variables before and after job-position mean imputations

A.3 Results Tables

Table A.5 – Generalized IV results (accounting for education level)

	Endogenous effect (1)	Individual effects (1)	Contextual effects (1)
Peer Y	-0.574***		
	(0.193)		
	[0.260]		
Hours worked per week		0.0006	0.0005
		(0.00056)	(0.0021)
		[0.0007]	[0.0027]
Education (max)		0.0186**	0.0549
		(0.0082)	(0.0336)
		[0.0102]	[0.0431]
Tenure ($18 years = 0$)		-0.0065*	-0.027**
		(0.0039)	(0.013)
		[0.0049]	[0.016]
Responsible midwife or doctor		0.399***	-0.259
		(0.068)	(0.258)
		[0.078]	[0.308]
Responsible nurse		0.120 *	-0.315
		(0.087)	(0.314)
		[0.116]	[0.419]
Simple midwife or doctor		0.234***	-0.170
		(0.079)	(0.278)
		[0.097]	[0.35]
Simple Nurse		0.078	-0.434*
		(0.076)	(0.253)
		[0.098]	[0.324]
Test of weak instruments	43.18 ***		
Wu-Hausman test (endogeneity of WY)	64.44***		

Significant at 1%(***) – 5%(**) – 10%(*) SE in brackets ( )
Significant at 1%(***) – 5%(**) – 10%(*) Cluster-Robust SE in brackets [ ].

Table A.6 – Generalized IV results (no education level)

	Endogenous effect (2)	Individual effects (2)	Contextual effects (2)	Endogenous effect (3)	Individual effects (3)	Contextual effects (3)
Peer Y	−0.661***			−0.638***		
	(0.190)			(0.190)		
	[0.252]			[0.242]		
Hours worked per week		0.0009	0.00143		0.0008	0.0011
		(0.00056)	(0.0021)		(0.00054)	(0.002)
		[0.00071]	[0.0027]		[0.00071]	[0.0027]
Tenure ($18years = 0$)		−0.0072*	−0.029**		−0.008**	−0.0305**
		(0.00387)	(0.013)		(0.004)	(0.0135)
		[0.0048]	[0.016]		[0.005]	[0.0167]
Responsible midwife or doctor		0.493***	0.538 **		0.498***	0.558**
		(0.058)	(0.230)		(0.057)	(0.234)
		[0.068]	[0.278]		[0.072]	[0.291]
Simple midwife or doctor		0.331***	0.1523		0.341***	0.169
		(0.071)	(0.284)		(0.073)	(0.296)
		[0.086]	[0.356]		[0.088]	[0.37]
Responsible nurse		0.166 *	−0.156		0.181**	−0.119
		(0.086)	(0.329)		(0.092)	(0.364)
		[0.114]	[0.439]		[0.12]	[0.46]
Simple Nurse		0.107	−0.303		0.115	−0.287
		(0.076)	(0.261)		(0.079)	(0.279)
		[0.099]	[0.338]		[0.102]	[0.345]
log(1+ unpaid wage)		–	–		−0.008**	−0.0305**
					(0.004)	(0.0135)
					[0.005]	[0.017]
log(1+ children)		–	–		−0.008**	−0.0305**
					(0.004)	(0.0135)
					[0.005]	[0.00167]

	Statistics	Statistics
Test of weak instruments	41.50 ***	63.88 ***
Wu-Hausman test (endogeneity of WY)	25.85***	41.18***

Significant at 1%(***) – 5%(**) – 10%(*) SE in brackets ( )

Significant at 1%(***) – 5%(**) – 10%(*) Cluster-Robust SE in brackets [ ].

Table A.7 – Generalized IV results with additional single individuals HF groups in sample

	Endogenous effect (4)	Individual effects (4)	Contextual effects (4)	Endogenous effect (5)	Individual effects (5)	Contextual effects (5)
Peer Y	−0.693***			−0.743***		
	(0.187)			(0.177)		
	[0.242]			[0.227]		
Hours worked per week		0.001*	0.0018		0.001 *	0.0018
		(0.00056)	(0.002)		(0.00054)	(0.002)
		[0.00070]	[0.0025]		[0.00068]	[0.0025]
Tenure ($18 years = 0$)		−0.0081**	−0.031**		−0.0088**	−0.032***
		(0.00369)	(0.012)		(0.0038)	(0.0122)
		[0.0044]	[0.014]		[0.0046]	[0.0148]
Responsible midwife or doctor		0.471***	0.485**		0.476***	0.518**
		(0.053)	(0.223)		(0.052)	(0.220)
		[0.062]	[0.268]		[0.062]	[0.267]
Simple midwife or doctor		0.270***	0.018		0.285***	0.083
		(0.073)	(0.281)		(0.076)	(0.292)
		[0.089]	[0.350]		[0.093]	[0.367]
Responsible nurse		0.143	−0.219		0.161*	−0.147
		(0.089)	(0.313)		(0.090)	(0.321)
		[0.118]	[0.415]		[0.120]	[0.425]
Simple Nurse		0.117	−0.262		0.134*	−0.192
		(0.079)	(0.269)		(0.080)	(0.272)
		[0.101]	[0.343]		[0.102]	[0.347]
log(1+ children)		–	–		0.044	0.088
					(0.043)	(0.137)
					[0.053]	[0.169]
	Statistics			Statistics		
Test of weak instruments	35.47 ***			44.94 ***		
Wu-Hausman test (endogeneity of WY)	20.43***			27.12***		

Significant at 1%(***) – 5%(**) – 10%(*) SE in brackets ( )

Significant at 1%(***) – 5%(**) – 10%(*) Cluster-Robust SE in brackets [ ].

Table A.8 – GIV results with additional single individuals and without contextual effects

	Endogenous effect (6)	Individual effects (6)	Endogenous effect (5)	Individual effects (5)
Peer Y	−0.275*		−0.279*	
	(0.167)		(0.163)	
Hours worked per week		0.0004		0.0005
		(0.00044)		(0.0003)
Education		.010**		0.009**
		(0.006)*		(0.005)
Responsible midwife or doctor		0.329***		0.333***
		(0.042)		(0.035)
Simple midwife or doctor		0.276***		0.282***
		(0.049)		(0.042)
Responsible nurse		0.212***		0.217***
		(0.065)		(0.05)
Simple Nurse		0.201***		0.205***
		(0.064)		(0.049)
log(1+ children)		–		0.026
				(0.022)

	Statistics	Statistics
Test of weak instruments	10.46 ***	10.16 ***
Wu-Hausman test (endogeneity of WY)	56.23***	58.2***

Significant at 1%(***) – 5%(**) – 10%(*) SE in brackets ( )

Significant at 1%(***) – 5%(**) – 10%(*) Cluster-Robust SE in brackets [ ].

Appendix B

Appendix to Chapter 2

B.1 Production Technology for MCH workers

I estimate the parameter γ, using a log-linear form of equation (2.1), as given by the regression model:

$$\ln(q_i) = \gamma \ln(h_i) + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, \sigma^2) \tag{B.1}$$

Assuming that the technology of production of a worker is also affected by a constant term $A \neq 1$ which also influences the returns to worker's hours, we have: $q_i = Ah_i^\gamma$. Then, the return to scale parameter γ can be estimated using a log-linear form of the modified technology, as in the regression model below:

$$\ln(q_i) = \ln(A) + \gamma \ln(h_i) + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, \sigma^2) \tag{B.2}$$

The summary of the OLS estimations of models (B.1) and (B.2) are given in the following table.

OLS	$Model(B.1)$			$Model(B.2)$		
	Coeff.		95% conf. int.	Coeff.		95% conf. int.
$\hat{\gamma}$	.323	$***$	$[.307, .341]$	.328	$**$	$[.098, .559]$
	(.009)			(.117)		
$Constant: \ln(A)$	–			−.028		$[-1.39, 1.33]$
	–			(.693)		
$R^2(adjusted)$	.768			.016		
N	428			428		

Standard errors in brackets (). $***$Significant at 1‰. $**$ Significant at 1%.

The results in table B.1, clearly show that the marginal return parameter γ is effectively below 1, and more precisely is around .3. Overall, the estimations results are more in favour of the

Table B.1 – OLS estimates of the marginal return γ

model (B.1), which is related to the form specified in the theoretical description of the model (see equation 2.1). In model (B.2), the constant term estimate is not significantly different from 0. In addition, the R^2 is very low, suggesting a bad specification. On the other hand, the value estimated for the constant $\ln(A)$ suggests that (if it were significant) $A \approx .97$, which is very close to 1.

B.2 Nash-bargaining problem with heterogeneous bargaining powers

It is also possible to consider different bargaining powers among co-workers within a same HF. To do so, the Nash-bargaining problem to solve takes the form:

$$\max_{u_1,...,u_n} \prod_{i=1}^{n} (u_i - d_i)^{\alpha_i}, \ s.t. \ (u_1, ..., u_n) \in U, \ (u_1, ..., u_n) > (d_1, ..., d_n) \tag{B.3}$$

where α_i denotes the bargaining power of individual i. In model (B.3) defined as such, we cannot directly estimate workers bargaining powers. However, if one is up to the challenge, it is possible to use refined Bayesian methods like the Monte Carlo Markov Chains or Gibbs sampling to estimate the parameters defining workers disagreement outcome.

B.3 Proof of theorem 2.1

For a simplicity of notations, let:

$$\begin{aligned} b_i = & \ \beta ln(y_i + 1) - d_0 - \lambda_s s_i - \lambda_t t_i - \lambda_{t2} t_i^2 - \lambda_{ch} ch_i \\ B_i = & \ B(r_i) = b_i - (1-\beta) r_i Q \end{aligned}$$

The existence and uniqueness of a solution for problem (2.7) over the feasible set $\Omega_r = \{(r_1, ..., r_n) \in \mathbb{R}_+ : \sum_i r_i = 1, B(r_i) > 0\}$ comes from the strict concavity of the objective function over the convex region Ω_r.

Proof that Ω_r is a convex set:
Let $\mathbf{x} = (x_1, ..., x_n)$, $\mathbf{z} = (z_1, ..., z_n) \in \Omega_r$, $\mathbf{x} \neq \mathbf{z}$, and $\alpha \in (0, 1)$. Then $\sum_i ((1-\alpha)x_i + \alpha z_i) = (1-\alpha)\sum_i x_i + \alpha \sum_i z_i = 1$; and $(1-\alpha)x_i + \alpha z_i > 0$.

Proof that $\prod_{i=1}^n B(r_i)$ is strictly concave over Ω_r:

$$\begin{aligned} \prod_{i=1}^n B((1-\alpha)x_i + \alpha z_i) = & \ \prod_{i=1}^n (b_i - (1-\alpha)(1-\beta)Qx_i - \alpha(1-\beta)Qz_i) \\ = & \ \prod_{i=1}^n ((1-\alpha)(b_i - (1-\beta)Qx_i) + \alpha(b_i - (1-\beta)Qz_i)) \\ > & \ (1-\alpha)\prod_{i=1}^n B(x_i) + \alpha \prod_{i=1}^n B(z_i) \end{aligned}$$

The last inequality results from the fact that $B(x_i) > 0$ and $B(z_i) > 0$ for all $\mathbf{x}, \mathbf{z} \in \Omega_r$. Therefore a unique maximum of the objective function exists.

Form of the interior solution Given the constraints $B_i > 0 \; \forall \; i$ in the optimization problem (2.7), its solution is equivalent to that of the natural logarithm of the objective function, given the same constraints. Replacing the binding constraint $\sum_i r_i = 1$ in the natural logarithm of the objective, the Lagrangian is expressed as follows:

$$\begin{aligned} \mathcal{L}(\boldsymbol{r}, \boldsymbol{\mu}) = \; & \textstyle\sum_{\substack{j=1 \\ j \neq i}}^{n} ln \left(\beta ln(y_j + 1) - (1-\beta) Q (1 - r_i - \sum_{\substack{k=1 \\ k \neq i,j}}^{n} r_k) - d_0 - \lambda_s s_j - \lambda_t t_j - \lambda_{t2} t_j^2 - \lambda_{ch} ch_j \right) \\ & + ln(B_i) + \textstyle\sum_{i=1}^{n} \mu_i B_i \end{aligned}$$

where $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)$ represents Lagrange parameters. The first order (Kuhn and Tucker) conditions give the following:

$$\begin{aligned} &(1)\; 0 = \frac{\partial \mathcal{L}}{\partial r_i} \; \forall \; i && \Leftrightarrow \; 0 = (1-\beta) Q \left(\textstyle\sum_{j \neq i} B_j^{-1} - B_i^{-1} - \sum_i \mu_i \right) \; \forall \; i \\ & && \Leftrightarrow r_i = \frac{\beta ln(y_i + 1) - d_0 - \lambda_s s_i - \lambda_t t_i - \lambda_{t2} t_i^2 - \lambda_{ch} ch_i - \left(\sum_{j \neq i} B_j^{-1} - \sum_i \mu_i \right)^{-1}}{Q(1-\beta)} \\ &(2)\; \mu_i \geq 0 \; \forall \; i && \\ &(3)\; 0 = \mu_i \frac{\partial \mathcal{L}}{\partial \mu_i} \; \forall \; i && \Leftrightarrow \; \mu_i = 0, \; B_i > 0 \; \forall \; i \end{aligned}$$

Replacing condition (3) in the solution in (1), the solution ensues, as defined in equation (2.8).

Appendix C

Appendix to Chapter 3

C.1 Mathematical Appendix

C.1.1 Miscellaneous Definitions and Propositions

Definition C.1. *Influence, influence set, and influence graph*
For any individual i in a network $\mathcal{N}$, the influence of a peer j on i after (or at) period t, is given by the $(i,j)^{th}$ element of the matrix product $\prod_{m=0}^{t} \boldsymbol{W}_{t-m}$.[1] *I say j (or j's belief) influences i if that quantity is strictly positive. In the case where beliefs converge, the "long-term" influence of a peer j on an individual i is given by the limiting value of the $(i,j)^{th}$ element of $\prod_{m=0}^{t} \boldsymbol{W}_{t-m}$ as time goes to infinity. In the particular case when consensus is obtained, the long-term influence of an individual j on all their peers is the same.* DeMarzo et al. (2003) *call it the social influence of j.*
Similarly, I define the influence set of an individual i at time t as the set of all peers js who influence i at time t. An influence graph at time t is the digraph which represents the matrix $\prod_{m=0}^{t} \boldsymbol{W}_{t-m}$. The influence set of a group of individuals at a given period is the reunion of their individual influence sets at that period.

Proposition C.1. *Let $(\boldsymbol{W}_t, \boldsymbol{b}_t)_t$ a sequence of $n \times n$ listening matrices and beliefs vectors updated through processes 3.4 and 3.2, such that $\boldsymbol{W}_0$ is lower (resp. upper) triangular with $w_{11}^0 = 1,\ w_{ii}^0 = 0\ \forall\ i \neq 1$ (resp. $w_{nn}^0 = 1,\ w_{ii}^0 = 0\ \forall i \neq n$). Then for all a priori vector $\boldsymbol{b}_0$, and for all period t, $\boldsymbol{W}_t$ is also a lower (resp. upper) triangular matrix, such that $w_{11}^t = 1,\ w_{ii}^t = 0\ \forall\ i \neq 1$ (resp. $w_{nn}^t = 1,\ w_{ii}^t = 0\ \forall i \neq n$).*

1. Note here that, as defined in the paper by DeMarzo et al. (2003), for all periods $t \neq 0$, there is a clear distinction between the *influence* of a peer j on i after (or at) period t and their *direct influence* on i at period t, which is the listening weight w_{ij}^t.

Proof. Trivial. It comes from the logical implications:

$$w_{ij}^0 = 0 \Rightarrow w_{ij}^t = 0,\ \forall\ i,\ j,\ t$$
$$w_{ij}^0 = 1 \Rightarrow w_{ij}^t = 1,\ \forall\ i,\ j,\ t$$

□

C.1.2 Proofs of propositions and lemmas

Proposition 3.2

Let $\mathcal{N}_0$ a strongly connected and aperiodic *a priori* network. Then, for all finite periods, $\mathcal{N}_t$ is also strongly connected and aperiodic.[2] Consequently, $\lim_{t\to+\infty} b_i^t$ exists for all i, and so does $\lim_{t\to+\infty} w_{ij}^t$ for all $i,\ j$. To demonstrate the equivalence in the proposition above let's demonstrate each implication step by step:

(i) $\Rightarrow$ *(ii)*: Assume that beliefs converge to a consensus b such that $|b| \neq 1$.

By definition, from equation 3.4, because $\lim_{t\to+\infty} w_{ij}^{t+1}$ exists, then $\lim_{t\to+\infty} w_{ij}^{t+1} = \lim_{t\to+\infty} w_{ij}^t,\ \forall\ i,\ j$. Which implies:

$$\begin{aligned}
& \lim_{t\to+\infty} f(|b_i^t - b_j^t|) = && \lim_{t\to+\infty} [\sum_{k\neq i} w_{ik}^t f(|b_i^t - b_k^t|) + w_{ii}^t g(|b_i^t|)]\ \forall\ i,\ j \\
\Rightarrow\quad & f(0) = && f(0) \lim_{t\to+\infty}\left(\sum_{k\neq i} w_{ik}^t\right) + g(|b|)\left(\lim_{t\to+\infty} w_{ii}^t\right) \\
\Rightarrow\quad & 0 = && f(0)\left(\lim_{t\to+\infty}\left(\sum_{k\neq i} w_{ik}^t\right) - 1\right) + g(|b|)\left(\lim_{t\to+\infty} w_{ii}^t\right) \\
\Rightarrow\quad & 0 = && 2\,(|b|-1)\left(\lim_{t\to+\infty} w_{ii}^t\right) \\
\Rightarrow\quad & \lim_{t\to+\infty} w_{ii}^t = 0,\ \forall\ i &&
\end{aligned}$$

Lemma 3.1

Step 1: The product $[\Pi_{l,2}$ of two lower triangular $\mathbf{W}_{rt}$-type matrices $\mathbf{P}_{l1}$ and $\mathbf{P}_{l2}$ (respectively upper triangular, $\mathbf{P}_{u1}$ and $\mathbf{P}_{u2}$) gives a lower (respectively upper) triangular $\mathbf{W}_{rt}$-type matrix such that its elements on line j and column $j-1$ (respectively on column j and line $j-1$) are zeros, for all $j \in \{3, 4, \ldots, n\}$.

2. In the absolute, this statement is true. However, I acknowledge that *asymptotically*, $\mathcal{N}_t$ can "become assimilated" to a periodic, or weakly connected or unconnected network. Yet, it does not change the fact that the network will always stay strongly connected and aperiodic in finite time.

This is straightforward. Replacing $p_{31,l1} + p_{32,l1} = 1$ (elements of matrix $\mathbf{P}_{l1}$), we have:

$$\Pi_{l,2} = \mathbf{P}_{l1}\mathbf{P}_{l2} = \begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ 1 & 0 & \dots & \dots & \dots & 0 \\ 1 & 0 & 0 & \dots & \dots & \vdots \\ p_{41,l1} + p_{42,l1} + p_{43,l1}p_{31,l2} & p_{43,l1}p_{32,l2} & 0 & 0 & \dots & \vdots \\ \vdots & \dots & \dots & \ddots & \ddots & \vdots \\ p_{n1,l1} + p_{n2,l1} + \sum_{j=3}^{n-1} p_{nj,l1}p_{j1,l2} & \dots & \dots & \dots & 0 & 0 \end{bmatrix}$$

Step 2: Let $(\mathbf{P}_{lm})_m$ (respectively, $(\mathbf{P}_{um})_m$) a sequence of lower (respectively upper) triangular matrices of $\mathbf{W}_{rt}$-type. For all $k < n$, if $\Pi_{l,k-1}$ is such that all its elements on line j and columns $j-1$ ($j \in \{3, 4, \dots, n\}$) through column $j-k+2$ ($j \in \{k, \dots, n\}$) are zeros, then $\Pi_{l,k}$ is such that all its elements on line j and columns $j-1$ ($j \in \{3, 4, \dots, n\}$) through column $j-k+1$ ($j \in \{k+1, \dots, n\}$) are also zeros.

Let p'_{ij} denote the terms of the row-stochastic matrix $\Pi_{l,k-1} = \prod_{m=1}^{k-1} \mathbf{P}_{l\ k-m}$ which are non null and different from 1, and $p"_{ij}$ the terms of $\Pi_{l,k}$.

$$\begin{aligned} \Pi_{l,k} =& \mathbf{P}_{lk}\Pi_{l,k-1} \\ =& \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ 1 & 0 & \dots & \dots & 0 \\ p_{31,lk} & p_{32,lk} & 0 & \dots & \vdots \\ \vdots & \dots & \dots & \ddots & \vdots \\ p_{n1,lk} & \dots & \dots & p_{n\ n-1,lk} & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ \vdots & 0 & \dots & \dots & \dots & \vdots \\ 1 & 0 & \ddots & \ddots & \dots & \vdots \\ p'_{k+1\ 1} & p'_{k+1\ 2} & 0 & \ddots & \dots & \vdots \\ \vdots & \dots & \ddots & \ddots & \ddots & \vdots \\ p'_{n1} & \dots & p'_{n\ n-k} & 0 & \dots & 0 \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ \vdots & 0 & \dots & \dots & \dots & \vdots \\ 1 & 0 & \ddots & \ddots & \dots & \vdots \\ p"_{k+2\ 1} & p"_{k+2\ 2} & 0 & \ddots & \dots & \vdots \\ \vdots & \dots & \ddots & \ddots & \ddots & \vdots \\ p"_{n1} & \dots & p"_{n\ n-k-1} & 0 & \dots & 0 \end{bmatrix}$$

Consequently, for $k \geq n-1$, we have:

$$\Pi_{l,k} = \Pi_{l,n-1} = \begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ 1 & 0 & \dots & \dots & \dots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & 0 & \dots & \dots & \dots & 0 \end{bmatrix}$$

Using a similar mathematical induction, we can prove the result for upper triangular matrices of type $\mathbf{W}_{rt}$.

Proposition 3.3

Let $\mathcal{N}_0$ a connected network. Let us assume that $\mathcal{N}_0$ has both a negative ϵ-extreme pole $\mathcal{V}^-_{\epsilon_1,0} \neq \emptyset$ and a positive ϵ-extreme pole $\mathcal{V}^+_{\epsilon_2,0} \neq \emptyset$ at time 0, with $\epsilon_1,\ \epsilon_2 \in (0, \frac{1}{2})$. Let $i \in \mathcal{V}^-_{\epsilon_1,0},\ j \in \mathcal{V}^+_{\epsilon_2,0}$. We have $\mathcal{V}^-_{\epsilon_1,0} \cap \mathcal{V}^+_{\epsilon_2,0} = \emptyset$. Then:

$$\sum_{k \in \mathcal{V}^-_{\epsilon_1,0}} w^0_{ik} = 1 \Rightarrow w_{ij} = 0$$

and,

$$\sum_{k \in \mathcal{V}^+_{\epsilon_2,0}} w^0_{jk} = 1 \Rightarrow w_{ji} = 0$$

Which is absurd, because the network is connected.

Proposition 3.4

Consider a network in which beliefs converge to a consensus vector $b\boldsymbol{i}$. I want to show that b is ϵ-extreme $\iff$ there is a non-empty ϵ-extreme pole in the network. This is trivial.

If a non-empty ϵ-extreme pole exists in the network, then if beliefs have converged, it means they must have converged to an ϵ-extreme value within the set of values in the pole (i.e. $|b| \geq 1-\epsilon$). Otherwise, if beliefs have converged to a value $|b| < 1-\epsilon$, there is an absurdity given the fact that individuals' beliefs in an ϵ-extreme pole only listen to themselves.

Similarly, if beliefs converge to a value $|b| \geq 1-\epsilon$, that value is ϵ-extreme, and it means that there exists at least one individual whose belief *a priori* is greater or equal to b in absolute value. Then, it follows that a non-empty ϵ-extreme pole exists.

C.1.3 Miscellaneous proofs

Proof that W_t must be row-stochastic for the existence of a consensus

Assume a consensus exist and is given by the vector $b.\boldsymbol{i}$. Then, using equation 3.1, there exist $\bar{t} > 0$, such that $\forall\ t \geq \bar{t}$:

$$\begin{aligned} b.\boldsymbol{i} = b\mathbf{W}_t.\boldsymbol{i} \Rightarrow b &= b\sum_{j=1}^{n} w_{ij}^{t} \\ &\Rightarrow b\left(1 - \sum_{j=1}^{n} w_{ij}^{t}\right) = 0 \end{aligned}$$

From equation 3.4, it follows that, $\forall\ b \neq 0$:

$$\sum_{j=1}^{n} w_{ij}^{t} = 1 \iff \sum_{j=1}^{n} w_{ij}^{0} = 1$$

Therefore, if $\mathbf{W}_0$ is not row-stochastic, consensus never exist, unless at 0. [3]

3. Logic: $(p \Rightarrow q) \iff (\neg q \Rightarrow \neg p)$

C.1.4 On paracontracting matrices

Relatively to the network structure, the literature in linear algebra have already documented some properties for convergence and consensus in models similar to 3.2. Usually, the criteria of ergodicity of the matrices sequence $(\prod_{m=0}^{t} \mathbf{W}_{t-m})_t$ is the most redundant, and is necessary for consensus. However, because of the complex form of the updating process 3.4 of the matrices $\mathbf{W}_t$, it is not possible to find direct analytic conditions that guaranty consensus in beliefs. Yet, other concepts like the concept of *paracontracting matrices* could also be used to infer consensus. In particular, Nelson and Neumann (1987) introduce *paracontracting matrices* as follows:

Definition C.2. *(Nelson and Neumann, 1987):*
Let's consider the set $\mathbb{C}^{n\times n}$ of complex matrices, $\|.\|$ some vector norm, and $\boldsymbol{x} \in \mathbb{C}^n$ a vector. A matrix $\boldsymbol{M} \in \mathbb{C}^{n\times n}$ is called ***paracontracting*** *with respect to the vector norm $\|.\|$, if the ftimeollowing equivalence is true:*

$$\boldsymbol{Mx} \neq \boldsymbol{x} \Leftrightarrow \|\boldsymbol{Mx}\| < \|\boldsymbol{x}\| \tag{C.1}$$

Let consider a sequence of listening matrices and beliefs $(\mathbf{W}_t, \mathbf{b}_t)_t$ defined by 3.2

Proposition C.2. *Proof omitted*

(i) If each matrix $\boldsymbol{W}_t$ of the sequence is paracontracting, then beliefs converge to a consensus.

(ii) If each matrix $\boldsymbol{W}_t$ is not row-stochastic, then beliefs never converge to a consensus.

C.2 Additional figures and graphs

Graphs and plots related to proposition 3.1

Weakly connected *a priori* network

Consider the following *a priori* listening matrix and its graph :

$$\mathbf{W}_{0w} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{9}{10} & \frac{1}{10} & 0 \\ 0 & \frac{1}{5} & \frac{4}{5} \end{bmatrix}$$

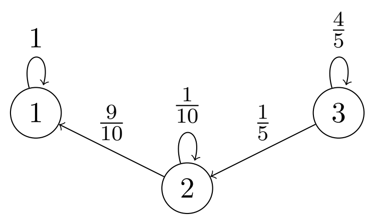
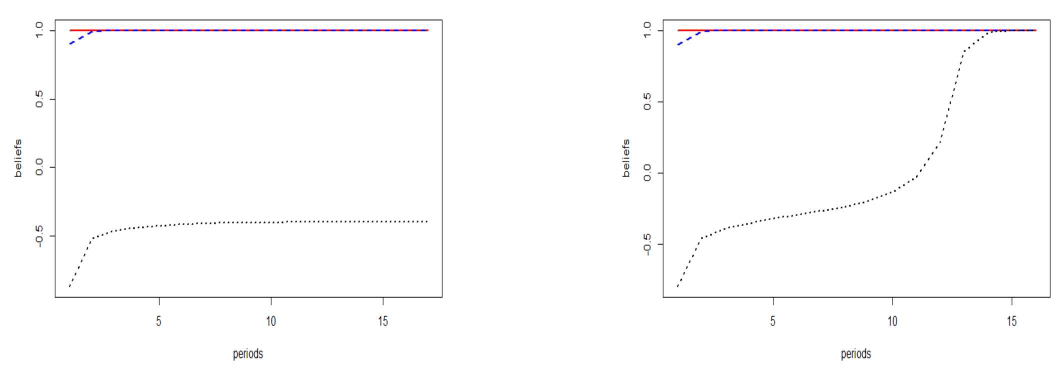


Figure C.1 – Network graph for listening matrix $\mathbf{W}_{0w}$

This network is obviously weakly connected. Using this listening matrix, the following figure shows the paths towards convergence for two different *a priori* beliefs vectors. In all future figures a 3-individuals network, the full line represents individual 1's beliefs, the dashed line individual 2 and the dotted line individual 3.



(a) $b_0 = (1, \frac{9}{10}, -\frac{13}{15}), \mathbf{b}_T = (1, 1, -.39)$, No consensus

(b) $b_0 = (1, \frac{9}{10}, -\frac{4}{5})$, $T = 26$, $\mathbf{b}_T = (1, 1, 1)$

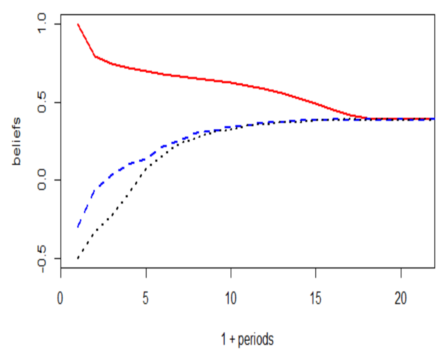
Figure C.2 – Examples of convergence for weakly connected initial listening matrix $\mathbf{W}_{0w}$

Strongly connected *a priori* network

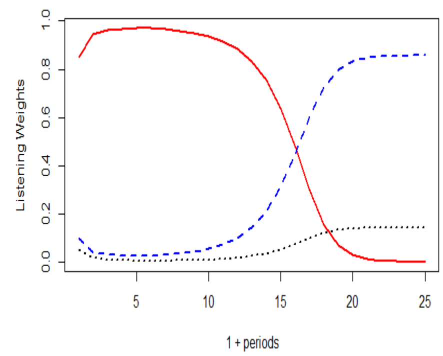
Let now consider the following *a priori* listening matrix and its graph :

$$\mathbf{W}_{0s} = \begin{bmatrix} \frac{17}{20} & \frac{1}{10} & \frac{1}{20} \\ \frac{1}{5} & \frac{7}{10} & \frac{1}{10} \\ \frac{1}{10} & \frac{1}{10} & \frac{4}{5} \end{bmatrix}$$

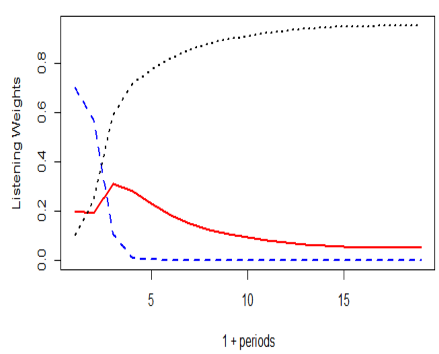
The following figure shows respectively the beliefs paths towards consensus for the *a priori* network $\mathcal{N}_{0,s}$, using the *a priori* beliefs vector $\mathbf{b}_0 = (1, -\frac{3}{10}, -\frac{1}{2})'$, and the convergence paths for individuals listening weights.



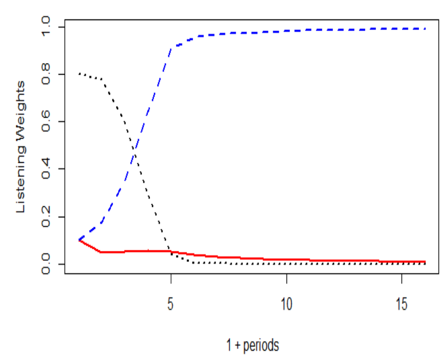
(a) $\mathbf{b}_0 = (1, -\frac{3}{10}, -\frac{1}{2})', \bar{\mathbf{b}} = (.39, .39, .39)'$



(b) $\mathbf{w}^1_{0,s} = (\frac{17}{20}, \frac{1}{10}, \frac{1}{20}), \bar{\mathbf{w}}^1_s = (0, .84, .16)$

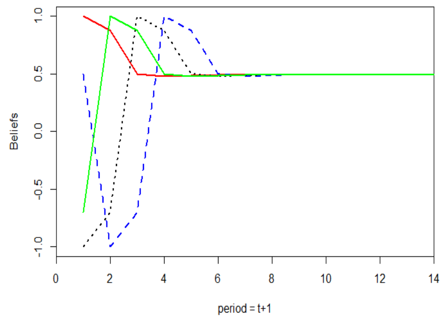


(c) $\mathbf{w}^2_{0,s} = (\frac{1}{5}, \frac{7}{10}, \frac{1}{10}), \bar{\mathbf{w}}^2_s = (.05, 0, .95)$

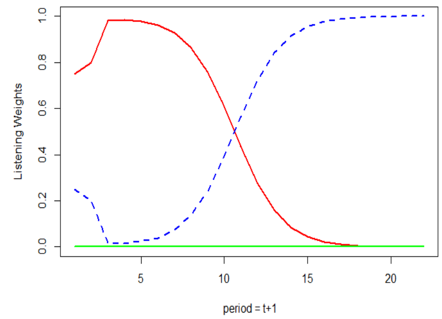


(d) $\mathbf{w}^3_{0,s} = (\frac{1}{10}, \frac{1}{10}, \frac{4}{5}), \bar{\mathbf{w}}^3_s = (.01, .99, 0)$

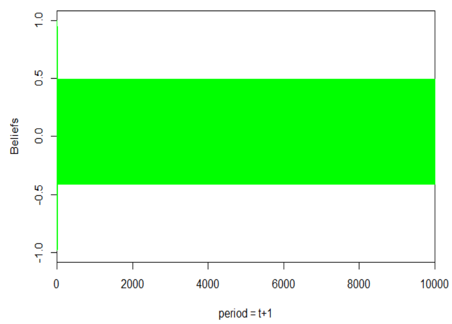
Figure C.3 – Examples of convergence for strongly connected initial listening matrix $\mathbf{W}_{0s}$



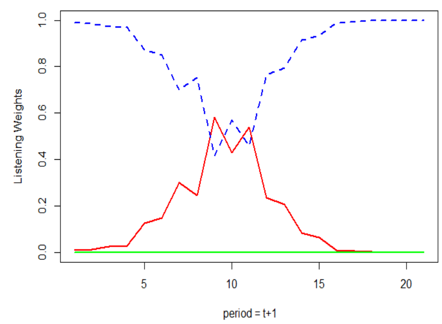
(a) $w^1_{0,aper} = (\frac{3}{4}, \frac{1}{4}, 0, 0), b \to .49$



(b) Individual 1, $\mathbf{w}^1_{0,aper} = (\frac{3}{4}, \frac{1}{4}, 0, 0)$, $\bar{\mathbf{w}}^1_{t,aper} \to (0, 1, 0, 0)$



(c) $w^1_{0,aper} = (\frac{1}{100}, \frac{99}{100}, 0, 0)$, no convergence



(d) Individual 1, $\mathbf{w}^1_{0,aper} = (\frac{1}{100}, \frac{99}{100}, 0, 0)$, $\bar{\mathbf{w}}^1_{t,aper} \to (0, 1, 0, 0)$

Figure C.4 – Examples of beliefs and listening weights paths using *a priori* beliefs vector $\mathbf{b}_0 = (1, \frac{1}{2}, -1, -\frac{7}{10})'$ and aperiodic network structure $\mathcal{N}_{aper}$

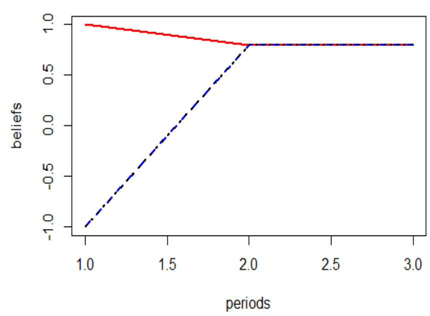
A strongly connected network aperiodic "enough": $\mathcal{N}_{aper}$, $\mathbf{b}_0 = (1, \frac{1}{1}, -1, -\frac{7}{10})'$

An ergodic *a priori* listening matrix

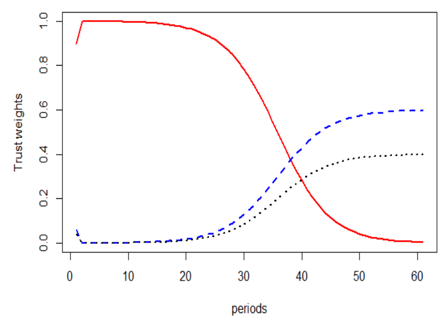
Let $\mathbf{b}_0 = (1, -1, -1)'$ a vector of *a priori* beliefs and

$$\mathbf{W}_0 = \begin{bmatrix} \frac{9}{10} & \frac{3}{50} & \frac{1}{25} \\ \frac{9}{10} & \frac{3}{50} & \frac{1}{25} \\ \frac{9}{10} & \frac{3}{50} & \frac{1}{25} \end{bmatrix}$$

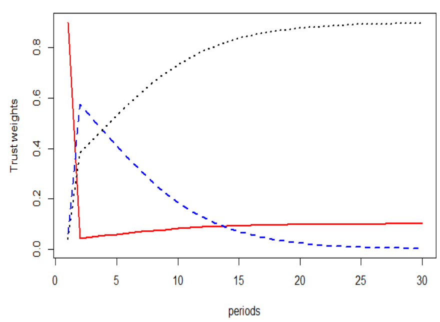
Here are the related convergence paths of individuals beliefs and listening weights.



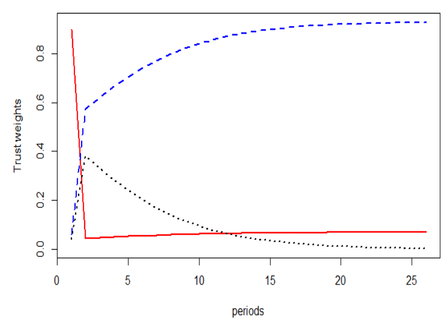
(a) $\mathbf{b}_0 = (1, -1, -1)'$, $T = 1, \mathbf{b}_T = (.8, .8, .8)'$



(b) Individual 1, $\mathbf{w}_{10} = (\frac{9}{10}, \frac{3}{50}, \frac{1}{25})$, $T = 207$, $\mathbf{w}_{1T} = (0, \frac{3}{5}, \frac{2}{5})$

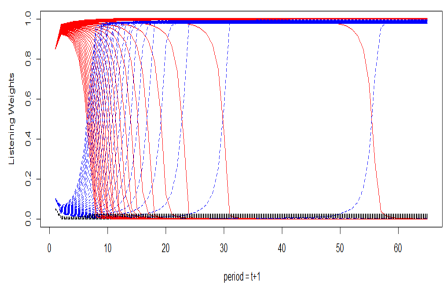


(c) Individual 2, $\mathbf{w}_{20} = (\frac{9}{10}, \frac{3}{50}, \frac{1}{25})$, $T = 190$, $\mathbf{w}_{2T} = (\frac{1}{10}, 0, \frac{9}{10})$

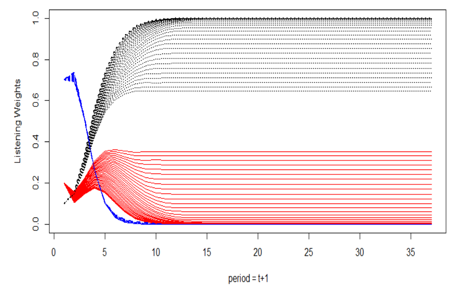


(d) Individual 3, $\mathbf{w}_{30} = (\frac{9}{10}, \frac{3}{50}, \frac{1}{25})$, $T = 183$, $\mathbf{w}_{3T} = (\frac{7}{100}, \frac{93}{100}, 0)$

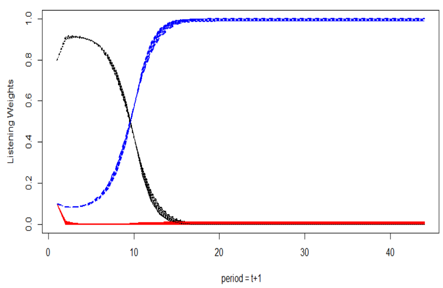
Figure C.5 – Consensus with an ergodic *a priori* listening matrix



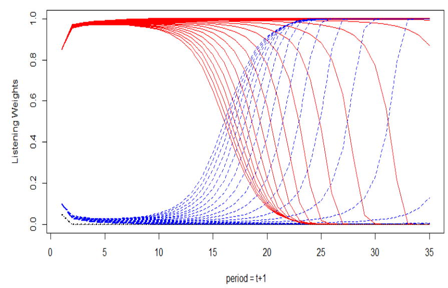
(a) Listening weights for **individual 1** - Pronounced ERS bias: $b1^0 \in [\frac{7}{10}, 1]$, $b_2^0 = -\frac{1}{2}$, $b_3^0 = -1$



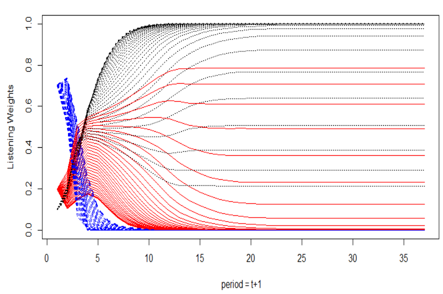
(b) Listening weights for **individual 2** - Pronounced ERS bias: $b1^0 \in [\frac{7}{10}, 1]$, $b_2^0 = -\frac{1}{2}$, $b_3^0 = -1$



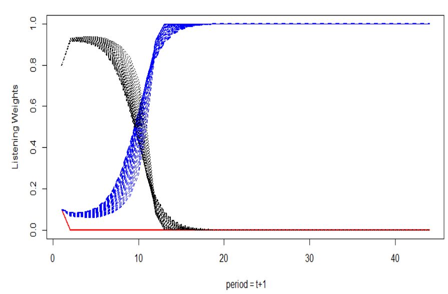
(c) Listening weights for **individual 3** - Pronounced ERS bias: $b1^0 \in [\frac{7}{10}, 1]$, $b_2^0 = -\frac{1}{2}$, $b_3^0 = -1$



(d) Listening weights for **individual 1** - Pronounced Confirmation bias: $b1^0 = 1$, $b_2^0 \in [-\frac{1}{2}, -\frac{1}{5}]$, $b_3^0 = -1$



(e) Listening weights for **individual 2** - Pronounced Confirmation bias: $b1^0 = 1$, $b_2^0 \in [-\frac{1}{2}, -\frac{1}{5}]$, $b_3^0 = -1$



(f) Listening weights for **individual 3** - Pronounced Confirmation bias: $b1^0 = 1$, $b_2^0 \in [-\frac{1}{2}, -\frac{1}{5}]$, $b_3^0 = -1$

Figure C.6 – Effect of the ERS and confirmation biases on convergence rates and consensus: Listening weights

C.3 Analysis of simulated data

Probability of consensus in groups of size $n = 3$

	β		$z-$value
$\log(\lambda_2(\mathbf{W}_0))$	-2.32	$***$	-39.55
	(.06)		
σ^2	-6.59	$***$	-29.52
	(.22)		
$\|\kappa\|$	0.77	$***$	12.47
	(.06)		
α	4.74	$***$	30.21
	(.16)		

Standard errors in brackets (). $***$ Significant at 1‰.

AIC criterion: 7686

Null deviance: 11466 on 99999 DF.

Residual deviance: 7678 on 99996 DF

Table C.1 – *Probit* model for the probability of consensus

	$\lvert\kappa\rvert = 1$					$\lvert\kappa\rvert = .5$					$\kappa = 0$				
Variance $\sigma^2 =$	.5	.6	.7	.8	.9	.5	.6	.7	.8	.9	.5	.6	.7	.8	.9
$\lambda_2(\mathbf{W}_0) = .01$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\lambda_2(\mathbf{W}_0) = .05$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\lambda_2(\mathbf{W}_0) = .1$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\lambda_2(\mathbf{W}_0) = .25$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.98
$\lambda_2(\mathbf{W}_0) = .5$	1.00	1.00	0.99	0.97	0.88	1.00	1.00	0.98	0.93	0.79	1.00	0.99	0.96	0.86	0.66
$\lambda_2(\mathbf{W}_0) = .75$	1.00	0.99	0.94	0.82	0.60	0.99	0.97	0.88	0.70	**0.44**	0.98	0.93	0.79	0.55	**0.30**
$\lambda_2(\mathbf{W}_0) = .9$	0.99	0.96	0.87	0.68	**0.43**	0.98	0.92	0.77	0.54	**0.29**	0.95	0.85	0.64	**0.39**	**0.17**
$\lambda_2(\mathbf{W}_0) = .95$	0.99	0.95	0.84	0.64	**0.38**	0.97	0.90	0.73	**0.49**	**0.25**	0.94	0.82	0.60	**0.34**	**0.14**

Table C.2 – Predicted probabilities of consensus by *a priori* network's characteristics: $\lambda_2(\mathbf{W}_0)$, σ^2 and κ