

Strategic Interaction and Socio-Economic Structure

Dissertation zur Erlangung des
Doktorgrades der Wirtschaftswissenschaften (Dr. rer. pol.)
an der Fakultät für Wirtschaftswissenschaften der Universität Bielefeld

vorgelegt von

Herrn Florian Gauer, M.Sc.

Februar 2016

Gutachter

1. Gutachter: Prof. Christoph Kuzmics, Ph. D.
2. Gutachter: Jr.-Prof. Tim Hellmann

Für meine Eltern und Jenny

Acknowledgments

It is a matter of great importance to me to thank some people who helped me with my research.

First of all, I would like to express my deepest gratitude to my supervisors, Christoph Kuzmics and Tim Hellmann, for their great support. In countless discussions about my research ideas and problems at any stage, I have profited enormously from their expertise. Christoph's thirst for research, his openness, friendliness and motivating nature were truly inspiring and it was both an honor and a pleasure to co-author a paper with him. Tim's enthusiasm for network theory was absolutely contagious and without his guidance this thesis would not have been possible either.

Further, I am very grateful to Herbert Dawid and Frank Riedel for their advice and support I could always count on. Also, it is thanks to them that I had the chance to present my work at several conferences, workshops and seminars which was of great benefit for this thesis. Moreover, I acknowledge the generous financial support of the German Research Foundation (DFG) within the doctoral program Economic Behavior and Interaction Models (EBIM) as well as the grant I received from the Bielefeld Young Researchers' Fund.

Having had three and a half amazing years at the Center for Mathematical Economics (IMW) at Bielefeld University, I am also indebted to all other members of the center, the Faculty of Economics, and the Bielefeld Graduate School of Economics and Management (BiGSEM) for incorporating me into a pleasant environment for academic work. The comments, suggestions and ideas in various seminars have significantly helped me to develop my projects (and to work harder), but also all the social events have made this time memorable for me.

In this regard, I want to particularly mention my dear colleagues and friends Michael Günther and Jakob Landwehr. I cannot imagine anyone better to share an office or to pass a six-month research stay in Paris with. Frequently discussing each other's work or even writing a paper together was very enjoyable and stimulating. It has been a pleasure. Likewise, I would like to thank Oliver Claas for his permanent helpfulness which is only excelled by his boundless knowledge.

Beyond that, I am obliged to Mihai Manea, Dunia López-Pintado, Agnieszka Rusinowska, Emily Tanimura, and Fernando Vega-Redondo for providing their valuable feedback on various occasions. Similarly, this applies for participants of seminars and conferences in Paris, Graz, Cambridge and Münster.

Finally, this thesis would not have been possible without the support of my family. In particular, I owe sincere and earnest thankfulness to my parents for all they have done for me and to my (future) wife for putting up with me all these years. You are my backbone and my driving force.

Contents

1	Introduction	1
1.1	Scientific Context	1
1.2	Contributions	6
2	Cognitive Empathy in Conflict Situations	11
2.1	Introduction	11
2.2	The Model	17
2.3	A Non-Conflict Example	19
2.4	Equilibrium Empathy Acquisition	20
2.5	Discussion and Conclusion	29
2.5.1	Empathy Acquisition at Zero Costs	29
2.5.2	Equilibrium Payoffs	30
2.5.3	The Timing of Decisions	34
2.5.4	Degrees of Cognitive Empathy	35
	Appendix 2.A Proofs	36
	Appendix 2.B A Continuum of Bayesian Nash Equilibria	51
	Appendix 2.C On the Uniqueness of Completely Mixed Nash Equilibria	53
3	Strategic Formation of Homogeneous Bargaining Networks	57
3.1	Introduction	57
3.2	The Model	61
3.3	Characterization of Stable Networks	68
3.3.1	Stability and Bargaining Outcomes	78
3.3.2	Singular Pairwise Stability	79
3.4	Efficiency	81
3.5	Effects of Time Discount	84
3.6	Conclusion	87
	Appendix 3.A Proofs	88

Appendix 3.B Relation to the Works of Hellmann (2013) and Hellmann and Landwehr (2014)	111
Appendix 3.C Relation to the Work of O'Donnell (2011)	115
4 Continuous Homophily and Clustering in Random Networks	123
4.1 Introduction	123
4.2 The Model	127
4.3 Basic Properties of Homophilous Random Networks	129
4.4 Clustering	133
4.5 The Small-World Phenomenon	136
4.6 An Example of the Labor Market	141
4.7 Conclusion	144
Appendix 4.A Proofs	145
References	157

Chapter 1

Introduction

This doctoral thesis in economic theory studies the interaction between socio-economic agents. On the one hand, it focuses on the fundamentals of strategic interaction, on the other it analyzes the induced socio-economic structures, in particular the formation of networks.

When it comes to strategic interaction in economic settings, several questions arise immediately. Why and how do agents interact? Which strategies will they pursue? To what extent do agents want to acquire additional information if there is uncertainty? How does the behavior of counterparts influence rational agents' (re)actions? Is it possible that no one has incentives to deviate from a given behavior? And if so, what are the characteristics of such a situation that is either called stable or an equilibrium? Beyond these questions concerning strategic interaction, we can also ask what these interactions imply. Which kinds of networks of agents typically emerge in certain settings? Which properties do they have? What effects does the formation of these networks have on socio-economic outcomes? To what extent do theoretic results fit real-world data?

These are major questions which this doctoral thesis considers in detail within the framework of three different models being developed and analyzed. Thus, it contributes to a variety of research fields in economics but also in sociology and mathematics.

1.1 Scientific Context

Thinking about strategic interaction, one is at the very heart of game theory. “The study of mathematical models of conflict and cooperation between intelligent rational

decision-makers”, as Myerson (1991, p. 1) describes it, is central to economic sciences nowadays. The important concepts of game theory find application wherever a certain outcome does not only depend on one’s own decision but also on the behavior of others. Common economic examples are decisions about prices or quantities by competing firms, the choice of technological standards or the behavior of competitors in auctions, to mention only a few.

First recorded steps towards the development of this field trace back to a discussion about the card game “Le Her” initiated by Charles Waldegrave in a letter he wrote to Pierre Rémond de Montmort in the early 18th century (see e.g. de Montmort, 1713; Bellhouse, 2007). However, it took until the 20th century for the mathematical theory of games to get established as a unique field by von Neumann (1928) (see also von Neumann and Morgenstern, 1944). A natural way to think about these situations of conflict or cooperation is to seek for a status quo where each agent’s strategy is individually optimal such that no one wants to deviate from it. This leads to the solution concept of “Nash equilibria” which was invented by Nash (1950b, 1951) and refined by Selten (1965) through his work on “subgame perfect equilibria” in a dynamic context.¹

Until then, in the game-theoretic literature, it was assumed that agents involved in a situation of strategic interaction are always perfectly informed about the state of the world. In sequential games, this means that each agent always knows all developments and decisions made at previous stages. In many real-world examples, however, this is very rarely the case. It was Harsanyi (1967, 1968a,b) who developed the concepts of incomplete information and Bayesian games giving rise to the solution concept of “Bayesian Nash equilibrium”.² Hereby, Harsanyi also provided a theoretical foundation for the economics of information.

During that time, the concepts of game theory also found their way into other disciplines such as biology. Since then and initiated by the paper of Maynard Smith and Price (1973), much research has been devoted to evolutionary game theory used as a tool to analyze Darwinian competition. Some of the results derived in this context are now used by the more recent literature on the evolution of preferences in more concrete economic settings. Güth and Yaari (1992) and Güth (1995) were the

¹From 1972 until 1984, Reinhard Selten worked at the Center for Mathematical Economics at Bielefeld University.

²For their contributions to economic game theory, John Nash, John Harsanyi and Reinhard Selten were awarded the Nobel Memorial Prize in Economic Sciences in 1994. With Robert Aumann and Thomas Schelling, two further game theorists received the prize in 2005. Two years later, Leonid Hurwicz, Eric Maskin and Roger Myerson received the prize for having laid the foundations of mechanism design theory, which is closely related to game theory.

first to work on this.

In general, both the literature on Bayesian games and on the evolution of preferences posit an exogenously given structure of information. However, in settings of strategic behavior where an agent's information typically has an impact on her payoffs, it seems appropriate and enlightening to consider individual information acquisition as an endogenous decision variable. Yet, not much work has been done in this direction. Chapter 2 of this doctoral thesis takes up this idea and contributes to a better understanding of endogenized acquisition of information or, as we call it, "cognitive empathy" in conflict situations.

Beyond that, another important application of game theory in general is the one to bargaining problems. In such problems, typically two or more agents try to find an agreement regarding how to distribute an object or a monetary amount. For example, this is the case when the personnel director of a company negotiates wages with workers' unions, firms bargain with other firms over prices or collaborations, or politicians over environmental or trade agreements. In such situations, each agent usually seeks for an outcome which is as favorable as possible for herself, however, without threatening an agreement. A first notable idea to tackle this problem in economics was the axiomatic, cooperative approach by Nash (1950a) (see also Nash, 1953, for a non-cooperative approach). Given agents' disagreement points, feasible utility values and bargaining power, this provides what we today call the "(generalized) Nash bargaining solution". In reality, however, agents might use elaborate bargaining tactics and, moreover, there is no certainty that an agreement will be reached. While the above approaches abstract away from this, these open points were already addressed to some extent by the work of Schelling (1956). From today's perspective, however, the strategic approach introduced by Rubinstein (1982) (see also Rubinstein and Wolinsky, 1985) was probably the most important contribution in this regard. It proposes a reasonable dynamic specification of the bargaining game and provides a unique, subgame perfect equilibrium as a solution. While most other approaches are even completely silent on the origin of agents' bargaining power, here, it is simply determined by her level of patience.

In general, however, bargaining power can be influenced by many different factors. In a context where bilateral agreements can be reached with different bargaining partners, one's bargaining power should heavily depend on the number and kinds of alternatives among which one can choose. Such a structure of bilateral links between agents can be interpreted as a network. As an example for this, one could consider a setting of project collaborations between companies. Given such a bargaining

network, a link between two agents would then represent a potential collaboration between the two. Capturing this idea, Manea (2011) sets up and analyzes a model which can be regarded as a microfoundation of the above seminal papers. As a main result, he establishes that all subgame perfect equilibria of the network bargaining game are payoff equivalent. Here, networks are assumed to be exogenously given.³ In a setting of strategic interaction where agents' expected payoffs are determined by the structure of their network, however, it seems reasonable to assume that each agent would strive to maximize her anticipated profit by optimizing her network position. This is where Chapter 3 on strategic formation of homogeneous bargaining networks comes into play so as to examine such network structures with regard to stability and efficiency.

Network theory and, in particular, strategic network formation is a relatively young research field in economics which is, however, not restricted to bargaining frameworks by far. To name just a few, modeling trade and exchange of goods in non-centralized markets (see e.g. Goyal and Joshi, 2006), firms involved in R&D networks (see e.g. Goyal and Joshi, 2003) or personal contacts in the context of job search (see e.g. Calvó-Armengol, 2004) are further insightful applications. Early works of Boorman (1975), Aumann and Myerson (1988) and Myerson (1991) mark the beginning of the economic literature on strategic network formation. However, only after Jackson and Wolinsky (1996) introduced the seminal equilibrium concept of “pairwise stability”, this field became a most active research area.⁴ In models of this literature, networks are assumed to induce explicit benefits and costs for each contained agent. Usually, this then gives rise to individual incentives to add or delete links unless the network is stable.⁵ Furthermore, considering the collective of all agents in the network provides a measure of (utilitarian) welfare. Based on this, analyzing the tension between stable networks and “(strongly) efficient” ones, that is networks being optimal from a society's perspective, often yields further interesting insights (see again Jackson and Wolinsky, 1996).⁶

Prior to this, social networks have already been an object of research in other fields and disciplines such as labor economics and sociology. However, research ques-

³At least, this applies for the actual paper. For details concerning Manea (2011, Online Appendix) see Chapter 3.

⁴Jackson (2005), Jackson (2008b, Chapter 6) and Goyal (2012, Chapter 7) provide a nice overview of the literature and basic concepts.

⁵Apart from pairwise stability, several variations, refinements and alternative notions of stability such as, for example, “Nash stability”, “pairwise Nash stability” and “pairwise stability with transfers” have been developed. For an overview see Bloch and Jackson (2006).

⁶Note that an efficient network always exists whereas this is not guaranteed for pairwise stable ones.

tions differ substantially from the strategic approach considered above. Here, the focus is rather on features of given networks and socio-economic implications thereof. Moreover, in a large part of this literature, it is either not feasible or considered unnecessary to take explicit representations of whole networks in a graph-theoretic sense as a basis (see above and below). Abstracting from this notion, sociological studies have shown, for example, that social contacts and interactions play an important role in finding jobs and filling vacancies (see e.g. Rees, 1966; Granovetter, 1973, 1995). Starting with Montgomery (1991), labor economic models then try to explain why this is the case and what consequences it has for earnings, abilities of employees, firms' profits, unemployment, etc. Beyond that, certain stylized facts about real-world networks are well-established: They typically exhibit "homophily", that is the tendency of individuals to connect to similar others, "clustering" and the "small-world phenomenon" (see e.g. Lazarsfeld and Merton, 1954; Milgram, 1967; McPherson et al., 2001).⁷ These properties have also been incorporated into some economic papers (see e.g. Currarini et al., 2009). However, in these models, homophily is usually captured by a binary or discrete measure, thus rather in terms of equality than similarity. Further simplifications are often due to the abstract notion of networks mentioned above. One should be very careful about such simplifications as it is probably not very rare that "the structure of the social network then turns out to be a key determinant" (p. 12 Jackson, 2005).

Another discipline in which networks play an important role is the mathematical field of graph theory. Considering this literature leads us back to an explicit notion of graphs (synonymously for networks) consisting of vertices and edges. There also exists a strand of literature here which investigates network formation. However, in contrast to the strategic approach mentioned above, it considers networks which form at random. This means that the formation of links does not result from individual incentives and strategic interaction but is assumed to follow some probabilistic rule.⁸ On the one hand, such random network models serve as an approximating tool to examine and handle real-world networks which are usually quite large and remain unknown for an analysis. On the other hand, random networks can be used, for example, to understand and predict diffusion processes in societies. This might be

⁷More precisely, a network is called homophilous if for any two individuals the likelihood to be linked is the higher the more similar they are in terms of one or several characteristics. A network is said to exhibit clustering if two individuals with a common neighbor have an increased probability of being linked. Finally, the small-world phenomenon describes the observation that even in large networks on average there exist relatively short paths between two individuals.

⁸For a general introduction into graph theory see Bollobás (1998) and West (2001). Moreover, see Bollobás (2001) and Jackson (2008b, Chapter 4) as well as Jackson (2006, Section 3.1) to get an overview of random graphs respectively networks.

of importance if one wants to estimate how information or a disease will spread. To provide meaningful results, such a model should be designed as realistically as possible, that is in a way to ensure it complies with the stylized facts known from sociology (see above).⁹ One that is frequently considered until today and which was already examined in the seminal paper of Erdős and Rényi (1959) is the “Bernoulli Random Graph model” in which links are formed uniformly at random. This popular model exhibits the small-world phenomenon but fails on homophily and clustering (see e.g. Bollobás, 2001). Chapter 4 of this doctoral thesis addresses this issue and proposes a tractable random network model which can be seen as a generalization of the Bernoulli Random Graph model exhibiting all of the stylized facts mentioned above.

1.2 Contributions

In Chapter 2, we build a simple model of strategic interaction with two players having the option to acquire information about their respective opponent’s preferences which are *ex ante* uncertain. We show that, for sufficiently small positive costs of information acquisition, in any Bayesian Nash equilibrium of the resulting conflict game of incomplete information the probability of getting informed about the opponent’s preferences is bounded away from zero and one. For the evolutionary population interpretation of the game this result implies that we would expect that there are people who are “cognitively empathic”, i.e. who know their opponents’ preferences, and that there are others who are not. Even if the cost of empathy acquisition is zero, besides a full empathy equilibrium, the game still has such an equilibrium with mixing between acquiring empathy and not acquiring it. Moreover, we show that for small costs there is always an equilibrium in which the lower bound on the probability of empathy acquisition is achieved for both players. Finally, we establish that in certain cases the partial empathy equilibrium is the only equilibrium.

In the model, each of the two players can be one of a finite number of different preference types. The distribution over all preference types is commonly known (to avoid that our results confound with higher-order belief effects). Both players, before learning their own types (this is for convenience), simultaneously decide whether or not to pay a small amount of cost in order to acquire empathy, that is to learn the opponent’s type. Anyway, players do not observe their opponent’s choice of empathy

⁹For formal definitions of these stylized facts in a random network setting see Jackson (2010, Section 3.3).

acquisition. After learning their own and, if appropriate, their opponent's type, players then choose their (possibly mixed) action as a function of what they know. For the main results, we investigate “two-action Bayesian conflict games”. Here, both players have two actions available and, if we assume players' types to be common knowledge, then any such complete information “realized type game” must have a unique Nash equilibrium and that Nash equilibrium must be in completely mixed strategies.

There are at least two different interpretations we can give for our model. One is that players are highly rational but have some small costs of reasoning about their respective opponent's preferences. Our model could then be about two individuals engaged in the penalty kick, two firms engaged in conflict or military generals engaged in war. In this context we talk about players “acquiring information” about their opponents. On the contrary, in its evolutionary interpretation, there is mother nature (or evolution) who works on everyone of her subjects independently and has their material interests at her heart. Nature knows that her subjects will be involved in all sorts of conflict situations throughout their life. She individually decides whether or not she should spend some small amount of fitness cost to endow her subjects with cognitive empathy, which would then allow the respective subject to always learn (in fact, to always know) the opponent's preferences. In this context we talk about “acquiring (cognitive) empathy”.

This chapter is joint work with Christoph Kuzmics. He initially contributed the general research idea and at the end contributed to the actual writing of the main parts of the paper. We worked together on finding and concretizing the actual final choice of model and on identifying which results we want to pursue. I provided the proofs for essentially all results and worked through all examples (thus, identifying possible results we can pursue). In the final stages of the project, I concentrated on the content and technical proofs and Christoph Kuzmics concentrated on the marketing of our paper. However, all this has been carried out in close consultation and double-checking each other's work.

In Chapter 3 (of which I am the sole author), we analyze a model of strategic network formation prior to a Manea (2011) bargaining game. Assuming patient players, we provide a complete characterization of non-singularly pairwise stable networks. More precisely, we show that specific unions of separated pairs, odd circles and isolated players constitute this class. As a byproduct, this implies that a pairwise stable network always exists, that is at each level of linking costs. We also show that many generic structures are not even singularly pairwise stable. As an important

implication, this reveals that the diversity of possible bargaining outcomes is substantially narrowed down provided pairwise stability. Moreover, we establish that for sufficiently high linking costs, the networks being efficient in terms of the utilitarian welfare criterion coincide with the pairwise stable ones. However, this does not hold if costs are low or at an intermediate level. As a robustness check, we finally study the case of time-discounting players as well.

Our model consists of two stages. First, players are assumed to form undirected bilateral links among each other which results in a network. These players are assumed to be *ex ante* homogeneous, meaning that they are equal apart from their potentially differing network positions. Further, we consider explicit linking costs which players have to bear for each link they form. In this context, one shall think of one-time initiation or communication costs. Benefits from linking are induced by the network in the second stage. Here, given the network that has formed in the first stage, players are supposed to play an infinite horizon network bargaining game à la Manea (2011). Thus, they sequentially bargain with neighbors for the division of a mutually generated unit surplus. According to Manea (2011) all subgame perfect equilibria of this bargaining game are payoff equivalent. Players anticipate these outcomes during the preceding stage of network formation and choose their actions accordingly. To state our results, we introduce a novel classification of pairwise stable networks. A network is said to be singularly (non-singularly) pairwise stable if it has this property at exactly (more than) one cost level. As only non-singularly pairwise stable networks can be robust with respect to slight changes of costs, we focus on this class throughout our analysis.

As a possible application of our model, we outline a setting of project collaboration. Here, players represent similar firms which can mutually generate an (additional) surplus within bilateral projects by exploiting synergy potentials. For instance, this possibility might be based on capacity constraints or cost-saving opportunities. In this context, one-time costs might arise to prepare each two firms for mutual projects (adjustment of IT, joint training for workers, etc.).

In Chapter 4, we propose and examine a random network model incorporating heterogeneity of agents and a continuous notion of homophily. As a main result, we show that for any positive level of homophily, our “Homophilous Random Network model” exhibits clustering. Moreover, simulations indicate that the small-world phenomenon is preserved even at high levels of homophily. Finally, we provide a possible application within a stylized labor market setting. We consider a firm which has to choose whether to hire a new employee via the social network or via the formal job

market and obtain a simple decision rule.

Our model is a two-stage random process. First, agents are assigned characteristics independently drawn from a continuous interval. Second, a random network realizes where linking probabilities are contingent on a homophily parameter and the pairwise distance between agents' characteristics. More precisely, the probability of linkage between two agents continuously decreases in the distance of their characteristics and the homophily parameter directly determines the strength of this effect. In the limit case of no homophily, we reproduce the Bernoulli Random Graph model. Insofar, our setting can be regarded as a generalization of this seminal model.

Our approach enables us to account for homophily in terms of similarity rather than equality of agents, capturing the original sociological definition instead of the stylized version that has been commonly used in the economic literature up to now. In this regard, observe that in reality people are in many respects neither completely "equal" nor completely "different". We therefore believe that a notion that provides an ordering of the "degree of similarity" with respect to which an agent orders his preference for connections can capture real-world effects more accurately. Besides, a major distinction of our approach compared to the literature is the sequential combination of two random processes, where agents' characteristics are considered as random variables that influence the random network formation. We thus account for the fact that in many applications in which the network remains unobserved, it seems unnatural to assume that individual characteristics, which in fact may depict attitudes, beliefs or abilities, are perfectly known.

This chapter is joint work with Jakob Landwehr. I contributed the original research idea which was closely related to the application of our model we provide for the labor market. However, we started discussing this idea at an early stage, changed the focus and jointly developed the model. During our six-month research stay at the University Paris 1 Panthéon-Sorbonne, we worked together on identifying the results we want to pursue, proved them and finally wrote the paper. Here, Jakob Landwehr's expertise in using MATLAB (2014) was of great benefit. As essentially all the work was carried out jointly, both authors contributed equally to this project.

Chapter 2

Cognitive Empathy in Conflict Situations

“If you know the enemy and know yourself, you need not fear the result of a hundred battles. If you know yourself but not the enemy, for every victory gained you will also suffer a defeat. If you know neither the enemy nor yourself, you will succumb in every battle.”

— Sun Tzu, *The Art of War*, approximately 500BC

2.1 Introduction

It is probably rare in a conflict situation that we know the exact cardinal preferences of our opponent.¹⁰ Consider, for instance, a penalty kick in soccer. This is as close as one can imagine to a pure conflict (i.e. zero-sum) situation. The kicker wants to score, the goalkeeper wants to prevent that. Now imagine that the goalkeeper incurred, earlier in the game, a slight injury, a bruising on her left side, which might induce her to have a slight additional preference of jumping to the right.

If we now assume that there is a distribution of such preferences, commonly known to both players, perhaps “centered” around the original zero-sum preference, then we get as a Nash equilibrium of the game (the unique one if we think of the original

¹⁰In the quote from Sun Tzu stated above, it is difficult to know what he meant with “knowing yourself” and “knowing your enemy”. The last sentence of the quote seems to suggest that it is in fact impossible that both warring generals know neither themselves nor their enemy, as presumably we cannot have that both “succumb” in the battle between them.

zero-sum game as the classic game of matching pennies) a “purified” version of the original Nash equilibrium (see Harsanyi, 1973).^{11,12}

We are here interested, however, in the possibility of the players, here especially the kicker, to possibly acquire, at some small cost, information about the opponent’s true preferences, here about the goalkeeper’s small injury. Alternatively, one can think of there being a small cost for the players to think about the opponent’s preferences. The latter interpretation leads us to the term “cognitive empathy” in our title, as defined in psychology as the process of understanding another person’s perspective (see e.g. Davis, 1983), which can be traced back to at least Köhler (1929), Piaget (1932), and Mead (1934).¹³ Building this possibility of empathy acquisition (or, respectively, information acquisition) into such a conflict game with incomplete information, we are then interested in the following questions. To which extent do players acquire empathy in equilibrium? In the context of the penalty kick, suppose the kicker is aware of the goalkeeper’s small injury. Does she reason through what consequences this fact has for the goalkeeper’s preferences and strategy? How does the possibility of empathy acquisition affect players’ action choices in the game? Finally, how do the answers to these questions depend on the value of the cost of empathy acquisition?

To answer these questions we build a simple model. There are two players and (for the main result) two actions for each player. Each player can be one of a finite number of different preference types. The distribution over all preference types is commonly known (to avoid confounding our results with higher-order belief effects). Both players, before learning their own types (this is for convenience), simultaneously decide whether or not to pay a small amount of cost $c \geq 0$ (simply subtracted from their payoffs) in order to learn the opponent’s type. Players do not observe their opponent’s choice of empathy acquisition. After learning their own and, if appropriate, their opponent’s type, players then choose, as a function of what they

¹¹In a “purified” equilibrium (almost) all types of players use a pure strategy, albeit different types use different pure strategies. Nevertheless all players face a mixed strategy because they do not know their opponent’s type.

¹²In other contexts, that of coordination games, uncertainty over the opponents’ preferences, provided it is severe enough (to include dominant strategy types), has led to the “global games” literature on bank runs, etc. (see e.g. Carlsson and Van Damme, 1993; Morris and Shin, 1998), and technically, to a refinement of even pure Nash equilibria of the original full information game. See also Weinstein and Yildiz (2007) on the possibility of getting almost any possible refinement depending on how the model is closed in terms of its higher-order belief assumptions.

¹³This is in contrast to “affective empathy” which is defined as a person’s emotional response to the emotional state of others (see again Davis, 1983) and the two are not necessarily related. Shamay-Tsoory et al. (2009) find that different areas of the human brain are responsible for “cognitive” and “affective” empathy. Rogers et al. (2007) find that people with Asperger syndrom lack “cognitive” but not “affective” empathy.

know, one of the two actions (or a mixed action). We investigate Bayesian Nash equilibria of this game.

We first provide, as a point of reference, an example of a non-conflict game, in which in equilibrium all players always acquire empathy as long as the corresponding costs are not too large. In this game, each player has three preference types, two dominant strategy preference types (one for each action) and one coordination preference type. It is easy to see that the coordination preference type clearly benefits from learning her opponent's type.

We then investigate two-player two-action Bayesian conflict games. These are such that if the types of players were common knowledge, then any such complete information “realized type game” must have a unique Nash equilibrium and that Nash equilibrium must be in completely mixed strategies. For such games we show that, for sufficiently low positive costs of empathy acquisition, the probability of empathy acquisition is strictly bounded away from zero and one in any Bayesian Nash equilibrium of this game (Theorem 2.1).¹⁴ These bounds do not depend on the cost of empathy acquisition beyond the requirement that this cost is sufficiently small. In other words, in any equilibrium of this game, players randomize strictly between acquiring empathy and not acquiring it. It turns out that even if the cost is zero, the game, besides a “full empathy equilibrium” (Proposition 2.1), still has such an equilibrium with mixing between acquiring empathy and not acquiring it. Beyond that, we show that there is, for small costs, always an equilibrium in which the lower bound on the probability of empathy acquisition is achieved for both players (Proposition 2.2). This equilibrium is referred to as the “partial empathy equilibrium”. Finally, we establish that for two-action Bayesian conflict games with either two types for both players or a single type for one player this partial empathy equilibrium is the only equilibrium if costs are sufficiently small but positive (Proposition 2.3).

There are at least two different interpretations we can give for our model. One, along the lines as suggested above, is such that players are highly rational but have some small costs of reasoning about their respective opponent's preferences. This model could then be about the two individuals engaged in the penalty kick, but could also be about firms engaging in conflict or indeed, as in the quote by Sun Tzu above, military generals engaged in war. In this context we talk about players “acquiring information” about their opponents.

We prefer to think of this model, however, in its evolutionary interpretation. That is there is mother nature (or evolution) who works on everyone of her subjects

¹⁴In fact, for a player's equilibrium probability of empathy acquisition to be strictly greater than zero, her opponent must have (at least two) distinct payoff types.

independently and has their material interests at her heart. Nature knows that her subjects will be involved in all sorts of conflict situations throughout their life. She individually decides whether or not she should spend some small amount of fitness cost to endow her subjects with cognitive empathy, which would allow the respective subject to always learn (in fact, to always know) the opponent's preferences. In this context we talk about "acquiring (cognitive) empathy". For convenience and to avoid confusion, this is the phrasing which we mainly use throughout the chapter.

Under the latter interpretation, our results imply that, in general, nature (who is assumed to guide play to a Bayesian Nash equilibrium) endows some but not all of her subjects with cognitive empathy even if the costs of doing so are essentially zero.

Various strands of literature have motivated us to write down and study the model we analyze in this chapter.

One strand is the literature on the evolution of preferences for strategic interaction, initiated by the now sometimes called "indirect evolutionary approach" of Güth and Yaari (1992) and Güth (1995). Individuals who are randomly matched to engage in some form of strategic interaction (some game) are first given a preference (or utility function) by mother nature. Mother nature works on every player separately and does this with the view in mind to maximize this player's material preferences (number of offspring or fitness). Players then evaluate outcomes of play given these preferences given to them by mother nature. There are two kinds of results in this literature. Assuming that individuals (automatically) observe their opponents' preferences, in many settings non-material preferences arise as mother nature's optimal choice (see e.g. Koçkesen et al., 2000a,b; Heifetz et al., 2007a,b; Dekel et al., 2007; Herold and Kuzmics, 2009). On the other hand, assuming that individuals cannot observe their opponents' preferences essentially only allows material preferences as mother nature's optimal choice (see e.g. Ely and Yilankaya, 2001; Ok and Vega-Redondo, 2001). This induced Robson and Samuelson (2010) to wish that the potential observability of preferences is also subject to evolutionary forces.¹⁵ Some work in that direction has recently been begun by Heller and Mohlin (2015a,b).¹⁶ Our model can be seen as to tackle the question of the evolution of observability of preferences without evolution.

¹⁵Similarly, Samuelson (2001, p. 228) states "Together, these papers highlight the dependence of indirect evolutionary models on observable preferences, posing a challenge to the indirect evolutionary approach that can be met only by allowing the question of preference observability to be endogenously determined within the model."

¹⁶The former is a model in which, while individual preferences evolve, so do individuals' abilities to deceive their opponents. The latter asks the question whether cooperation can be a stable outcome of the evolution of preferences in the prisoners' dilemma when players can observe and condition their play on some of their opponent's past actions (in encounters with other people).

One such model is given in Robalino and Robson (2012, 2015). In their model, individuals are interacting in ever changing environments. An individual with “theory of mind” (synonymous to cognitive empathy) is able to use past experiences of opponent play to predict more quickly how her opponent will play. Thus, even if it is somewhat costly, in such a setting there is a strict benefit from having a “theory of mind”. One could argue that the incomplete information (about opponents’ preferences) in our model is somewhat akin to the ever changing environment in Robalino and Robson (2015). Our model has no explicit learning. One could perhaps argue it is implicit in our use of Bayesian Nash equilibrium. Our example of a non-conflict game provides a similar result as that in Robalino and Robson (2015) in that any Bayesian Nash equilibrium must exhibit “full” cognitive empathy, i.e. with probability one. In contrast, when we focus on conflict games alone, we find a starkly contrasting result in that any Bayesian Nash equilibrium must exhibit “partial” cognitive empathy, i.e. the probability of acquiring empathy is bounded from below as well as from above, even when costs of acquiring empathy tend to zero.

Another strand of literature started with Aumann and Maschler (1972), who provide an example of a complete information bimatrix game, due to John Harsanyi, that can be used to discuss the relative normative appeal of maxmin and Nash equilibrium strategies. The game is a two-player two-action game and not quite zero sum with a unique Nash equilibrium which is in completely mixed strategies. In this game, Nash equilibrium strategies and maxmin strategies differ for both players. Yet the expected payoff to a given player in the Nash equilibrium is the same as the expected payoff that this player can guarantee herself by playing her maxmin strategy. Pruzhansky (2011) provides a large class of complete information bimatrix games that has this feature. If this is the case, would one not, for this class, recommend players to use their maxmin strategies? In our model, in which players have uncertainty about their opponent’s preferences, and therefore in some sense greater uncertainty about their opponent’s strategy, one might think that the appeal of maxmin strategies is even greater. Yet, in our model there may be a strict benefit from deviating from maxmin strategies.

The literature on level-k thinking typically finds that individuals engaged in game theory experiments do not all reason in the same way as they have different “theories of mind”. See e.g. Stahl and Wilson (1994, 1995); Nagel (1995); Ho et al. (1998); Costa-Gomes et al. (2001); Crawford (2003); Costa-Gomes and Crawford (2006); Crawford and Iriberri (2007). In that sense, our work can be loosely interpreted as a model to understand why there may be individuals of different levels of strategic

thinking.

There is a purely decision theoretic literature on “rational inattention” (e.g. Sims, 2003, 2006; Matějka and McKay, 2012, 2015). In these models, individuals can obtain costly information, where costs are proportional to some measure of informativeness of the possible information to be acquired, before making their ultimate decisions. Our work can be interpreted as an attempt to introduce these considerations into a model of strategic interaction. The individuals in our model can, however, only choose between having perfect information or none.

Moreover, there is a literature on information acquisition in oligopoly models as in e.g. Li et al. (1987), Hwang (1993), Hauk and Hurkens (2001), Dimitrova and Schlee (2003), and Jansen (2008), where firms can acquire information about the uncertain market demand before engaging in oligopoly competition. Market demand enters all agents’ profit functions, whereas in our model the information a player might acquire is exclusively about the opponent’s preferences. More general models in which players acquire information about an uncertain parameter affecting all players’ preferences are given in Hellwig and Veldkamp (2009), Myatt and Wallace (2012), and Amir and Lazzati (2014), as well as in Persico (2000) and Bergemann et al. (2009) in a mechanism design context.

Solan and Yariv (2004) consider a sequential model of two-player two-action interaction in which one player chooses a (possibly mixed) action first, then a second player can buy, at some cost, information about the first player’s (realized) action before finally then also choosing an action herself. The second player can also choose the precision of the information purchased. The structure of the game is common knowledge. In particular the first player is fully aware that she might be spied upon. Thus “spying” in their model is about the opponent’s already determined action with complete information regarding payoffs, whereas in our model “spying” (or cognitive empathy as we call it) is about the opponent’s preferences.

Closest is perhaps Mengel (2012), who studies a model in which individuals play many games and ex ante do not know which game they are playing. Individuals can partition the set of games in any way they like, with the understanding that any two games in the same partition element cannot be distinguished. The individual can condition her action only on the partition element. Adopting a partition comes at some cost, called reasoning costs, and finer partitions are more costly than coarser ones. One difference between Mengel (2012) and what we do here is, therefore, that in our model players always learn their own payoff type, while in Mengel (2012) individuals do not necessarily even learn their own payoff type. Another difference is in

the choice of solution concept, we study Bayesian Nash equilibria while Mengel (2012) studies asymptotically stable strategy profiles under some evolutionary process. Both these differences are probably only superficial. The real difference between the two papers is the class of games they study within their respective models. Our main results deal with the case of conflict games. Mengel (2012) does not explicitly study this class. Therefore, the nature of our results is also different.¹⁷

The rest of the chapter is organized as follows. Section 2.2 states the model. Section 2.3 provides an example of a non-conflict game. Section 2.4 then defines and focuses Bayesian conflict games and provides the main result. In this section we also characterize equilibrium strategy profiles further and provide a uniqueness result. Finally, Section 2.5 concludes with a discussion of further properties of equilibria in Bayesian conflict games as well as a discussion of possible variations of the model. The more complex proofs of results in this chapter are delegated to the appendix, while sketches of these proofs are provided as part of the main text.

2.2 The Model

There are two players $p \in \{B, R\}$. One might think for example of a blue and a red player. Each player p can have one of a finite number n^p of possible (*payoff*) *types* $\theta^p \in \Theta^p$. There are commonly known full support probability distributions over types given by $\mu^p : \Theta^p \rightarrow (0, 1]$ for both players $p \in \{B, R\}$. Abusing notation slightly we sometimes write μ^{θ^p} instead of $\mu^p(\theta^p)$. The types of the two players are drawn from the respective distribution statistically independently from each other. Every type of every player has the same finite set of possible *actions* at her disposal, given by $A = \{a_1, \dots, a_m\}$.¹⁸ *Payoffs* to player $p \in \{B, R\}$ are given by the utility function

¹⁷The main results in Mengel (2012) are that strict Nash equilibria, while (evolutionarily) stable if the game is commonly known, can be made unstable under learning across games; that weakly dominated strategies, while unstable if the game is commonly known, can be stable under learning across games; and that, if all games have distinct Nash equilibrium supports, learning across games under small reasoning costs leads to individuals holding the finest partition with probability one. Our paper is silent on all these results as our conflict games do not have strict Nash equilibria, do not have weakly dominated strategies, and are such that all (what we call realized type) games are such that their Nash equilibria all have full support. All our results, thus, add to the results in Mengel (2012). One could probably translate our main result into the language of Mengel (2012) as follows. If having the finest partition in the model of Mengel (2012) is essentially the same as acquiring cognitive empathy in our model, then our result, that in conflict games we expect proper mixing between acquiring empathy and not acquiring it, suggests that, in conflict games, learning across games as in Mengel (2012) would lead to individuals properly mixing between different partitions, including the finest as well as the coarsest.

¹⁸In principle, one could consider action sets of different cardinality for both players. However, in this chapter we focus on what we call “conflict games” later on and Remark 2.2 (see Appendix 2.C)

$u^{\theta^p} : A \times A \rightarrow \mathbb{R}$, where the first argument depicts the action taken by player p and the second the one taken by her opponent $-p$. Note that different types have different utility functions and that utility functions only depend on the chosen action pair and not directly on the opponent's type.

Before players learn their own type, i.e. at the complete ex-ante stage, each of them can independently and secretly invest a cost of $c \geq 0$ in order to acquire cognitive empathy. This cost, which we refer to as the *cost of empathy acquisition*, is simply subtracted from the player's payoff. A player who acquires empathy then, at the interim stage, learns not only her own type but also the type of her opponent. These player types are then called *informed*. Note, however, that an informed type is not able to observe her opponent's choice of empathy acquisition. We further assume that there is only no empathy or full empathy. When we speak of a player having "partial empathy" we mean that this player randomizes between no and full empathy.¹⁹ A player who does not acquire empathy learns, at the interim stage, only her own type. The corresponding player types are then called *uninformed*.

A *strategy* of player $p \in \{B, R\}$ is then given by a pair $(\rho^p, (\sigma^{\theta^p})_{\theta^p \in \Theta^p})$ where $\rho^p \in [0, 1]$ is the *information strategy*, which we usually refer to as the *probability of empathy acquisition*, and $\sigma^{\theta^p} : \Theta^{-p} \cup \{\emptyset\} \rightarrow \Delta(A)$, the *action strategy*, is the (mixed) action to be played by player p of type $\theta^p \in \Theta^p$ against any opponent of known type $\theta^{-p} \in \Theta^{-p}$, when informed, and of unknown type (which is indicated by the player receiving the uninformative "signal" \emptyset), when uninformed.

Our solution concept is Bayesian Nash equilibrium. Our favorite interpretation of equilibrium is that it is the outcome of a long and slow evolutionary process. It is by now well-known that if any strategy profile is the outcome of a reasonable evolutionary process, it must be an equilibrium.²⁰ As our main result holds for all equilibria of the game, it is therefore true for all candidates of an evolutionary stable outcome.²¹

shows that for this class of games it is inevitable that players' action sets have the same cardinality.

¹⁹Throughout this chapter, "partial empathy" usually comprises the case that the corresponding player acquires empathy with probability zero while always excluding empathy acquisition with probability one.

²⁰See e.g. Weibull (1995) for a textbook treatment for this and all other statements in this paragraph.

²¹It is also well-known that not all games have evolutionary stable outcomes. There can, for instance, be cycles in behavior. Such cycles tend to cycle around equilibria (see e.g. Hofbauer and Sigmund, 1998, Chapter 7.6).

2.3 A Non-Conflict Example

In this section we provide, as a point of contrast to our main results, a non-conflict example.

Example 2.1. Consider a symmetric setup in which both players $p \in \{B, R\}$ can have one of three types $\Theta^B = \Theta^R = \{\theta_1, \theta_2, \theta_3\}$ chosen uniformly (i.e. $\mu^\theta = \frac{1}{3}$ for all $\theta \in \Theta^p$) for the two players. Both players can choose between two actions H and T . Type θ_1 finds action H strictly dominant, type θ_3 finds action T strictly dominant, and type θ_2 has pure coordination preferences. These payoffs can be written in matrix form as given in Figure 2.1.

		H	T			H	T			H	T		
u^{θ_1} :	H	1	1		u^{θ_2} :	H	1	0		u^{θ_3} :	H	0	0
	T	0	0			T	0	1			T	1	1

Figure 2.1: Payoffs of the non-conflict game in Example 2.1

It is straightforward to see that for costs of empathy acquisition sufficiently low (in fact for $c < \frac{1}{9}$) the Bayesian game has no Bayesian Nash equilibrium in which a player acquires empathy with probability less than one. Suppose a player (say blue) does not acquire empathy and fix her opponent's (red) strategy. Then blue does not learn red's preferences. Red, however, makes her choice of action dependent on her own type. Obviously, dominant action types play their dominant actions. Now consider the coordination type of blue. The best she can do in terms of payoffs is to play a best response to the given (mixed) action of the coordination type of red. W.l.o.g. let this best response action be H . The coordination type of blue then receives a payoff of zero against the red type having dominant action T . For blue switching to acquiring empathy and playing T against the T dominant action type of red is then beneficial if $c < \frac{1}{9}$. Thus, for $c < \frac{1}{9}$ any Bayesian Nash equilibrium of this game has both players acquiring empathy with probability one.²²

²² Suppose we reverse the timing of learning one's type and acquiring empathy in this example. That is individuals choose whether or not to acquire empathy after they learn their own type. Of course the two dominant action types do not acquire empathy now but for the coordination type the result is much the same as before: For $c < \frac{1}{3}$ coordination types acquire empathy with probability one in any Bayesian Nash equilibrium of that game.

2.4 Equilibrium Empathy Acquisition

For any pair of types $\theta^B \in \Theta^B$ and $\theta^R \in \Theta^R$ we define the *realized type game* as the complete information game that would result if it were common knowledge among the two players that they are of exactly these two types.

We call the Bayesian game described in Section 2.2 a *Bayesian conflict game* if every possible realized type game has a unique Nash equilibrium and if this Nash equilibrium is in completely mixed strategies.²³ We first show that for positive costs of empathy acquisition there cannot be an equilibrium of a conflict game in which both players choose to acquire empathy with probability one.

Proposition 2.1. *Consider a Bayesian conflict game. If costs of empathy acquisition are positive, then no strategy profile with full empathy, i.e. with $(\rho^B, \rho^R) = (1, 1)$, can be a Bayesian Nash equilibrium. On the contrary, if costs are zero, there is such a full empathy equilibrium.*

Proof of Proposition 2.1. Suppose a conflict game has an equilibrium with $(\rho^B, \rho^R) = (1, 1)$. Then whenever two types $\theta^B \in \Theta^B$ and $\theta^R \in \Theta^R$ meet, it is common knowledge that this is the case and, as this happens with positive probability, they must play a Nash equilibrium of the corresponding realized type game. Any realized type game by definition has a unique Nash equilibrium and this Nash equilibrium is in completely mixed strategies. Thus, every type of every player in every situation is always indifferent between all her pure actions. Hence, when costs are positive, any player would be better off not acquiring empathy, thus saving $c > 0$, and playing any (mixed) action. Arriving at a contradiction, we therefore have the proof for $c > 0$. Observe however that this saving opportunity disappears for $c = 0$, meaning that in this case the above strategy profile is indeed an equilibrium of the conflict game. Throughout the chapter we refer to this as the *full empathy equilibrium*. \square

Note that Proposition 2.1 leaves open the possibility that one (and only one) player acquires empathy with probability one. Turning to a population interpretation of (mixed) equilibrium (as in evolutionary game theory), Proposition 2.1 can be read to say that we expect at least a fraction of the population for at least one player position to not acquire empathy in equilibrium. For instance, if these games are

²³In our main theorem and propositions we write “Bayesian conflict game”, to ensure that a reader who only browses the chapter understands that the conflict games studied in this work have incomplete information. Everywhere else in the chapter we simply write “conflict game” with the understanding that we are nevertheless dealing with a Bayesian conflict game. Analogously, we refer to Bayesian Nash equilibria of Bayesian conflict games simply as equilibria of conflict games.

always played between one man and one woman (both randomly drawn from their respective population), then for at least one of these two populations we expect that, if costs of empathy acquisition are positive, some individuals do not have cognitive empathy.

Suppose we consider symmetric conflict games, such as a Bayesian version of the well-known rock-scissors-paper game. Suppose we are interested in the single population evolutionary model. That is, there is one population of individuals from which repeatedly two are randomly drawn to play the game. Then the appropriate solution concept is symmetric Bayesian Nash equilibrium and Proposition 2.1 implies that this population has a fraction of individuals without cognitive empathy.

In what follows we focus on *two-action Bayesian conflict games*, that is on conflict games in which each player has two actions available. In two-action conflict games we must have that one player always wants to coordinate actions while the other wants to mis-coordinate actions. Throughout the chapter, the former is player B (or blue) and the latter is player R (or red) for convenience. The Bayesian uncertainty is then only about the intensity of these preferences. One could thus alternatively describe a two-action conflict game as a non-zero-sum version of matching pennies with incomplete information. One such game is given in the following example.

Example 2.2. Consider the two-action Bayesian conflict game with action set $A = \{H, T\}$, type sets $\Theta^B = \{\theta_1^B, \theta_2^B\}$ and $\Theta^R = \{\theta_1^R, \theta_2^R\}$, probability distributions over types $\mu^B = \mu^R = (\frac{1}{2}, \frac{1}{2})$, and the payoffs as given in Figure 2.2 (where player B chooses rows and R chooses columns).

		H	T			H	T	
$u^{\theta_1^B}$:	H	1	-1		$u^{\theta_1^R}$:	H	-1	1
	T	-1	1			T	1	-1
		H	T			H	T	
$u^{\theta_2^B}$:	H	3	-1		$u^{\theta_2^R}$:	H	-2	1
	T	-1	1			T	1	-1

Figure 2.2: Payoffs of the conflict game in Example 2.2

Note that the blue player has coordination preferences for all her types, while the red player has mis-coordination preferences. Any realized type game, thus, has only one Nash equilibrium, and that is in completely mixed strategies. The game considered in the example is, therefore, a two-action conflict game. We are interested in the Bayesian Nash equilibria of this game as a function of the cost of empathy

acquisition. We know from Proposition 2.1 that, for positive costs, in equilibrium at least one player does not acquire empathy with probability one. We are particularly interested in how the probability of empathy acquisition in equilibrium changes when we change the corresponding costs. Again from Proposition 2.1, we know that for zero costs the conflict game has an equilibrium in which both players acquire empathy with probability one. Therefore, one would expect that as costs tend to zero, the probability of empathy acquisition of both players in all equilibria of the conflict game tends to one. Surprisingly, this is not the case. This can be seen in the example by computing the (unique) equilibrium of the conflict game for various cost levels.²⁴ These are given in the following table.²⁵

c	ρ^B	ρ^R	$\sigma_H^{\theta^B}(\emptyset)$	$\sigma_H^{\theta^B}(\emptyset)$	$\sigma_H^{\theta^R}(\emptyset)$	$\sigma_H^{\theta^R}(\emptyset)$
10	0	0	0	1	$\frac{2}{3}$	0
1	0	0	0	1	$\frac{2}{3}$	0
$\frac{9}{10}$	0	$\frac{1}{6}$	$\frac{2}{45}$	$\frac{43}{45}$	$\frac{4}{5}$	0
$\frac{4}{5}$	$\frac{1}{10}$	$\frac{1}{6}$	$\frac{4}{81}$	$\frac{68}{81}$	$\frac{98}{125}$	$\frac{2}{125}$
$\frac{1}{2}$	$\frac{1}{10}$	$\frac{1}{6}$	$\frac{16}{81}$	$\frac{56}{81}$	$\frac{16}{25}$	$\frac{4}{25}$
$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{6}$	$\frac{32}{81}$	$\frac{40}{81}$	$\frac{56}{125}$	$\frac{44}{125}$
0	$\frac{1}{10}$	$\frac{1}{6}$	$\frac{4}{9}$	$\frac{4}{9}$	$\frac{2}{5}$	$\frac{2}{5}$

Table 2.1: Equilibria of the conflict game in Example 2.2 for different cost levels $c \geq 0$. We here provide only the equilibrium (mixed) action strategies players use when they do not acquire empathy. For $c \leq \frac{4}{5}$ both players acquire empathy with positive probability and the corresponding types, when informed, always play pure actions.

For high costs, empathy acquisition is strictly dominated and players therefore do not acquire empathy in equilibrium. Moreover, for every player $p \in \{B, R\}$ there seems to be a positive cost level C^p below which this player in equilibrium acquires empathy with positive probability. This equilibrium probability of empathy acquisition ρ^p seems to be strictly greater than zero and strictly less than one and to remain constant for all cost levels lower than C^p . Even in the limit as costs tend to zero the equilibrium still has the same probability of empathy acquisition. Also there seems to be a unique equilibrium for all positive cost levels. In the remainder of

²⁴Uniqueness follows from Proposition 2.3.

²⁵These equilibria were computed using the game theory software Gambit by McKelvey et al. (2014).

this chapter we aim to see which of these statements are generally true in two-action conflict games.

An informed and concerned reader might wonder how it is possible that for all positive cost levels all equilibria of this conflict game have a probability of empathy acquisition that is bounded away from one, given that we know that for zero costs there is an equilibrium with probability one of empathy acquisition and given that we know that the Nash equilibrium correspondence is upper hemi continuous in the space of games (see e.g. Ritzberger, 2002, p. 292). The answer must be that indeed even in the conflict game with zero costs there is a Bayesian Nash equilibrium in which players acquire empathy with a probability that is less than one (see Table 2.1).

While, at this point, it is not at all clear why there would be a unique equilibrium in this game (as long as costs are positive and sufficiently small), one can at least understand the nature of this equilibrium. The key is to understand how the indifference principle, i.e. the fact that in any mixed equilibrium a player must be indifferent between all actions in its support, applies here. By randomizing between acquiring empathy and not doing so, a player does not make her opponent indifferent between acquiring empathy or not. If that were the case, the equilibrium probability of acquiring empathy ρ^p would have to depend on the cost c (the opponent's cost in fact). But this is apparently not the case. In this equilibrium a player mixes between acquiring empathy and not doing so apparently in order to make the uninformed opponent types indifferent between the two actions (see Lemma 2.1). On the other hand, player types, when uninformed, randomize between the two actions in order to make the opponent indifferent between acquiring empathy and refraining from doing so. This is apparent if we consider the mixed actions of the uninformed player types. These mixed actions very much depend on the cost of empathy acquisition. The higher the cost the more diverse are the mixed actions of the uninformed types. This is done in such a way as to keep the player just indifferent between acquiring empathy at costs c and not doing so (see again Table 2.1).

The following theorem is the main result of this chapter. It establishes that in any equilibrium of a two-action conflict game for any of the two players the probability of empathy acquisition is bounded away from zero (if the considered player's opponent has at least two distinct types) and, even more importantly, bounded away from one for all sufficiently small positive costs. In order to state this theorem we require one additional piece of notation. In a two-action conflict game, for any player $p \in \{B, R\}$ of any type $\theta^p \in \Theta^p$ denote by $x(\theta^p)$ the probability of action H that, if played by the

opponent, makes θ^p indifferent between the two actions.²⁶ One could call $x(\theta^p)$ the *indifference probability* of type θ^p . Note that by assumption we have $x(\theta^p) \in (0, 1)$ for all $\theta^p \in \Theta^p$ and $p \in \{B, R\}$. Further, denote by θ_{max}^p (θ_{min}^p) the type which maximizes (minimizes) the indifference probability $x(\theta^p)$.²⁷

Theorem 2.1 (Bounds on Empathy). *Consider a two-action Bayesian conflict game. There exists $C > 0$ such that for all $p \in \{B, R\}$ we have in any Bayesian Nash equilibrium that*

- (i) $\rho^p \geq x(\theta_{max}^{-p}) - x(\theta_{min}^{-p})$ if $c \in [0, C)$ and
- (ii) $\rho^p < \max \{x(\theta_{max}^{-p}), 1 - x(\theta_{min}^{-p})\}$ if $c \in (0, C)$.

The proof of this theorem is somewhat lengthy and, like all other more complex proofs, provided in the appendix. The proof rests on two lemmas that are of some independent interest. We shall now state these lemmas, one after the other, give their respective proof (or a sketch thereof with the full proof in the appendix), and then sketch how they combine with some additional work to establish that equilibrium empathy acquisition probabilities are bounded away from zero and one.

We first show that in equilibrium any uninformed player type must be indifferent between both actions. Just as we do this for the indifference probabilities, we omit the subscript H for ease of notation when considering action strategies in two-action conflict games from here on.

Lemma 2.1. *Consider a two-action Bayesian conflict game. Then there exists $C > 0$ such that for all $c \in [0, C)$, $p \in \{B, R\}$ and $\theta^p \in \Theta^p$ it is*

$$\sum_{\theta^{-p} \in \Theta^{-p}} \mu^{\theta^{-p}} \left(\rho^{-p} \sigma^{\theta^{-p}}(\theta^p) + (1 - \rho^{-p}) \sigma^{\theta^{-p}}(\emptyset) \right) = x(\theta^p) \quad (2.1)$$

in any Bayesian Nash equilibrium.

Sketch of Proof of Lemma 2.1. Suppose there is a player, w.l.o.g. blue, of some type that is uninformed and not indifferent between her two actions. Suppose, w.l.o.g. that she prefers action H . As she is uninformed, she is facing a (mixed) action that is a convex combination of all (mixed) actions of all opponent (red player) types. As she prefers H against this mixture, and as blue is the coordination type, this mixture must place a relatively high probability on H . But as this mixture is a convex combination of mixed actions of all red types there must be one red type who also plays H with

²⁶For a player $p \in \{B, R\}$ we call two types $\theta_1^p, \theta_2^p \in \Theta^p$ *distinct* if it is $x(\theta_1^p) \neq x(\theta_2^p)$.

²⁷If any of these extreme types is not unique, simply choose one maximizer (minimizer) arbitrarily.

higher probability. Thus, the same blue player type, when informed and facing that red type, also plays H . But then the red player, the mis-coordination player, of this type, when informed and playing against the considered blue type, must play T as she is facing the pure action H . This finally can be used to argue that this implies on the one hand that the red player is not acquiring empathy with high probability and on the other hand that she is not playing close to T when of the considered type and uninformed. But then, as costs are small, she should deviate to acquiring empathy with probability one and playing T when meeting this given blue type. \square

The second intermediate result we need is that in equilibrium for each of the two players there must be at least one type who, when informed and playing against certain opponent types, cannot be indifferent between both actions as long as costs are positive.

Lemma 2.2. *Consider a two-action Bayesian conflict game.²⁸ If $c > 0$, then for any Bayesian Nash equilibrium and $p \in \{B, R\}$ with $\rho^p > 0$ there must exist $\bar{\theta}^p \in \Theta^p$ and $\hat{\theta}^{-p}, \tilde{\theta}^{-p} \in \Theta^{-p}$ such that*

$$\rho^{-p}\sigma^{\hat{\theta}^{-p}}(\bar{\theta}^p) + (1 - \rho^{-p})\sigma^{\hat{\theta}^{-p}}(\emptyset) > x(\bar{\theta}^p), \quad (2.2a)$$

$$\rho^{-p}\sigma^{\tilde{\theta}^{-p}}(\bar{\theta}^p) + (1 - \rho^{-p})\sigma^{\tilde{\theta}^{-p}}(\emptyset) < x(\bar{\theta}^p). \quad (2.2b)$$

For $p = B$ ($p = R$) this induces $\sigma^{\bar{\theta}^B}(\hat{\theta}^R) = 1$ and $\sigma^{\bar{\theta}^B}(\tilde{\theta}^R) = 0$ ($\sigma^{\bar{\theta}^R}(\hat{\theta}^B) = 0$ and $\sigma^{\bar{\theta}^R}(\tilde{\theta}^B) = 1$).

Proof of Lemma 2.2. This proof is similar to that of Proposition 2.1. Suppose a player $p \in \{B, R\}$ acquires empathy with some positive probability $\rho^p > 0$ in equilibrium while costs are positive, i.e. $c > 0$. Now assume that every type θ^p of player p , when informed, is indifferent between the two actions H and T against any opponent type θ^{-p} . Then player p could benefit strictly from deviating to acquiring empathy with probability zero (thus, saving costs $c > 0$ with probability $\rho^p > 0$) and playing any (mixed) action (not losing anything because of the complete indifference). Arriving at a contradiction, we therefore have that there must be at least one player type $\bar{\theta}^p$ strictly preferring H or T against some opponent type here. Together with Lemma 2.1 this concludes the proof. \square

²⁸The reader may feel that we use an overabundance of different types in the statement of this lemma. This is, unfortunately, necessary. There are three different types for each player, denoted $\bar{\theta}^p$, $\hat{\theta}^p$, and $\tilde{\theta}^p$. It is important to realize that generally it may well be that all three types on each side are different from each other. In the case where there are only two types for one player, some of these three types naturally must coincide. This additional structure allows us to prove more in such cases. See Proposition 2.3.

Note that, for the special case that the opponent $-p$ only has one possible type, Lemma 2.2 implies that for any positive cost of empathy acquisition player p does not acquire empathy, i.e. $\rho^p = 0$, in any equilibrium. This must be true as in this case player p already knows her opponent's only possible type (this is implicit in the assumption of common knowledge of the conflict game).

Consider first Part (i) of Theorem 2.1, which states that there is a specific lower bound on the equilibrium probabilities of empathy acquisition.

Sketch of Proof of Theorem 2.1(i). The key to this part is Lemma 2.1. It states that every type of any player, when uninformed, must be indifferent between both actions as long as costs are sufficiently small. Consider, w.l.o.g., the red player and assume that $x(\theta_{max}^R) > x(\theta_{min}^R)$ (otherwise the lower bound is trivially satisfied). Now both player types θ_{max}^R and θ_{min}^R must be indifferent between both actions when uninformed. These two red types, however, face the same distribution over actions if the blue player's probability of empathy acquisition is zero. Why? If the blue player did not acquire empathy, she cannot recognize the red player's type and cannot condition her action strategy on that information. On the other hand, the two (extreme) red types cannot be both indifferent between the two actions if they are facing the same distribution. Thus, arriving at a contradiction, it must be that the blue player acquires empathy with positive probability. In fact the exact lower bound can be obtained by taking the difference between equation(s) (2.1) for the extreme types θ_{max}^R and θ_{min}^R . \square

The key statement in Part (ii) of the theorem is that it establishes an upper bound, strictly below one, for each player's equilibrium probability of empathy acquisition. What this upper bound is, is less important. In the appendix we, in fact, prove two results that imply the existence of an upper bound strictly below one. One is as stated in Theorem 2.1(ii), the other is stated in the appendix as Theorem 2.1(ii)'. The respective statements are similar but neither implies the other. The former is more elegant in its expression, the latter is more intuitive in its proof. Therefore, we choose to present the sketch of proof for the more intuitively explainable Theorem 2.1(ii)' here.

Sketch of Proof of Theorem 2.1(ii)'. Let us, w.l.o.g., focus on the blue player (the coordination preference player). Assume that the blue player acquires empathy with a probability greater than the stated bound, say close to one. By Lemma 2.2 we know that, in equilibrium, there must be a type of blue player who, when informed, has the strict preference to play H against some type of red player. This type of red

player, when informed herself and meeting the given type of blue player, then faces with high likelihood an informed blue player who plays H . Her best response then (being the mis-coordination player) is to play T against this blue type. But as the informed blue type's equilibrium action against this red type is H , two things must be true about the red player. First, she cannot be too informed, i.e. her probability of acquiring empathy must be low, and second, when she is of the considered type and uninformed, she must play H with a high probability. But this means, as the cost of empathy acquisition is small, that the red player could strictly benefit from deviating to acquiring empathy and then, when she is of this red type, play T against this blue type. Arriving at a contradiction, in any equilibrium the assumption of a highly empathic blue player cannot hold. A similar argument can be made for the red player. \square

While generally we do not know whether there can be equilibria in two-action conflict games in which a player's probability of empathy acquisition is close to the upper bound(s) stated in Theorem 2.1, we can establish that there is, for small costs, always an equilibrium in which the lower bound is achieved.

Proposition 2.2. *For every two-action Bayesian conflict game there exists $C > 0$ such that for all $c \in [0, C)$ it has a Bayesian Nash equilibrium with*

$$\rho^p = x(\theta_{max}^{-p}) - x(\theta_{min}^{-p})$$

for both $p \in \{B, R\}$.

The proof of this proposition is constructive and given in the appendix. The key to reaching the lower bound for the probability of empathy acquisition is to let all informed types of player $p \in \{B, R\}$ play H against opponent type θ_{max}^{-p} and T against type θ_{min}^{-p} . Taking into account Lemma 2.1, this immediately pins down the equilibrium probability of empathy acquisition of player p and it is exactly the lower bound. The equilibrium is then further constructed by letting uninformed types mix in a way that makes the opponent $-p$ indifferent between acquiring empathy and not doing so. This is similar as in the discussion of the equilibria in Example 2.2.

With more than two types for one player and at least two for the other, this kind of equilibrium can be constructed in different ways. In general, this gives rise to a continuum of equilibria that differ in terms of players' action strategies but not in terms of their information strategies. See Corollary 2.3 in Appendix 2.B (and the proof of Proposition 2.2 which additionally captures the case in which a player does not have distinct types) for a characterization of this class of equilibria. In what

follows, any representative of this class is called a *partial empathy equilibrium*. Note, however, that a player acquires empathy with probability zero in such an equilibrium if her opponent does not have distinct types.

We do not know whether general two-action conflict games with positive costs of empathy acquisition can actually have equilibria in which a player's probability of empathy acquisition is strictly greater than this lower bound. For any such game with either two types for both players or a single type for one player we can show, however, that the partial empathy equilibrium considered in the proof of Proposition 2.2 is indeed the only equilibrium.²⁹

Proposition 2.3. *For every two-action Bayesian conflict game with only one type for one player and more than one type for the other player or with exactly two types for both players there exists $C > 0$ such that for all $c \in (0, C)$ and $p \in \{B, R\}$ the probability of empathy acquisition is $\rho^p = x(\theta_{max}^{-p}) - x(\theta_{min}^{-p})$ in any Bayesian Nash equilibrium. In these cases the Bayesian Nash equilibrium is uniquely determined by the strategy profile considered in the proof of Proposition 2.2 if there is a unique maximal type θ_{max}^p and a unique minimal type θ_{min}^p for both players $p \in \{B, R\}$.³⁰*

The proof is again given in the appendix. The key for this proposition is to realize that with a limited number of types for at least one player we can pin down behavior of all types of this player fairly quickly with the help of Theorem 2.1. Things turn out to be much more complex if both players have many types, as then all we know is, for instance, that there is one type of player red who plays H against some type of player blue, but we do not know which types these are. If there are only two types on both sides, for instance, then by starting with one type who does something specific against one opponent type all other types' behavior follows.

Finally, note that applying Proposition 2.3 to the game considered in Example 2.2 implies that the partial empathy equilibrium we listed in Table 2.1 for $c = \frac{1}{10}, \frac{1}{2}, \frac{4}{5}$ is unique for $c > 0$ sufficiently small.

²⁹While, generally, Corollary 2.3 characterizes a continuum of equilibria, it is easy to see that, in the case with two (distinct) types per player, this condenses to the equilibrium considered in the proof of Proposition 2.2.

³⁰In the case in which one player $p \in \{B, R\}$ has a single type the equilibrium is in fact unique up to variations of the action strategies $\sigma^{\theta^{-p}}(\theta^p)$ of the informed opponent types which are played with probability $\rho^{-p} = 0$.

2.5 Discussion and Conclusion

In this work, we study two-player conflict situations with ex-ante uncertainty over (the exact) opponent preferences for both players. We allow players, before learning their own payoff type, to acquire cognitive empathy at some (small) cost. Cognitive empathy enables a player to learn the preferences of her opponent in all situations. There are at least two ways we can interpret this model. The first interpretation is that there are indeed two strategic opponents (the two soccer players from the introduction, two firms, two military generals, etc.) who are involved in a conflict situation and who can acquire information about their opponent's ex ante unknown preferences. Given this interpretation, we find that in equilibrium these strategic players do not fully acquire information about their opponent's preferences, even if the cost of doing so is vanishingly small. A second interpretation is that there are many individuals who are often and somewhat randomly engaged in pairwise conflict situations and mother nature can endow these individuals (each individual separately) with cognitive empathy, i.e. with the ability to understand opponents' preferences, at some positive cost (e.g. by providing an additional brain function). Under the assumption that nature guides play to an evolutionary stable state, which must be a Bayesian Nash equilibrium of this game, our results can be read to imply that nature endows some but not all of her subjects with cognitive empathy, even if the costs of doing so are essentially zero.

Our model is simple and sparse and many alterations and additions are conceivable. In what follows we discuss additional consequences of our results as well as some possible modifications of our model and what we know or believe about how these change our main results.

2.5.1 Empathy Acquisition at Zero Costs

In this subsection we provide a corollary to (the proof of) Proposition 2.2 for the special case of zero costs of empathy acquisition that allows us to provide additional intuition for our main result.

Corollary 2.1. *Any two-action Bayesian conflict game with $c = 0$ has a Bayesian Nash equilibrium with partial empathy, i.e. $\rho^p \in [0, 1)$, and*

$$\rho^p \sigma^{\theta^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta^p}(\emptyset) = x(\theta^{-p})$$

for all $p \in \{B, R\}$, $\theta^p \in \Theta^p$, $\theta^{-p} \in \Theta^{-p}$.

Why is this of interest? This corollary implies that an outside observer who can observe the two players' types would observe, for any pair of types, a frequency of actions exactly as given by the Nash equilibrium of the realized type game (in which that game is common knowledge). In other words, even though when the two types meet, players are far from having common knowledge that the two are of these particular types, they nevertheless manage to play "as if" they had common knowledge of this fact.³¹ Expressed differently, the two equilibria we identify for two-action conflict games with zero costs are, in this sense, observationally equivalent. That is to say an outside observer who observes all types would see the same frequency of actions (as a function of type pairs) in both equilibria.

In another sense, the two equilibria are observationally distinct. Consider the evolutionary interpretation of this game. Then individuals would either have empathy or not and both kinds would exist in the partial empathy equilibrium but not in the full empathy equilibrium. In the partial empathy equilibrium, an outside observer, who observes all types and could follow individuals' behavior in many conflict situations, could quickly identify which of the individuals are empathic (always uses the same pure action against certain opponent types) and which are not (these always mix between pure actions against the same type of opponent). In the full empathy equilibrium the observer would note that all types always mix (in a certain way). Thus, the observer could tell whether one or the other equilibrium is played (provided the observer can observe many interactions of the same individuals).

2.5.2 Equilibrium Payoffs

In this subsection we turn to a discussion of equilibrium payoffs in two-action conflict games with the (costly) possibility of empathy acquisition. Note that, when we talk about the payoff to an (informed) type, we mean the payoff without taking into account the costs of empathy acquisition that players have to bear. In contrast, these costs are included when considering players' ex-ante expected payoffs. In what follows, the full empathy equilibrium under zero costs, which we know from Proposition 2.1, is referred to as the benchmark case. Consider first the case of costs being small.

Corollary 2.2. *Consider a two-action Bayesian conflict game. There exists $C > 0$ such that for all $c \in (0, C)$ we have that in any Bayesian Nash equilibrium*

- (i) *every player obtains an ex-ante expected payoff equal to her ex-ante expected payoff in the benchmark case,*

³¹This insight could be useful if one were to attempt to generalize our result to more than two actions.

- (ii) every uninformed type for each player obtains the same expected payoff as she does in the benchmark case, and
- (iii) for each player acquiring empathy with positive probability there is at least one type who, when informed, obtains a strictly higher expected payoff than she obtains in the benchmark case.

This set of statements is a corollary to Lemma 2.1, Theorem 2.1, and Proposition 2.1. In particular, Part (i) follows from the fact that all types, when uninformed, are indifferent between both actions in any equilibrium by Lemma 2.1 and all players acquiring empathy with positive probability are ex ante indifferent between acquiring empathy and not doing so by Theorem 2.1(ii). Part (ii) follows from Lemma 2.1 alone. Part (iii) follows from Part (i) and the fact that players have to bear a cost of $c > 0$ for acquiring empathy.

This corollary, thus, states that there is a sense in which in two-action conflict games for all cost levels, provided they are small enough, all equilibria are ex ante payoff equivalent (if we consider payoffs net of costs). As costs are positive, this implies that some types of players must, when informed, expect higher payoffs than they expect when they are uninformed. However, unlike in the proof of Proposition 2.2, this must not be the case for all types. According to Corollary 2.3 there are in general even equilibria in which only one type of a player obtains a payoff being strictly greater than in the benchmark case.

Before we turn to the case of large costs, it is fruitful to partition the class of conflict games into two subclasses. These are inspired by Pruzhansky (2011). For every type $\theta^p \in \Theta^p$ of a player $p \in \{B, R\}$ define the *type-induced zero-sum game* as the complete information game in which player p has preferences given by her type, i.e. given by u^{θ^p} , and her opponent has preferences $-u^{\theta^p}$. We call a type *immunizable* if the type-induced zero-sum game has no strictly dominated (mixed) strategy for both players. Moreover, we call a type *robustly immunizable* if the type-induced zero-sum game has no weakly dominated (mixed) strategy for both players.³² Let us further call a conflict game *immunizable* if every type of every player is immunizable. On the contrary, if in a conflict game there is at least one type being not immunizable, then this game is called *non-immunizable*. Example 2.2 is an example of an immunizable (two-action) conflict game. There are, however, conflict games that are non-immunizable. Consider again Example 2.2 and modify the preferences

³²Note that in a conflict game no type has a dominated action strategy. A type in a conflict game is therefore (robustly) immunizable if her fictitious zero-sum opponent in the type-induced zero-sum game has no strictly (weakly) dominated strategy.

of the second type of player B as given in Figure 2.3.

	H	T
H	3	-1
T	2	1

Figure 2.3: Modified payoffs to player θ_2^B in the conflict game provided in Example 2.2.

The game, so modified, is still a conflict game (i.e. every realized type game has a unique Nash equilibrium and that is in mixed strategies). However, the game is non-immunizable as the modified second type of player B is not immunizable: the type-induced zero-sum game for this modified second type of player B is such that, for the fictitious zero-sum opponent, action T strictly dominates action H .

We choose the label “immunizable” because of a result due to Pruzhansky (2011, p. 355). He shows that in any complete information game with two immunizable players (in the above sense) both players have “equalizer” strategies. If a player adopts an “equalizer” strategy, she gets the same expected payoff regardless of the action taken by the opponent. He then shows in his Lemma 1, that in any complete information game with immunizable players on both sides every equalizer strategy is a maxmin strategy. Moreover, he shows in his Lemma 2 that equalizer strategies guarantee the player the Nash equilibrium payoff in such games. This generalizes the insight found by Aumann and Maschler (1972) in their example.

With this partition of conflict games in hand we can now turn to the discussion of payoffs in equilibria for large costs of empathy acquisition.

Remark 2.1. *Consider a two-action Bayesian conflict game with costs of empathy acquisition so high that any strategy including empathy acquisition is dominated by one without empathy acquisition. If this game is immunizable, then in any Bayesian Nash equilibrium we have that*

- (i) *for every player every (necessarily uninformed) type obtains an expected payoff that is at least as large as in the benchmark case, and*
- (ii) *every player $p \in \{B, R\}$ having at least two robustly immunizable types $\theta_1^p, \theta_2^p \in \Theta^p$ with $x(\theta_1^p) \neq x(\theta_2^p)$ obtains an ex-ante expected payoff strictly larger than her ex-ante expected payoff in the benchmark case.*

Further, if this game is non-immunizable, then in any Bayesian Nash equilibrium we have that

(iii) for every player every (necessarily uninformed) type being not immunizable obtains an expected payoff that is strictly larger than her maxmin payoff.

However, in a Bayesian Nash equilibrium of such a non-immunizable game there can be

(iv) (necessarily uninformed) types of a player who obtain an expected payoff that is strictly larger, respectively strictly lower, than her expected payoff in the benchmark case, and even

(v) players who have an ex-ante expected payoff that is strictly larger, respectively strictly lower, than her ex-ante expected payoff in the benchmark case.

To see Part (i) of the remark one can use the result of Pruzhansky (2011) (see above) that in such games any type of any player's maxmin payoff is equal to her Nash equilibrium payoff in any realized type game. The latter payoff is the payoff this type of player obtains in the benchmark case. As she can always guarantee herself this payoff by playing her maxmin action strategy, she can certainly never receive less in any equilibrium for any cost level. Moreover, as players are uninformed here, each type faces the same average opponent action strategy. Under the additional assumption of Part (ii) this means that in any equilibrium at least one of the two robustly immunizable types must have incentives to play a pure action strategy which makes her strictly better off than in the benchmark case. This, together with Part (i), proves Part (ii).

Part (iii) of the remark follows from the observation that in a two-action conflict game, to prevent a player type that is not immunizable from obtaining more than the maxmin payoff (which she can of course guarantee herself), the opponent needs to play a pure action. However, one can show that in any equilibrium of such a game the opponent, on average, does not use a pure action strategy. Therefore every such player type must receive a payoff being strictly larger than her maxmin payoff.

Finally, to see Parts (iv) and (v) of the remark, we consider the following example.

Example 2.3. Consider the two-action Bayesian conflict game with action set $A = \{H, T\}$, type sets $\Theta^B = \{\theta_1^B, \theta_2^B\}$ and $\Theta^R = \{\theta_1^R, \theta_2^R\}$, probability distributions over types $\mu^B = \mu^R = \left(\frac{1}{2}, \frac{1}{2}\right)$, and the payoffs as given in Figure 2.4 with $a, b \in \mathbb{R}$ (where player B chooses rows and R chooses columns).

In this example, if we set $a = -1$ and $b = 1$ we obtain Example 2.2. Now consider $a = 2$ and $b = -\frac{3}{2}$. Note first that this is still a conflict game but that it is non-immunizable as, given these parameter values, types θ_2^B and θ_2^R are not immunizable.

	<table border="1" style="border-collapse: collapse; margin: auto;"> <tr><td style="padding: 5px;"></td><td style="padding: 5px; text-align: center;">H</td><td style="padding: 5px; text-align: center;">T</td></tr> <tr><td style="padding: 5px; text-align: center;">$u^{\theta_1^B}$:</td><td style="padding: 5px; text-align: center;">H</td><td style="padding: 5px; text-align: center;">T</td></tr> <tr><td style="padding: 5px;"></td><td style="padding: 5px; text-align: center;">1</td><td style="padding: 5px; text-align: center;">-1</td></tr> <tr><td style="padding: 5px;"></td><td style="padding: 5px; text-align: center;">-1</td><td style="padding: 5px; text-align: center;">1</td></tr> </table>		H	T	$u^{\theta_1^B}$:	H	T		1	-1		-1	1	<table border="1" style="border-collapse: collapse; margin: auto;"> <tr><td style="padding: 5px;"></td><td style="padding: 5px; text-align: center;">H</td><td style="padding: 5px; text-align: center;">T</td></tr> <tr><td style="padding: 5px; text-align: center;">$u^{\theta_1^R}$:</td><td style="padding: 5px; text-align: center;">H</td><td style="padding: 5px; text-align: center;">T</td></tr> <tr><td style="padding: 5px;"></td><td style="padding: 5px; text-align: center;">-1</td><td style="padding: 5px; text-align: center;">1</td></tr> <tr><td style="padding: 5px;"></td><td style="padding: 5px; text-align: center;">1</td><td style="padding: 5px; text-align: center;">-1</td></tr> </table>		H	T	$u^{\theta_1^R}$:	H	T		-1	1		1	-1
	H	T																								
$u^{\theta_1^B}$:	H	T																								
	1	-1																								
	-1	1																								
	H	T																								
$u^{\theta_1^R}$:	H	T																								
	-1	1																								
	1	-1																								
	<table border="1" style="border-collapse: collapse; margin: auto;"> <tr><td style="padding: 5px;"></td><td style="padding: 5px; text-align: center;">H</td><td style="padding: 5px; text-align: center;">T</td></tr> <tr><td style="padding: 5px; text-align: center;">$u^{\theta_2^B}$:</td><td style="padding: 5px; text-align: center;">H</td><td style="padding: 5px; text-align: center;">T</td></tr> <tr><td style="padding: 5px;"></td><td style="padding: 5px; text-align: center;">3</td><td style="padding: 5px; text-align: center;">-1</td></tr> <tr><td style="padding: 5px;"></td><td style="padding: 5px; text-align: center;">a</td><td style="padding: 5px; text-align: center;">1</td></tr> </table>		H	T	$u^{\theta_2^B}$:	H	T		3	-1		a	1	<table border="1" style="border-collapse: collapse; margin: auto;"> <tr><td style="padding: 5px;"></td><td style="padding: 5px; text-align: center;">H</td><td style="padding: 5px; text-align: center;">T</td></tr> <tr><td style="padding: 5px; text-align: center;">$u^{\theta_2^R}$:</td><td style="padding: 5px; text-align: center;">H</td><td style="padding: 5px; text-align: center;">T</td></tr> <tr><td style="padding: 5px;"></td><td style="padding: 5px; text-align: center;">-2</td><td style="padding: 5px; text-align: center;">b</td></tr> <tr><td style="padding: 5px;"></td><td style="padding: 5px; text-align: center;">1</td><td style="padding: 5px; text-align: center;">-1</td></tr> </table>		H	T	$u^{\theta_2^R}$:	H	T		-2	b		1	-1
	H	T																								
$u^{\theta_2^B}$:	H	T																								
	3	-1																								
	a	1																								
	H	T																								
$u^{\theta_2^R}$:	H	T																								
	-2	b																								
	1	-1																								

Figure 2.4: Payoffs of the conflict game in Example 2.3.

Further, note that the indifference probabilities are given by $x(\theta_1^B) = \frac{1}{2}$, $x(\theta_2^B) = \frac{2}{3}$, $x(\theta_1^R) = \frac{1}{2}$, and $x(\theta_2^R) = \frac{4}{5}$.

One can verify that the following is an equilibrium of this game under large costs. Obviously, we need to have $\rho^B = \rho^R = 0$, i.e. no empathy is acquired. Furthermore, let $\sigma^{\theta_1^B}(\emptyset) = \sigma^{\theta_2^R}(\emptyset) = 1$ and $\sigma^{\theta_2^B}(\emptyset) = \sigma^{\theta_1^R}(\emptyset) = 0$.

One can then verify that type θ_2^B receives an equilibrium payoff of $\frac{3}{2}$ while in any realized type game her payoff in the unique Nash equilibrium would be $\frac{5}{3}$. Her payoff in the considered equilibrium of the conflict game under large costs is thus strictly lower than her payoff in the benchmark case. On the other hand, type θ_2^R receives an equilibrium payoff of $-\frac{1}{2}$ which is strictly larger than her payoff of $-\frac{7}{5}$ which she obtains in any realized type game and, thus, in the benchmark case. As in the considered equilibrium all other types expect the same payoff (of zero) as in the benchmark case, player B receives a lower ex-ante expected payoff here than in the benchmark case, while for player R the opposite is true.

2.5.3 The Timing of Decisions

Given the evolutionary interpretation of our model and the idea that nature's subjects play many conflict games with often different preferences throughout their life, it seems appropriate that nature makes the decision about empathy acquisition at the very beginning. Also in the other interpretation, in which players are consciously strategic about their choice of information acquisition, it can make sense to have the information acquisition decision before knowing the exact nature of the conflict situation. A soccer team may study the opposing goalkeeper for the eventuality of a penalty kick before knowing whether the goalkeeper will incur an injury or which of their own players will actually take the penalty kick. A military general might want to spy on her opponent's preferences before knowing the future strength of the own troop or on which terrain, in which place, at which state of the war etc. the actual battle will take place. A firm might decide on research activities on another firm's

motives before they know whether they are facing a merger or a hostile takeover.

There are certainly cases, however, in which the reverse timing is just as plausible. That is we could envision a version of our model in which players consider acquiring the information about their opponent's preferences only after they know their own preferences.

In Footnote 22 we have already thought about this issue and found that, for the given non-conflict example (see Example 2.1), the main insight does not change. We have also looked at the reverse-timing model for the conflict example given as Example 2.2. We shall not go through this here but it suffices to say that, while small details change, the main result, that for small positive costs of empathy acquisition any equilibrium has partial cognitive empathy, seems to remain unchanged.³³ In fact, all types acquire partial empathy in this example: the probability of empathy acquisition is, as in our main result, bounded from below and above. It is, however, unlike in our model with two actions and two types, not constant for all small costs of empathy acquisition. Still, we expect our main theorem to hold in principal also in the model in which the timing of empathy acquisition and of learning one's type is reversed.

2.5.4 Degrees of Cognitive Empathy

Another issue, especially for the evolutionary interpretation of our model, is this. If nature has to make her decision on cognitive empathy at the beginning once and for all possible situations, then these "all possible situations" should probably cover more than just conflict games. And, if these situations include, for instance, the three possible types (for both players) as given in our non-conflict example, then for small costs nature would always endow her subjects with full empathy. One could now state that it is therefore a question of which is smaller, the cost of empathy acquisition or the probability of these three types, but this is not where we want to go in this discussion. Instead, we think that a better model in such cases would be one in which nature can give her subjects degrees of empathy. For instance, nature could give us enough cognitive empathy to always check whether or not our opponent has a dominant action strategy, but if our opponent does not, nature may not give us more cognitive empathy to differentiate our opponent's preferences further. The result would then be as in our model.

A similar response could be made to the ultimate implication of the following

³³To be precise, we used Gambit by McKelvey et al. (2014) and found exactly one equilibrium. We have not attempted to prove that this equilibrium is unique but we conjecture that it is.

consideration. Consider, for convenience, our result for two-action conflict games with two types per player. For these games Proposition 2.3 implies that a player’s probability of empathy acquisition is exactly given by the difference of the two indifference probabilities of her two opponent types. This means that the more similar her two opponent types are, the more similar are their indifference probabilities and the less empathy is acquired by her. This is also true for the lower bound established for the probability of empathy acquisition in our main theorem. In particular, this also implies that the more different kinds of situations a person faces, i.e. the bigger the possible difference between the possible opponent types, the more empathy is acquired. If this goes as far as to include even dominant strategy types, she has to acquire full empathy. To tackle this issue one could build a model of empathy acquisition more like that of “rational (in)attention” as in the decision theoretic models of Sims (2003, 2006); Matějka and McKay (2012, 2015). Adapting these models to our strategic interaction setting could be done by allowing players to buy signals about their opponent’s preferences of any precision but where the costs of these signals are increasing in the information content of these signals, as measure, for instance, by their entropy. Another model would be to allow individuals to acquire multiple signals of whatever precision, one after the other, about their opponent’s preferences, before making their final action decision. While we do not think that the main insight of our work would change in such a model, it might nevertheless add substantial additional insights, the pursuit of which we leave to future research.

Appendix 2.A Proofs

Throughout this section we again abuse notation of action strategies in two-action conflict games slightly by denoting by $\sigma^{\theta^p}(\cdot) \in [0, 1]$ the probability of H chosen by player p of type θ^p . For ease of notation, when it comes to the arguments of utility functions u^{θ^p} , we also only mention the probabilities of action H . And finally, let $\mathcal{U}_{Info}^{\theta^p}$ denote the (interim) expected payoff of a type $\theta^p \in \Theta^p$ of player p if she acquired empathy and before she learns her opponent’s type. Similarly, $\mathcal{U}_N^{\theta^p}$ denotes the expected payoff of a type θ^p of player p who did not acquire empathy.

2.A.1 Proof of Lemma 2.1

For $p \in \{B, R\}$, $\theta^p \in \Theta^p$ we define

$$C^{BH}(\theta^B, \theta^R) := u^{\theta^R}(0, 1) - u^{\theta^R}(x(\theta^B), 1),$$

$$\begin{aligned}
 C^{B_T}(\theta^B, \theta^R) &:= u^{\theta^R}(1, 0) - u^{\theta^R}(x(\theta^B), 0), \\
 C^{R_H}(\theta^B, \theta^R) &:= u^{\theta^B}(1, 1) - u^{\theta^B}(x(\theta^R), 1), \text{ and} \\
 C^{R_T}(\theta^B, \theta^R) &:= u^{\theta^B}(0, 0) - u^{\theta^B}(x(\theta^R), 0).
 \end{aligned}$$

Notice that $C^B(\theta^B, \theta^R) > 0$ ($C^R(\theta^B, \theta^R) > 0$) for all $\theta^B \in \Theta^B, \theta^R \in \Theta^R$ as player R wants to mis-coordinate (as player B wants to coordinate). Based on this let

$$C := \min_{a \in \{B_H, B_T, R_H, R_T\}} \min_{\theta^B, \theta^R} \mu^{\theta^B} \mu^{\theta^R} C^a(\theta^B, \theta^R).$$

W.l.o.g. we consider player $p = B$ and assume that we have

$$\sum_{\theta^R} \mu^{\theta^R} (\rho^R \sigma^{\theta^R}(\bar{\theta}^B) + (1 - \rho^R) \sigma^{\theta^R}(\emptyset)) > x(\bar{\theta}^B)$$

for some $\bar{\theta}^B \in \Theta^B$.³⁴ Since player B wants to coordinate actions, this implies $\sigma^{\bar{\theta}^B}(\emptyset) = 1$ (if $\rho^B < 1$). Furthermore, if a probability weighted sum of terms exceeds $x(\bar{\theta}^B)$, then at least one term must exceed $x(\bar{\theta}^B)$ as well. Thus, there must exist a type $\bar{\theta}^R$ such that

$$\rho^R \sigma^{\bar{\theta}^R}(\bar{\theta}^B) + (1 - \rho^R) \sigma^{\bar{\theta}^R}(\emptyset) > x(\bar{\theta}^B). \quad (2.3)$$

In turn, this implies $\sigma^{\bar{\theta}^B}(\bar{\theta}^R) = 1$ (if $\rho^B > 0$), meaning that

$$\rho^B \sigma^{\bar{\theta}^B}(\bar{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset) = 1 > x(\bar{\theta}^R).$$

Moreover, it is obvious that this equality and inequality also hold for $\rho^B = 0$ and $\rho^B = 1$. As player R wants to mis-coordinate, this implies $\sigma^{\bar{\theta}^R}(\bar{\theta}^B) = 0$ (if $\rho^R > 0$). Inserting the latter into inequality (2.3) gives

$$(1 - \rho^R) \sigma^{\bar{\theta}^R}(\emptyset) > x(\bar{\theta}^B).$$

Again, it is obvious that this inequality is satisfied for $\rho^R = 0$ as well. It follows from this that $1 - \rho^R > x(\bar{\theta}^B)$ and $\sigma^{\bar{\theta}^R}(\emptyset) > x(\bar{\theta}^B)$. Hence, for $c \in [0, C)$ player R can improve her payoff by deviating to a strategy with $\hat{\rho}^R = 1$ and obtaining an

³⁴Observe that the subsequent line of argument is almost identical for the reversed inequality as well as for $p = R$. Thus, we can omit these cases.

additional payoff of at least

$$\begin{aligned}
& (1 - \rho^R) \left(\mu^{\bar{\theta}^R} \mu^{\bar{\theta}^B} \left(u^{\bar{\theta}^R}(0, 1) - u^{\bar{\theta}^R}(\sigma^{\bar{\theta}^R}(\emptyset), 1) \right) - c \right) \\
& > x(\bar{\theta}^B) \left(\mu^{\bar{\theta}^R} \mu^{\bar{\theta}^B} \left(u^{\bar{\theta}^R}(0, 1) - u^{\bar{\theta}^R}(x(\bar{\theta}^B), 1) \right) - c \right) \\
& > 0.
\end{aligned}$$

We thus arrive at a contradiction. \square

2.A.2 Proof of Theorem 2.1

We need one additional, purely technical lemma, in order to prove Theorem 2.1.

Lemma 2.3. *Consider $\alpha, \beta', \beta'', \gamma \in \mathbb{R}$ where $\beta' - \beta'' \leq \alpha$. Then (at least) one of the following three conditions must be satisfied:*

$$\alpha + (1 - \alpha)\gamma = \beta' \text{ and } (1 - \alpha)\gamma = \beta'', \quad (2.4a)$$

$$\alpha + (1 - \alpha)\gamma > \beta' \text{ or } \quad (2.4b)$$

$$(1 - \alpha)\gamma < \beta''. \quad (2.4c)$$

Proof of Lemma 2.3. Suppose none of the three conditions is satisfied. Then we obtain

$$\alpha + (1 - \alpha)\gamma < (\leq)\beta' \text{ and}$$

$$(1 - \alpha)\gamma \geq (>)\beta''.$$

In either case, subtracting the second from the first inequality yields

$$\alpha < \beta' - \beta'' \leq \alpha,$$

a contradiction. \square

Having established this lemma, we can now turn to the proof of the theorem.

Proof of Theorem 2.1. In the following we distinguish between the two parts.

Part (i): Lower Bound

From Lemma 2.1 and equation (2.1) it follows immediately that for all $p \in \{B, R\}$,

$\theta^p \in \Theta^p$

$$\rho^p \left(\underbrace{\sum_{\theta^p} \mu^{\theta^p} \sigma^{\theta^p}(\theta_{max}^{-p})}_{\leq 1} - \underbrace{\sum_{\theta^p} \mu^{\theta^p} \sigma^{\theta^p}(\theta_{min}^{-p})}_{\geq 0} \right) = x(\theta_{max}^{-p}) - x(\theta_{min}^{-p}).$$

Hence, we have that $\rho^p \geq x(\theta_{max}^{-p}) - x(\theta_{min}^{-p})$.

Part (ii): Upper Bound

We here prove the upper bound for player p . The statement is trivially satisfied if $\rho^p = 0$. Thus, suppose that $\rho^p > 0$. We need to distinguish two different cases.

Case 1: $\rho^{-p} = 0$

Given the lower bound we proved in Part (i), we then must have that $x(\theta_{max}^p) = x(\theta_{min}^p)$. Lemma 2.2 then implies that there are two opponent types $\hat{\theta}^{-p}$ and $\tilde{\theta}^{-p}$ such that

$$\begin{aligned} \sigma^{\hat{\theta}^{-p}}(\emptyset) &> x(\theta^p), \\ \sigma^{\tilde{\theta}^{-p}}(\emptyset) &< x(\theta^p) \end{aligned}$$

for all θ^p . For $p = B$ ($p = R$) this induces $\sigma^{\theta^p}(\hat{\theta}^{-p}) = 1$ and $\sigma^{\theta^p}(\tilde{\theta}^{-p}) = 0$ ($\sigma^{\theta^p}(\hat{\theta}^{-p}) = 0$ and $\sigma^{\theta^p}(\tilde{\theta}^{-p}) = 1$) for all $\theta^p \in \Theta^p$. Applying Lemma 2.1 yields

$$\rho^p \left(\underbrace{\sum_{\theta^p} \mu^{\theta^p} \sigma^{\theta^p}(\hat{\theta}^{-p})}_{=1 \text{ (=0)}} - \underbrace{\sum_{\theta^p} \mu^{\theta^p} \sigma^{\theta^p}(\tilde{\theta}^{-p})}_{=0 \text{ (=1)}} \right) = x(\hat{\theta}^{-p}) - x(\tilde{\theta}^{-p}).$$

Taking into account Part (i) this gives

$$\begin{aligned} \rho^p = x(\theta_{max}^{-p}) - x(\theta_{min}^{-p}) &< \min \left\{ x(\theta_{max}^{-p}), 1 - x(\theta_{min}^{-p}) \right\} \\ &\leq \max \left\{ x(\theta_{max}^{-p}), 1 - x(\theta_{min}^{-p}) \right\}. \end{aligned} \quad (2.5)$$

Case 2: $\rho^{-p} > 0$

The reasoning is very similar for both players and w.l.o.g. we consider the case $p = B$. Again, Lemma 2.2 implies that there is a type $\bar{\theta}^B$ and that there are two opponent types $\hat{\theta}^R$ and $\tilde{\theta}^R$ such that

$$\begin{aligned} \alpha + (1 - \alpha)\gamma &= \rho^B \sigma^{\bar{\theta}^B}(\hat{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset), \\ (1 - \alpha)\gamma &= \rho^B \sigma^{\bar{\theta}^B}(\tilde{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset) \end{aligned}$$

with $\alpha = \rho^B$ and $\gamma = \sigma^{\bar{\theta}^B}(\emptyset)$. As we have already seen that $x(\theta_{max}^R) - x(\theta_{min}^R)$ is a lower bound for ρ^B , according to Lemma 2.3 one of the following three subcases must apply:

Subcase 2(a): $\rho^B \sigma^{\hat{\theta}^B}(\hat{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset) = x(\hat{\theta}^R)$ and
 $\rho^B \sigma^{\check{\theta}^B}(\check{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset) = x(\check{\theta}^R)$

This subcase is straightforward. We simply have

$$\rho^B = x(\hat{\theta}^R) - x(\check{\theta}^R) \leq x(\theta_{max}^R) - x(\theta_{min}^R) < \max \{x(\theta_{max}^R), 1 - x(\theta_{min}^R)\}.$$

Subcase 2(b): $\rho^B \sigma^{\hat{\theta}^B}(\hat{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset) > x(\hat{\theta}^R)$

This subcase implies that $\sigma^{\hat{\theta}^R}(\bar{\theta}^B) = 0$. Moreover, by Lemma 2.1 there must exist $\check{\theta}^B \neq \bar{\theta}^B$ such that

$$\rho^B \sigma^{\check{\theta}^B}(\hat{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset) < x(\hat{\theta}^R). \quad (2.6)$$

This induces $\sigma^{\hat{\theta}^R}(\check{\theta}^B) = 1$. Furthermore, according to inequality (2.2a) we have

$$x(\bar{\theta}^B) < \rho^R \underbrace{\sigma^{\hat{\theta}^R}(\bar{\theta}^B)}_{=0} + (1 - \rho^R) \sigma^{\hat{\theta}^R}(\emptyset) = (1 - \rho^R) \sigma^{\hat{\theta}^R}(\emptyset).$$

Applying Lemma 2.3 again – here with $\alpha = \rho^R$, $\beta' = x(\check{\theta}^B)$, $\beta'' = x(\bar{\theta}^B)$, and $\gamma = \sigma^{\hat{\theta}^R}(\emptyset)$ – then gives

$$x(\check{\theta}^B) < \rho^R + (1 - \rho^R) \sigma^{\hat{\theta}^R}(\emptyset) = \rho^R \sigma^{\hat{\theta}^R}(\check{\theta}^B) + (1 - \rho^R) \sigma^{\hat{\theta}^R}(\emptyset). \quad (2.7)$$

In turn, this induces $\sigma^{\check{\theta}^B}(\hat{\theta}^R) = 1$. By inserting the latter into inequality (2.6) we get

$$\rho^B + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset) < x(\hat{\theta}^R),$$

which implies

$$\rho^B < x(\hat{\theta}^R) \leq \max \{x(\theta_{max}^R), 1 - x(\theta_{min}^R)\}.$$

Subcase 2(c): $\rho^B \sigma^{\check{\theta}^B}(\check{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset) < x(\check{\theta}^R)$

This subcase is in large parts quite similar to the previous one. It implies that $\sigma^{\check{\theta}^R}(\bar{\theta}^B) = 1$. Moreover, again by Lemma 2.1 there must exist $\check{\theta}^B \neq \bar{\theta}^B$ such that

$$\rho^B \sigma^{\check{\theta}^B}(\check{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset) > x(\check{\theta}^R). \quad (2.8)$$

This induces $\sigma^{\tilde{\theta}^R}(\check{\theta}^B) = 0$. According to inequality (2.2b) it is

$$x(\bar{\theta}^B) > \rho^R \underbrace{\sigma^{\tilde{\theta}^R}(\bar{\theta}^B)}_{=1} + (1 - \rho^R) \sigma^{\tilde{\theta}^R}(\emptyset) = \rho^R + (1 - \rho^R) \sigma^{\tilde{\theta}^R}(\emptyset).$$

Applying again Lemma 2.3 – here with $\alpha = \rho^R$, $\beta' = x(\bar{\theta}^B)$, $\beta'' = x(\check{\theta}^B)$, and $\gamma = \sigma^{\tilde{\theta}^R}(\emptyset)$ – then gives

$$x(\check{\theta}^B) > (1 - \rho^R) \sigma^{\tilde{\theta}^R}(\emptyset) = \rho^R \sigma^{\tilde{\theta}^R}(\check{\theta}^B) + (1 - \rho^R) \sigma^{\tilde{\theta}^R}(\emptyset). \quad (2.9)$$

In turn, this induces $\sigma^{\check{\theta}^B}(\tilde{\theta}^R) = 0$. By inserting the latter into inequality (2.8) we get

$$\begin{aligned} (1 - \rho^B) \sigma^{\check{\theta}^B}(\emptyset) &> x(\tilde{\theta}^R) \\ \Leftrightarrow \sigma^{\check{\theta}^B}(\emptyset) - x(\tilde{\theta}^R) &> \rho^B \sigma^{\check{\theta}^B}(\emptyset) \\ \Rightarrow \rho^B < 1 - x(\tilde{\theta}^R) &\leq \max \{x(\theta_{max}^R), 1 - x(\theta_{min}^R)\}. \end{aligned}$$

□

2.A.3 An Alternative to Theorem 2.1

Theorem 2.1(ii)'. *Consider a two-action Bayesian conflict game. For all $\epsilon > 0$ there exists $C > 0$ such that for all $p \in \{B, R\}$ and $c \in (0, C)$ we have*

$$\rho^p < \min \{x(\theta_{max}^{-p}), 1 - x(\theta_{min}^{-p})\} + \epsilon$$

in any Bayesian Nash equilibrium.

Proof of Theorem 2.1(ii)'. From inequality (2.5) in Case 1 of the proof of Part (ii) of Theorem 2.1 we already know for $p \in \{B, R\}$ that

$$\rho^p < \min \{x(\theta_{max}^{-p}), 1 - x(\theta_{min}^{-p})\}$$

in any equilibrium with $\rho^{-p} = 0$. Therefore, we only need to consider the case $\rho^{-p} > 0$.

For $p \in \{B, R\}$, $\theta^p \in \Theta^p$ we define

$$\begin{aligned} C^{BH}(\epsilon, \theta^B, \theta^R) &:= u^{\theta^R}(0, x(\theta^R) + \epsilon) - u^{\theta^R}(x(\theta^B), x(\theta^R) + \epsilon), \\ C^{BT}(\epsilon, \theta^B, \theta^R) &:= u^{\theta^R}(1, x(\theta^R) - \epsilon) - u^{\theta^R}(x(\theta^B), x(\theta^R) - \epsilon), \\ C^{RH}(\epsilon, \theta^B, \theta^R) &:= u^{\theta^B}(1, x(\theta^B) + \epsilon) - u^{\theta^B}(x(\theta^R), x(\theta^B) + \epsilon), \text{ and} \end{aligned}$$

$$C^{RT}(\epsilon, \theta^B, \theta^R) := u^{\theta^B}(0, x(\theta^B) - \epsilon) - u^{\theta^B}(x(\theta^R), x(\theta^B) - \epsilon).$$

Notice that, as in the proof of Lemma 2.1, we have $C^B(\epsilon, \theta^B, \theta^R) > 0$ ($C^R(\epsilon, \theta^B, \theta^R) > 0$) for all $\theta^B \in \Theta^B, \theta^R \in \Theta^R$ as player R wants to mis-coordinate (as player B wants to coordinate). Based on this let

$$C(\epsilon) := \min_{a \in \{B_H, B_T, R_H, R_T\}} \min_{\theta^B, \theta^R} \mu^{\theta^B} \mu^{\theta^R} C^a(\epsilon, \theta^B, \theta^R).$$

Now assume that the statement of the theorem does not hold. Then there must exist $c \in (0, C(\epsilon))$ such that

$$(a) \quad \rho^p \geq x(\theta_{max}^{-p}) + \epsilon \text{ or}$$

$$(b) \quad \rho^p \geq 1 - x(\theta_{min}^{-p}) + \epsilon$$

for some $p \in \{B, R\}$ in an equilibrium. Again, the reasoning is almost identical for both players and w.l.o.g. we consider $p = B$.

Case (a): $\rho^B \geq x(\theta_{max}^R) + \epsilon$

As we are in the situation of Lemma 2.2, we know that there exist types $\bar{\theta}^B \in \Theta^B, \hat{\theta}^R \in \Theta^R$ such that $\sigma^{\bar{\theta}^B}(\hat{\theta}^R) = 1$. We then have

$$\rho^B \sigma^{\bar{\theta}^B}(\hat{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset) \geq x(\hat{\theta}^R) + \epsilon > x(\hat{\theta}^R). \quad (2.10)$$

This implies that $\sigma^{\hat{\theta}^R}(\bar{\theta}^B) = 0$ as player R wants to mis-coordinate and as $\rho^R > 0$. Inserting this into inequality (2.2a) gives

$$(1 - \rho^R) \sigma^{\hat{\theta}^R}(\emptyset) > x(\bar{\theta}^B).$$

From this we deduce that $1 - \rho^R > x(\bar{\theta}^B)$ and $\sigma^{\hat{\theta}^R}(\emptyset) > x(\bar{\theta}^B)$. Now consider an alternative strategy for player R with $\check{\rho}^R = 1$ and $\sigma^{\theta^R}(\theta^B)$ a best response for all $\theta^B \in \Theta^B, \theta^R \in \Theta^R$. Taking into account inequality (2.10) we find that by deviating to this strategy player R would obtain an additional payoff of at least

$$\begin{aligned} & (1 - \rho^R) \left(\mu^{\bar{\theta}^B} \mu^{\hat{\theta}^R} \left(u^{\hat{\theta}^R}(0, \rho^B \sigma^{\bar{\theta}^B}(\hat{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset)) \right. \right. \\ & \quad \left. \left. - u^{\hat{\theta}^R}(\sigma^{\hat{\theta}^R}(\emptyset), \rho^B \sigma^{\bar{\theta}^B}(\hat{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset)) \right) - c \right) \\ & \geq (1 - \rho^R) \left(\mu^{\bar{\theta}^B} \mu^{\hat{\theta}^R} \left(u^{\hat{\theta}^R}(0, x(\hat{\theta}^R) + \epsilon) - u^{\hat{\theta}^R}(\sigma^{\hat{\theta}^R}(\emptyset), x(\hat{\theta}^R) + \epsilon) \right) - c \right) \\ & > x(\bar{\theta}^B) \left(\mu^{\bar{\theta}^B} \mu^{\hat{\theta}^R} \left(u^{\hat{\theta}^R}(0, x(\hat{\theta}^R) + \epsilon) - u^{\hat{\theta}^R}(x(\bar{\theta}^B), x(\hat{\theta}^R) + \epsilon) \right) - c \right) > 0 \end{aligned}$$

as we have $c \in (0, C(\epsilon))$. We, thus, arrive at a contradiction.

Case (b): $\rho^B \geq 1 - x(\theta_{min}^R) + \epsilon$

Notice first that this inequality is equivalent to $1 - \rho^B \leq x(\theta_{min}^R) - \epsilon$. The approach here is similar to Case (a). Here, we have from Lemma 2.2 that there exist types $\bar{\theta}^B \in \Theta^B, \tilde{\theta}^R \in \Theta^R$ with $\sigma^{\bar{\theta}^B}(\tilde{\theta}^R) = 0$. This implies

$$\rho^B \sigma^{\bar{\theta}^B}(\tilde{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset) \leq x(\tilde{\theta}^R) - \epsilon < x(\tilde{\theta}^R). \quad (2.11)$$

Hence, we have $\sigma^{\tilde{\theta}^R}(\bar{\theta}^B) = 1$. Inserting this into inequality (2.2b) gives

$$\rho^R + (1 - \rho^R) \sigma^{\tilde{\theta}^R}(\emptyset) < x(\bar{\theta}^B).$$

From this we deduce that $\rho^R < x(\bar{\theta}^B)$ and $\sigma^{\tilde{\theta}^R}(\emptyset) < x(\bar{\theta}^B)$. The former is equivalent to $1 - \rho^R > 1 - x(\bar{\theta}^B)$. Again, we find that player R could improve by deviating to a strategy with probability of empathy acquisition one and obtaining an additional payoff of at least

$$\begin{aligned} & (1 - \rho^R) \left(\mu^{\bar{\theta}^B} \mu^{\tilde{\theta}^R} \left(u^{\tilde{\theta}^R}(1, \rho^B \sigma^{\bar{\theta}^B}(\tilde{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset)) \right. \right. \\ & \quad \left. \left. - u^{\tilde{\theta}^R}(\sigma^{\tilde{\theta}^R}(\emptyset), \rho^B \sigma^{\bar{\theta}^B}(\tilde{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset)) \right) - c \right) \\ & \geq (1 - \rho^R) \left(\mu^{\bar{\theta}^B} \mu^{\tilde{\theta}^R} \left(u^{\tilde{\theta}^R}(1, x(\tilde{\theta}^R) - \epsilon) - u^{\tilde{\theta}^R}(\sigma^{\tilde{\theta}^R}(\emptyset), x(\tilde{\theta}^R) - \epsilon) \right) - c \right) \\ & > (1 - x(\bar{\theta}^B)) \left(\mu^{\bar{\theta}^B} \mu^{\tilde{\theta}^R} \left(u^{\tilde{\theta}^R}(1, x(\tilde{\theta}^R) - \epsilon) - u^{\tilde{\theta}^R}(x(\bar{\theta}^B), x(\tilde{\theta}^R) - \epsilon) \right) - c \right) > 0 \end{aligned}$$

as we have $c \in (0, C(\epsilon))$. This is again a contradiction. \square

2.A.4 Proof of Proposition 2.2

The proof is by construction. We identify a particular strategy profile $(\rho^p, (\sigma^{\theta^p})_{\theta^p \in \Theta^p})_{p \in \{B, R\}}$ with the desired property and show that it is an equilibrium. Let

$$\rho^p = x(\theta_{max}^{-p}) - x(\theta_{min}^{-p}), \quad (2.12a)$$

$$\sigma^{\theta^p}(\emptyset) = \frac{1}{1 - \rho^p} x(\theta_{min}^{-p}) \quad \forall \theta^p \in \Theta^p \setminus \{\theta_{max}^p, \theta_{min}^p\} \quad (2.12b)$$

$$\sigma^{\theta^p}(\theta^{-p}) = \begin{cases} \frac{1}{\rho^p} (x(\theta^{-p}) - x(\theta_{min}^{-p})) & \text{if } \rho^p > 0 \\ 0 & \text{if } \rho^p = 0 \end{cases} \quad \forall \theta^p \in \Theta^p, \theta^{-p} \in \Theta^{-p}. \quad (2.12c)$$

Note that $\sigma^{\theta^p}(\theta_{max}^{-p}) = 1$ and $\sigma^{\theta^p}(\theta_{min}^{-p}) = 0$ for all $p \in \{B, R\}$ and $\theta^p \in \Theta^p$ if $x(\theta_{max}^{-p}) > x(\theta_{min}^{-p})$.³⁵ The strategy profile is, thus, almost fully specified. It only remains to be chosen how extreme types play when they are uninformed. In case that $x(\theta_{max}^p) > x(\theta_{min}^p)$ let $\sigma^{\theta_{max}^p}(\emptyset)$ and $\sigma^{\theta_{min}^p}(\emptyset)$ be chosen to satisfy

$$\sum_{\theta^{-p}} \mu^{\theta^{-p}} \left(u^{\theta^{-p}}(1, \rho^p \sigma^{\theta_{max}^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta_{max}^p}(\emptyset)) - u^{\theta^{-p}}(0, \rho^p \sigma^{\theta_{max}^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta_{max}^p}(\emptyset)) \right) = \frac{c}{\mu^{\theta_{max}^p}} \quad (2.13)$$

and

$$\frac{\mu^{\theta_{max}^p}}{\mu^{\theta_{max}^p} + \mu^{\theta_{min}^p}} \sigma^{\theta_{max}^p}(\emptyset) + \frac{\mu^{\theta_{min}^p}}{\mu^{\theta_{max}^p} + \mu^{\theta_{min}^p}} \sigma^{\theta_{min}^p}(\emptyset) = \frac{1}{1 - \rho^p} x(\theta_{min}^{-p}). \quad (2.14)$$

For $x(\theta_{max}^p) = x(\theta_{min}^p)$ however let

$$\sigma^{\theta_{max}^p}(\emptyset) = \sigma^{\theta_{min}^p}(\emptyset) = \frac{1}{1 - \rho^p} x(\theta_{min}^{-p}). \quad (2.15)$$

For the remainder of the proof we distinguish these two cases.

Case 1: $x(\theta_{max}^p) > x(\theta_{min}^p)$

Before we move on to prove that the considered strategy profile is indeed an equilibrium in this case, we need to make sure that it is well-defined. For this we need to show that equation (2.13) has a feasible solution for $c = 0$ and $c > 0$ sufficiently small. Consider

$$\sigma^{\theta_{max}^p}(\emptyset) = \frac{1}{1 - \rho^p} \left(x(\theta_{min}^{-p}) + \epsilon^p \right) = \frac{x(\theta_{min}^{-p}) + \epsilon^p}{1 - x(\theta_{max}^{-p}) + x(\theta_{min}^{-p})},$$

where $\epsilon^p \in \mathbb{R}$. For $c = 0$ let $\epsilon^p = 0$. We then have $\sigma^{\theta_{max}^p}(\emptyset) \in (0, 1)$ and

$$\text{LHS of (2.13)} = \sum_{\theta^{-p}} \mu^{\theta^{-p}} \left(u^{\theta^{-p}}(1, x(\theta^{-p})) - u^{\theta^{-p}}(0, x(\theta^{-p})) \right) = 0 = \text{RHS of (2.13)}$$

since player $-p$ of type θ^{-p} is indifferent between both actions if the opponent plays $x(\theta^{-p})$. Equation (2.14) then implies $\sigma^{\theta_{min}^p}(\emptyset) = \sigma^{\theta_{max}^p}(\emptyset)$.

Now consider $c > 0$. Notice first that the left-hand side of (2.13) is a linear function in ϵ^p which is strictly decreasing (increasing) for $p = B$ ($p = R$). To see this, consider temporarily and w.l.o.g. $-p = B$ and some type θ^B whose payoffs are

³⁵This means that in case that they are informed, both players of any type play pure action strategies against extreme type opponents.

represented by the matrix

	H	T
H	$u_{H,H}$	$u_{H,T}$
T	$u_{T,H}$	$u_{T,T}$

where $u_{H,H}, u_{H,T}, u_{T,H}, u_{T,T} \in \mathbb{R}$. As player B wants to coordinate actions, we must have $u_{H,H} > u_{T,H}$ and $u_{T,T} > u_{H,T}$. Further, we calculate $x(\theta^B) = \frac{u_{T,T} - u_{H,T}}{u_{H,H} - u_{T,H} + u_{T,T} - u_{H,T}}$. Our claim follows immediately as this gives

$$\begin{aligned}
 & u^{\theta^B}(1, x(\theta^B) + \epsilon^R) - u^{\theta^B}(0, x(\theta^B) + \epsilon^R) \\
 &= u_{H,H}(x(\theta^B) + \epsilon^R) + u_{H,T}(1 - x(\theta^B) - \epsilon^R) - u_{T,H}(x(\theta^B) + \epsilon^R) \\
 &\quad - u_{T,T}(1 - x(\theta^B) - \epsilon^R) \\
 &= (u_{H,H} - u_{T,H} + u_{T,T} - u_{H,T})(x(\theta^B) + \epsilon^R) - (u_{T,T} - u_{H,T}) \\
 &= \underbrace{(u_{H,H} - u_{T,H} + u_{T,T} - u_{H,T})}_{>0} \epsilon^R.
 \end{aligned}$$

So, generally speaking, we have that for every $c > 0$ sufficiently small there exists a unique $\epsilon^B < 0$ ($\epsilon^R > 0$) such that both equations (2.13) and (2.14) are fulfilled and $\sigma^{\theta^B_{max}}(\emptyset), \sigma^{\theta^B_{min}}(\emptyset) \in [0, 1]$.

After this is done we turn towards proving that the proposed strategy profile is indeed an equilibrium. Suppose that in the conflict game both players B and R are playing a strategy as considered above. Then player $-p \in \{B, R\}$ cannot improve by deviating if the following conditions are satisfied:

- $\sigma^{\theta^{-p}}(\theta^p)$ is a best response to $\rho^p \sigma^{\theta^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta^p}(\emptyset)$ for all $\theta^p \in \Theta^p, \theta^{-p} \in \Theta^{-p}$,
- $\sigma^{\theta^{-p}}(\emptyset)$ is a best response to $\sum_{\theta^p} \mu^{\theta^p} (\rho^p \sigma^{\theta^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta^p}(\emptyset))$ for all $\theta^{-p} \in \Theta^{-p}$,
- $\sum_{\theta^{-p}} \mu^{\theta^{-p}} \mathcal{U}_{Info}^{\theta^{-p}} = \sum_{\theta^{-p}} \mu^{\theta^{-p}} \mathcal{U}_N^{\theta^{-p}} + c$.

In the following let $c = 0$ or $c > 0$ sufficiently small as mentioned above. Further let $p = B$ ($p = R$). Consider first the action strategies that types of player $-p$ face when they are informed. We calculate for $\theta^{-p} \in \Theta^{-p}, \theta^p \in \Theta^p \setminus \{\theta^p_{max}, \theta^p_{min}\}$:

$$\begin{aligned}
 \rho^p \sigma^{\theta^p_{max}}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta^p_{max}}(\emptyset) &= x(\theta^{-p}) + \epsilon^p && \leq (\geq) x(\theta^{-p}), \\
 \rho^p \sigma^{\theta^p_{min}}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta^p_{min}}(\emptyset) &= x(\theta^{-p}) - \frac{\mu^{\theta^p_{max}}}{\mu^{\theta^p_{min}}} \epsilon^p && \geq (\leq) x(\theta^{-p}), \\
 \rho^p \sigma^{\theta^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta^p}(\emptyset) &&& = x(\theta^{-p}).
 \end{aligned} \tag{2.16}$$

Hence, $\sigma^{\theta^{-p}}(\theta_{max}^p) = 1$ and $\sigma^{\theta^{-p}}(\theta_{min}^p) = 0$ are indeed best responses in each case $p \in \{B, R\}$, $\theta^{-p} \in \Theta^{-p}$. Against all other types $\theta^p \in \Theta^p \setminus \{\theta_{max}^p, \theta_{min}^p\}$, any informed type $\theta^{-p} \in \Theta^{-p}$ is indifferent between both actions.

Beyond that, any uninformed player type $\theta^{-p} \in \Theta^{-p}$ faces

$$\begin{aligned} & \sum_{\theta^p} \mu^{\theta^p} \left(\rho^p \sigma^{\theta^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta^p}(\emptyset) \right) \\ &= \mu^{\theta_{max}^p} \left(x(\theta^{-p}) + \epsilon^p \right) + \mu^{\theta_{min}^p} \left(x(\theta^{-p}) - \frac{\mu^{\theta_{max}^p}}{\mu^{\theta_{min}^p}} \epsilon^p \right) + \sum_{\theta^p \notin \{\theta_{max}^p, \theta_{min}^p\}} \mu^{\theta^p} x(\theta^{-p}) \\ &= x(\theta^{-p}) \end{aligned}$$

and is therefore indifferent between both actions.

Finally, we have to examine the expected payoffs. For an uninformed player type $\theta^{-p} \in \Theta^{-p}$ we have

$$\begin{aligned} \mathcal{U}_N^{\theta^{-p}} &= u^{\theta^{-p}} \left(\sigma^{\theta^{-p}}(\emptyset), \sum_{\theta^p} \mu^{\theta^p} \left(\rho^p \sigma^{\theta^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta^p}(\emptyset) \right) \right) \\ &= u^{\theta^{-p}} \left(\sigma^{\theta^{-p}}(\emptyset), x(\theta^{-p}) \right). \end{aligned}$$

If θ^{-p} is informed, then her expected payoff (ex costs) is given by

$$\begin{aligned} \mathcal{U}_{Info}^{\theta^{-p}} &= \sum_{\theta^p} \mu^{\theta^p} u^{\theta^{-p}} \left(\sigma^{\theta^{-p}}(\theta^p), \rho^p \sigma^{\theta^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta^p}(\emptyset) \right) \\ &= \mu^{\theta_{max}^p} u^{\theta^{-p}} \left(1, \rho^p \sigma^{\theta_{max}^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta_{max}^p}(\emptyset) \right) \\ &\quad + \mu^{\theta_{min}^p} u^{\theta^{-p}} \left(0, \frac{1}{\mu^{\theta_{min}^p}} \left(x(\theta^{-p}) - \sum_{\theta^p \neq \theta_{min}^p} \mu^{\theta^p} \left(\rho^p \sigma^{\theta^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta^p}(\emptyset) \right) \right) \right) \\ &\quad + \sum_{\theta^p \notin \{\theta_{max}^p, \theta_{min}^p\}} \mu^{\theta^p} u^{\theta^{-p}} \left(\sigma^{\theta^{-p}}(\theta^p), x(\theta^{-p}) \right) \\ &= \mu^{\theta_{max}^p} \left(u^{\theta^{-p}} \left(1, \rho^p \sigma^{\theta_{max}^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta_{max}^p}(\emptyset) \right) \right. \\ &\quad \left. - u^{\theta^{-p}} \left(0, \rho^p \sigma^{\theta_{max}^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta_{max}^p}(\emptyset) \right) \right) + u^{\theta^{-p}} \left(0, x(\theta^{-p}) \right) \end{aligned}$$

Notice that according to (2.16) we have $\mathcal{U}_{Info}^{\theta^{-p}} \geq u^{\theta^{-p}} \left(0, x(\theta^{-p}) \right) = \mathcal{U}_N^{\theta^{-p}}$ for all $\theta^{-p} \in \Theta^{-p}$. Taken together we get

$$(2.13) \Leftrightarrow \sum_{\theta^{-p}} \mu^{\theta^{-p}} \left(\mu^{\theta_{max}^p} \left(u^{\theta^{-p}} \left(1, \rho^p \sigma^{\theta_{max}^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta_{max}^p}(\emptyset) \right) \right) \right)$$

$$\begin{aligned}
 & - u^{\theta^{-p}} \left(0, \rho^p \sigma^{\theta_{max}^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta_{max}^p}(\emptyset) \right) + u^{\theta^{-p}} \left(0, x(\theta^{-p}) \right) \\
 & = \sum_{\theta^{-p}} \mu^{\theta^{-p}} u^{\theta^{-p}} \left(\sigma^{\theta^{-p}}(\emptyset), x(\theta^{-p}) \right) + c \\
 & \Leftrightarrow \sum_{\theta^{-p}} \mu^{\theta^{-p}} \mathcal{U}_{Info}^{\theta^{-p}} = \sum_{\theta^{-p}} \mu^{\theta^{-p}} \mathcal{U}_N^{\theta^{-p}} + c.
 \end{aligned}$$

This means that player $-p$ is indeed indifferent between acquiring empathy and not acquiring it. Thus, we established for the conflict game that player $-p$ has no incentives to deviate from the considered strategy in this case.

Case 2: $x(\theta_{max}^p) = x(\theta_{min}^p)$

Suppose again that both players B and R are playing a strategy as considered above. As according to equation (2.12a) we have $\rho^{-p} = 0$ in this case, player $-p$ cannot improve by deviating if the following conditions are satisfied:

- $\sigma^{\theta^{-p}}(\emptyset)$ is a best response to $\sum_{\theta^p} \mu^{\theta^p} \left(\rho^p \sigma^{\theta^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta^p}(\emptyset) \right)$ for all $\theta^{-p} \in \Theta^{-p}$,
- $\sum_{\theta^{-p}, \theta^p} \mu^{\theta^{-p}} \mu^{\theta^p} u^{\theta^{-p}} \left(s^{\theta^{-p}}(\theta^p), \rho^p \sigma^{\theta^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta^p}(\emptyset) \right) \leq \sum_{\theta^{-p}} \mu^{\theta^{-p}} \mathcal{U}_N^{\theta^{-p}} + c$ for all $\left(s^{\theta^{-p}}(\theta^p) \right)_{\theta^{-p}, \theta^p} \in \Delta(A)^{n^{-p} \times n^p}$.

Taking into account equations (2.12) and (2.15), concerning the first condition we simply have

$$\sum_{\theta^p} \mu^{\theta^p} \left(\rho^p \sigma^{\theta^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta^p}(\emptyset) \right) = \sum_{\theta^p} \mu^{\theta^p} x(\theta^{-p}) = x(\theta^{-p}).$$

Hence, this condition is obviously fulfilled as any uninformed type θ^{-p} is indifferent between both actions.

The second condition states that the ex ante expected payoff of player $-p$ from not acquiring empathy must be greater than or equal to the maximal payoff (minus costs) she could get instead from acquiring empathy and playing freely choosable action strategies which can be conditioned on the opponent's type. For all θ^{-p} , $\left(s^{\theta^{-p}}(\theta^p) \right)_{\theta^p}$ we have

$$\sum_{\theta^p} \mu^{\theta^p} u^{\theta^{-p}} \left(s^{\theta^{-p}}(\theta^p), \underbrace{\rho^p \sigma^{\theta^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta^p}(\emptyset)}_{=x(\theta^{-p})} \right) = u^{\theta^{-p}} \left(\cdot, x(\theta^{-p}) \right).$$

On the contrary, type θ^{-p} receives

$$\mathcal{U}_N^{\theta^{-p}} = u^{\theta^{-p}} \left(\sigma^{\theta^{-p}}(\emptyset), x(\theta^{-p}) \right)$$

if she is uninformed. Thus, we have

$$\sum_{\theta^{-p}, \theta^p} \mu^{\theta^{-p}} \mu^{\theta^p} u^{\theta^{-p}} \left(s^{\theta^{-p}}(\theta^p), \rho^p \sigma^{\theta^p}(\theta^{-p}) + (1 - \rho^p) \sigma^{\theta^p}(\emptyset) \right) = \sum_{\theta^{-p}} \mu^{\theta^{-p}} \mathcal{U}_N^{\theta^{-p}}$$

for all $(s^{\theta^{-p}}(\theta^p))_{\theta^{-p}, \theta^p} \in \Delta(A)^{n^{-p} \times n^p}$. This concludes Case 2 and the proof as a whole. \square

2.A.5 Proof of Proposition 2.3

Recall the proof of Theorem 2.1. In Case 1 of Part (ii) we already established that we must have

$$\rho^p = x(\theta_{max}^{-p}) - x(\theta_{min}^{-p}) \quad (2.17)$$

if $\rho^{-p} = 0$ for $p \in \{B, R\}$. Notice that $\rho^p > 0$ then implies $x(\theta_{max}^{-p}) > x(\theta_{min}^{-p})$. Taking into account Theorem 2.1(i) in this situation we also have that

$$0 = \rho^{-p} \geq x(\theta_{max}^p) - x(\theta_{min}^p) \geq 0,$$

and thus $\rho^{-p} = x(\theta_{max}^p) - x(\theta_{min}^p)$. In what follows we distinguish the two cases considered in the proposition.

Part 1: $n^B = 1$ and $n^R > 1$ ($n^B > 1$ and $n^R = 1$, respectively)

W.l.o.g. consider the case $n^B = 1$ (such that $\Theta^B = \{\theta^B\}$) and $n^R > 1$ and let $c \in (0, C)$ sufficiently small. Assume that $\rho^R > 0$ in an equilibrium. Then according to Lemma 2.2 there must exist $\bar{\theta}^R$ and $\hat{\theta}^B, \tilde{\theta}^B$ fulfilling inequalities (2.2). This however implies $\hat{\theta}^B \neq \tilde{\theta}^B$ which is a contradiction as we have $n^B = 1$. Thus, we must have $\rho^R = 0$ which (together with the above considerations) establishes uniqueness of the empathy levels for this part of the proof.

By assumption we have that $x(\theta_{max}^R) > x(\theta^R) > x(\theta_{min}^R)$ for all $\theta^R \in \Theta^R \setminus \{\theta_{max}^R, \theta_{min}^R\}$. We now show that the equilibrium considered in the proof of Proposition 2.2 is unique up to variations of the action strategies $\sigma^{\theta^R}(\theta^B)$ which are played with probability $\rho^R = 0$. Notice first that according to Lemma 2.1 we must have

$$\rho^B \sigma^{\theta^B}(\theta^R) + (1 - \rho^B) \sigma^{\theta^B}(\emptyset) = x(\theta^R) \quad (2.18)$$

for all $\theta^R \in \Theta^R$. Taking into account equation (2.17) this gives

$$\begin{aligned} (x(\theta_{max}^R) - x(\theta_{min}^R))(\sigma^{\theta^B}(\theta_{max}^R) - \sigma^{\theta^B}(\theta_{min}^R)) &= \rho^B(\sigma^{\theta^B}(\theta_{max}^R) - \sigma^{\theta^B}(\theta_{min}^R)) \\ &= x(\theta_{max}^R) - x(\theta_{min}^R). \end{aligned}$$

Hence, we must have $\sigma^{\theta^B}(\theta_{max}^R) = 1$ and $\sigma^{\theta^B}(\theta_{min}^R) = 0$. Again according to Lemma 2.1 this implies that

$$\sigma^{\theta^B}(\theta^R) = \sigma^{\theta^B}(\theta^R) - \sigma^{\theta^B}(\theta_{min}^R) = \frac{1}{\rho^B}(x(\theta^R) - x(\theta_{min}^R)) \Leftrightarrow (2.12c)$$

for all $\theta^R \in \Theta^R$. Moreover, by equation (2.18) this induces

$$\sigma^{\theta^B}(\emptyset) = \frac{1}{1 - \rho^B}x(\theta_{min}^R) \Leftrightarrow (2.12b).$$

As $x(\theta_{max}^R) > x(\theta^R) > x(\theta_{min}^R)$ we have $\sigma^{\theta^B}(\theta^R) \in (0, 1)$ for all $\theta^R \in \Theta^R \setminus \{\theta_{max}^R, \theta_{min}^R\}$. This means that θ^B must be indifferent against any opponent type $\theta^R \in \Theta^R \setminus \{\theta_{max}^R, \theta_{min}^R\}$ if she is informed. Thus, we must have

$$\sigma^{\theta^R}(\emptyset) = x(\theta^B)$$

for all $\theta^R \in \Theta^R \setminus \{\theta_{max}^R, \theta_{min}^R\}$. Equation (2.1) of Lemma 2.1 then transforms to

$$\begin{aligned} \sum_{\theta^R} \mu^{\theta^R} \sigma^{\theta^R}(\emptyset) &= x(\theta^B) \\ \Leftrightarrow \frac{\mu^{\theta_{max}^R}}{\mu^{\theta_{max}^R} + \mu^{\theta_{min}^R}} \sigma^{\theta_{max}^R}(\emptyset) + \frac{\mu^{\theta_{min}^R}}{\mu^{\theta_{max}^R} + \mu^{\theta_{min}^R}} \sigma^{\theta_{min}^R}(\emptyset) &= x(\theta^B) \Leftrightarrow (2.14). \end{aligned}$$

Together with equation (2.13) (for $p = R$) this then uniquely determines $\sigma^{\theta_{max}^R}(\emptyset)$ and $\sigma^{\theta_{min}^R}(\emptyset)$.

Obviously, the reasoning is the same for $n^B > 1, n^R = 1$.

Part 2: $n^B = n^R = 2$

In this case we have $\Theta^p = \{\theta_{max}^p, \theta_{min}^p\}$ for $p = B, R$. We already know that uniqueness of the empathy levels follows immediately if we have $\rho^p = 0$ for some $p \in \{B, R\}$. So in this regard we only need to consider the case that $\rho^B, \rho^R > 0$. Again, we recall the proof of Theorem 2.1 and take Case 2 with $p = B$ as a starting point. Consider its three subcases.

Subcase (a): $\rho^B \sigma^{\bar{\theta}^B}(\hat{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset) = x(\hat{\theta}^R)$ and
 $\rho^B \sigma^{\bar{\theta}^B}(\tilde{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset) = x(\tilde{\theta}^R)$

This subcase is again straightforward as we simply have $\rho^B = x(\hat{\theta}^R) - x(\tilde{\theta}^R)$ and know already that it is $\rho^B \geq x(\theta_{max}^R) - x(\theta_{min}^R)$. Therefore it must be $\hat{\theta}^R = \theta_{max}^R$ and $\tilde{\theta}^R = \theta_{min}^R$.

Subcase (b): $\rho^B \sigma^{\bar{\theta}^B}(\hat{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset) > x(\hat{\theta}^R)$

Recall inequality (2.7). Lemma 2.1 then implies that

$$\rho^R \sigma^{\check{\theta}^R}(\check{\theta}^B) + (1 - \rho^R) \sigma^{\check{\theta}^R}(\emptyset) < x(\check{\theta}^B)$$

as here it is $\{\theta^R \in \Theta^R \mid \theta^R \neq \hat{\theta}^R\} = \{\tilde{\theta}^R\}$. In turn, this induces $\sigma^{\check{\theta}^B}(\tilde{\theta}^R) = 0$. Moreover, recall that it is $\sigma^{\bar{\theta}^B}(\tilde{\theta}^R) = 0$, $\sigma^{\bar{\theta}^B}(\hat{\theta}^R) = 1$ and $\sigma^{\check{\theta}^B}(\hat{\theta}^R) = 1$. Further, we know again by Lemma 2.1 that it must be

$$\rho^B \left(\sum_{\theta^B} \mu^{\theta^B} \sigma^{\theta^B}(\theta_{max}^R) - \sum_{\theta^B} \mu^{\theta^B} \sigma^{\theta^B}(\theta_{min}^R) \right) = x(\theta_{max}^R) - x(\theta_{min}^R).$$

If it were $\hat{\theta}^R = \theta_{min}^R$, $\tilde{\theta}^R = \theta_{max}^R$, then this would imply $\rho^B = x(\theta_{min}^R) - x(\theta_{max}^R) \leq 0$. So it must be $\hat{\theta}^R = \theta_{max}^R$, $\tilde{\theta}^R = \theta_{min}^R$ which implies $\rho^B = x(\theta_{max}^R) - x(\theta_{min}^R)$.

Subcase (c): $\rho^B \sigma^{\bar{\theta}^B}(\tilde{\theta}^R) + (1 - \rho^B) \sigma^{\bar{\theta}^B}(\emptyset) < x(\tilde{\theta}^R)$

The procedure is very similar to Subcase (b). By recalling (2.9) and applying Lemma 2.1 we can show that here it is $\sigma^{\check{\theta}^B}(\hat{\theta}^R) = 1$. Together with $\sigma^{\bar{\theta}^B}(\hat{\theta}^R) = 1$, $\sigma^{\bar{\theta}^B}(\tilde{\theta}^R) = 0$ and $\sigma^{\check{\theta}^B}(\tilde{\theta}^R) = 0$ this leads to the same result as in Subcase (b).

Obviously, by choosing $p = R$ one can show in a very similar way that $\rho^R = x(\theta_{max}^B) - x(\theta_{min}^B)$ is satisfied as well in each case. Also, we get that $\sigma^{\theta^R}(\theta_{max}^B) = 1$ and $\sigma^{\theta^R}(\theta_{min}^B) = 0$ for all $\theta^R \in \Theta^R$.

It remains to show that the equilibrium considered in the proof of Proposition 2.2 is indeed unique as long as $x(\theta_{max}^{-p}) > x(\theta_{min}^{-p})$ for both $p \in \{B, R\}$. Notice that we therefore only need to consider the case $\rho^B, \rho^R > 0$. So far, we found that the equations (2.12) must necessarily be satisfied here. If we insert $\sigma^{\theta^{-p}}(\theta_{min}^p) = 0$, $\theta^{-p} \in \Theta^{-p}$, into the indifference condition (2.1) with $\theta^p = \theta_{min}^p$ which we established in Lemma 2.1, we end up with equation (2.14). Finally, both players $p \in \{B, R\}$ must be indifferent between acquiring empathy and not acquiring it as we have $\rho^B, \rho^R > 0$.

Therefore, we must also have

$$\sum_{\theta^p} \mu^{\theta^p} \mathcal{U}_{Info}^{\theta^p} = \sum_{\theta^p} \mu^{\theta^p} \mathcal{U}_N^{\theta^p} + c \Leftrightarrow (2.13)$$

as in Case 1 of the proof of Proposition 2.2. Again, these conditions then uniquely determine $\sigma^{\theta_{max}^B}(\emptyset)$, $\sigma^{\theta_{min}^B}(\emptyset)$, $\sigma^{\theta_{max}^R}(\emptyset)$, and $\sigma^{\theta_{min}^R}(\emptyset)$. This concludes this part of the proof. \square

Appendix 2.B A Continuum of Bayesian Nash Equilibria

In the proof of Proposition 2.2 we construct a certain equilibrium for general two-action conflict games with small costs. Here, we show, however, that such a partial empathy equilibrium can in general be constructed in many different ways. In fact, this gives rise to a continuum of equilibria which have in common that they achieve the lower bound of empathy acquisition established in Theorem 2.1. Thus, they do not differ in terms of players' information strategies but in their action strategies.

In a partial empathy equilibrium in which we have $\rho^p > 0$ for some player $p \in \{B, R\}$, i.e. p is mixing properly between acquiring empathy and not acquiring it, this player must be indifferent between these two possibilities. As she has to bear the cost for acquiring empathy, the (weighted) sum of the expected payoffs of her types being informed must be larger than in the case when they are uninformed. In the proof of Proposition 2.2 this is achieved by letting extreme type opponents, when uninformed, play in such a way that the considered player is indeed indifferent between acquiring empathy and not acquiring it. In this equilibrium, moreover, all types of the opponent $-p$, when informed and facing an extreme type θ_{max}^p or θ_{min}^p , play pure action strategies. In general, i.e. if player p has more than two types, this leaves open the possibility to redesign strategies in such a way that the extreme type opponents, now however the ones being informed and playing against a non-extreme type of the player, are (still) indifferent between both actions and then choose their (mixed) actions such that the player is again indifferent between acquiring empathy and not acquiring it. In doing so, it is however important to retain that all player types, when uninformed, are indifferent between both actions (see Lemma 2.1). With this idea in mind (and given that costs are positive and at least one player has more than two types) we are now able to construct a continuum of partial empathy equilibria.

To ease notation in the following, at least to some extent, we define

$$\lambda^p := \frac{\mu^{\theta_{max}^p}}{\mu^{\theta_{max}^p} + \mu^{\theta_{min}^p}}$$

for $p \in \{B, R\}$. In principle, the following statement is then a corollary to the proof of Proposition 2.2.

Corollary 2.3. *Consider a two-action Bayesian conflict game. There exists $C > 0$ such that all strategy profiles satisfying the following conditions for all $p \in \{B, R\}$ constitute a Bayesian Nash equilibrium of the game if $c \in [0, C]$.³⁶*

- $\rho^p = x(\theta_{max}^{-p}) - x(\theta_{min}^{-p})$,
- $\sigma^{\theta^p}(\theta_{max}^{-p}) = 1$ and $\sigma^{\theta^p}(\theta_{min}^{-p}) = 0$ for all $\theta^p \in \Theta^p$,
- $\sigma^{\theta^p}(\theta^{-p}) = \frac{1}{\rho^p}(x(\theta^{-p}) - x(\theta_{min}^{-p}))$ for all $\theta^p \in \Theta^p \setminus \{\theta_{max}^p, \theta_{min}^p\}$, $\theta^{-p} \in \Theta^{-p} \setminus \{\theta_{max}^{-p}, \theta_{min}^{-p}\}$,
- $\sigma^{\theta^p}(\emptyset) = \frac{1}{1-\rho^p}x(\theta_{min}^{-p})$ for all $\theta^p \in \Theta^p \setminus \{\theta_{max}^p, \theta_{min}^p\}$.

Further, $\sigma^{\theta_{max}^p}$ and $\sigma^{\theta_{min}^p}$ are determined such that on the one hand

- $\lambda^p \sigma^{\theta_{max}^p}(\theta^{-p}) + (1 - \lambda^p) \sigma^{\theta_{min}^p}(\theta^{-p}) = \frac{1}{\rho^p}(x(\theta^{-p}) - x(\theta_{min}^{-p}))$ for all $\theta^{-p} \in \Theta^{-p} \setminus \{\theta_{max}^{-p}, \theta_{min}^{-p}\}$,
- $\lambda^p \sigma^{\theta_{max}^p}(\emptyset) + (1 - \lambda^p) \sigma^{\theta_{min}^p}(\emptyset) = \frac{1}{1-\rho^p}x(\theta_{min}^{-p})$,

where $\sigma^{\theta_{max}^p}(\theta^{-p}) \leq (\geq) \frac{1}{\rho^p}(x(\theta^{-p}) - x(\theta_{min}^{-p})) \leq (\geq) \sigma^{\theta_{min}^p}(\theta^{-p})$ and $\sigma^{\theta_{max}^p}(\emptyset) \leq (\geq) \frac{1}{1-\rho^p}x(\theta_{min}^{-p}) \leq (\geq) \sigma^{\theta_{min}^p}(\emptyset)$ for $p = B$ ($p = R$), and on the other hand

- $\sum_{\theta^{-p}} \mu^{\theta^{-p}} \mathcal{U}_{Info}^{\theta^{-p}} = \sum_{\theta^{-p}} \mu^{\theta^{-p}} \mathcal{U}_N^{\theta^{-p}} + c.$

Note that, in general, there exist (infinitely many) different strategy profiles fulfilling these requirements. However, for $c = 0$ this continuum collapses to a single strategy profile, namely the one where we have $\sigma^{\theta_{max}^p}(\theta^{-p}) = \sigma^{\theta_{min}^p}(\theta^{-p})$ and $\sigma^{\theta_{max}^p}(\emptyset) = \sigma^{\theta_{min}^p}(\emptyset)$ for all $p \in \{B, R\}$ and all $\theta^{-p} \in \Theta^{-p}$.

³⁶For ease of notation we assume here that it is $|\{\arg \max_{\theta^p} x(\theta^p)\}| = |\{\arg \min_{\theta^p} x(\theta^p)\}| = 1$ and $n^p > 1$ for all $p \in \{B, R\}$.

Appendix 2.C On the Uniqueness of Completely Mixed Nash Equilibria

In this chapter, we consider a certain class of games with incomplete information. Within this class we focus on what we call conflict games. By definition these games are such that any realized type game, which is in each case simply a complete information normal form game with two players, has a unique Nash equilibrium and that Nash equilibrium must be in completely mixed strategies. Throughout the chapter, we consider all types of both players having the same finite action set A available. One could therefore wish for a generalization of our model and results that allows two opposing types to possibly choose their (mixed) actions from action sets with different cardinalities. The following remark, however, establishes that in any realized type game of a conflict game the action sets of the opposing types necessarily have to have the same cardinality. Given this finding, it is of course w.l.o.g. to assume that, in a conflict game, types of both players have the same action set to choose from.

Remark 2.2. *Consider a complete information normal form game $\Gamma = (\{B, R\}, \Delta(A^B) \times \Delta(A^R), (u^B, u^R))$ such that for the finite action sets it is $|A^B| \neq |A^R|$. If this game has a completely mixed Nash equilibrium, then this Nash equilibrium is not unique. In fact, this game then even has a continuum of Nash equilibria.*

Proof of Remark 2.2. First, we introduce some notation. Let $A^B := \{a_1^B, \dots, a_{m^B}^B\}$ and $A^R := \{a_1^R, \dots, a_{m^R}^R\}$ and assume w.l.o.g. that $m^B < m^R$. Let $U^B := (b_i^j)_{i=1, \dots, m^B; j=1, \dots, m^R}$ with $b_i^j := u_B(a_i^B, a_j^R)$ denote the payoff matrix of player B and let $b^j := (b_1^j, \dots, b_{m^B}^j)$. Using this notation, it is $\text{rank } U^B \leq m^B < m^R$, meaning that the columns of U^B are not linearly independent, that is there exists $(\lambda_1, \dots, \lambda_{m^R}) \in \mathbb{R}^{m^R} \setminus \{0\}$ such that

$$\sum_{j=1}^{m^R} \lambda_j b^j = 0. \quad (2.19)$$

By assumption the game Γ has a completely mixed Nash equilibrium $(\alpha^B, \alpha^R) \in \Delta(A^B) \times \Delta(A^R)$. Then we can distinguish two cases: Either $(\lambda_1, \dots, \lambda_{m^R})$ can be scaled such that $\lambda_j \leq \alpha_j^R$ for all $j \in \{1, \dots, m^R\}$ and $\lambda_k = \alpha_k^R$, $\lambda_l < \alpha_l^R$ for some $k, l \in \{1, \dots, m^R\}$ or otherwise it must be $(\lambda_1, \dots, \lambda_{m^R}) = \bar{\gamma} \alpha^R$ for some $\bar{\gamma} \in \mathbb{R} \setminus \{0\}$.

Consider the first of the two cases and let $(\lambda_1, \dots, \lambda_{m^R})$ be scaled as mentioned. Since the Nash equilibrium is assumed to be completely mixed, player B must be

indifferent between all of her actions, meaning that there exists $\gamma \in \mathbb{R}$ such that $\sum_{j=1}^{m^R} \alpha_j^R b^j = \gamma(1, \dots, 1)$. Taking into account (2.19), we also have

$$\sum_{j=1}^{m^R} \alpha_j^R b^j = \sum_{j=1}^{m^R} (\alpha_j^R - \lambda_j) b^j.$$

We define $\tilde{\gamma} := \sum_{j=1}^{m^R} (\alpha_j^R - \lambda_j)$. Then it is $\tilde{\gamma} > 0$ and

$$\sum_{j=1}^{m^R} \frac{\alpha_j^R - \lambda_j}{\tilde{\gamma}} b^j = \frac{\gamma}{\tilde{\gamma}}(1, \dots, 1).$$

Notice that we have $\frac{\alpha_j^R - \lambda_j}{\tilde{\gamma}} \geq 0$ and $\sum_{j=1}^{m^R} \frac{\alpha_j^R - \lambda_j}{\tilde{\gamma}} = 1$. Moreover, it is easy to see that

$$\left(\frac{\alpha_1^R - \lambda_1}{\tilde{\gamma}}, \dots, \frac{\alpha_{m^R}^R - \lambda_{m^R}}{\tilde{\gamma}} \right) \neq (\alpha_1^R, \dots, \alpha_{m^R}^R)$$

as we have $\frac{\alpha_k^R - \lambda_k}{\tilde{\gamma}} = 0 \neq \alpha_k^R$. Thus, we found another strategy for player R against which player B is indifferent between all of her actions. Hence, the completely mixed Nash equilibrium is not unique in this case.

Now we consider the second of the two cases. This case implies that for all $(\lambda_1, \dots, \lambda_{m^R}) \in \mathbb{R}^{m^R} \setminus \{0\}$ fulfilling equation (2.19) we need to have $\lambda_j \neq 0$ for all $j \in \{1, \dots, m^R\}$ since otherwise we would be back in the first case. However, this means that b^1, \dots, b^{m^R-1} are linearly independent. As we have $m^R - 1 \geq m^B$, this implies $\text{span}((b^j)_{j=1, \dots, m^R-1}) = \mathbb{R}^{m^B}$ and it even must be $m^R - 1 = m^B$. So there exists $(\lambda'_1, \dots, \lambda'_{m^B}) \in \mathbb{R}^{m^B} \setminus \{0\}$ with $\lambda'_j \leq \alpha_j^R$ for all $j \in \{1, \dots, m^B\}$ and a scalar $\gamma' \neq 0$ such that

$$\sum_{j=1}^{m^B} \lambda'_j b^j = \gamma'(1, \dots, 1).$$

Further, define $\lambda'_{m^R} := 0$ and $\tilde{\gamma}' := \sum_{j=1}^{m^R} (\alpha_j^R - \lambda'_j)$. Notice that $\tilde{\gamma}' \geq \alpha_{m^R}^R - \lambda'_{m^R} = \alpha_{m^R}^R > 0$. Hence, taking into account equation (2.19) and recalling the scalar $\bar{\gamma} \in \mathbb{R} \setminus \{0\}$, we have

$$\sum_{j=1}^{m^R} \frac{\alpha_j^R - \lambda'_j}{\tilde{\gamma}'} b^j = \frac{1}{\tilde{\gamma}'} \sum_{j=1}^{m^R} \left(\frac{\lambda_j}{\bar{\gamma}} - \lambda'_j \right) b^j = \frac{1}{\tilde{\gamma}'} \left(\underbrace{\frac{1}{\bar{\gamma}} \sum_{j=1}^{m^R} \lambda_j b^j}_{=0} - \sum_{j=1}^{m^R} \lambda'_j b^j \right) = -\frac{\gamma'}{\tilde{\gamma}'}(1, \dots, 1).$$

Notice that, similarly to the first case, we have $\frac{\alpha_j^R - \lambda'_j}{\tilde{\gamma}'} \geq 0$ and $\sum_{j=1}^{m^R} \frac{\alpha_j^R - \lambda'_j}{\tilde{\gamma}'} = 1$. Also,

realize that

$$\left(\frac{\alpha_1^R - \lambda'_1}{\tilde{\gamma}'}, \dots, \frac{\alpha_{m^R}^R - \lambda'_{m^R}}{\tilde{\gamma}'} \right) \neq (\alpha_1^R, \dots, \alpha_{m^R}^R).$$

This is because we either have $\tilde{\gamma}' \neq 1$ implying $\frac{\alpha_{m^R}^R - \lambda'_{m^R}}{\tilde{\gamma}'} = \frac{\alpha_{m^R}^R}{\tilde{\gamma}'} \neq \alpha_{m^R}^R$ or $\tilde{\gamma}' = 1$ which means that $\frac{\alpha_j^R - \lambda'_j}{\tilde{\gamma}'} = \alpha_j^R - \lambda'_j \neq \alpha_j^R$ for some $j \in \{1, \dots, m^B\}$ since it is $(\lambda'_1, \dots, \lambda'_{m^B}) \neq 0$. Consequently, we found again another strategy for player R against which player B is indifferent between all of her actions.

Finally, notice that in both cases, any convex combination of the alternative strategy and α^R is again a strategy against which player B is indifferent between all of her actions. Thus, we even proved existence of a continuum of Nash equilibria. \square

Chapter 3

Strategic Formation of Homogeneous Bargaining Networks

3.1 Introduction

People often engage in bi- and multilateral bargaining: firms bargain with workers' unions over contracts, firms with other firms over prices or collaborations, politicians over environmental or trade agreements, or even friends and family members over household duties or other arrangements. However, in most of the situations that come to mind not everyone will be able or willing to bargain with anyone else. This idea can be expressed by means of a network. One's bargaining power in negotiations then commonly depends on the number and types of alternative partners as they present outside options. Agents typically intend to maximize their expected profit from bargaining, which suggests that beforehand they might want to influence and optimize their network of potential bargaining partners. This motivates that the underlying network should not be regarded as being exogenously given but as the outcome of strategic interaction among agents. However, establishing a connection to someone else usually costs some time and effort, which should be taken into account as well. This gives rise to an interesting trade-off between the costs of forming links and potential benefits from it. This consideration is the topic of this chapter.

We set up and analyze a sequential model of strategic network formation prior to a Manea (2011) infinite horizon network bargaining game. We consider *ex ante* homogeneous players who in the first stage strategically form undirected, costly links. In this

context, one might think of one-time initiation or communication costs that players have to bear. In the second stage, we take the resulting network as given and players sequentially bargain with a neighbor for the division of a mutually generated unit surplus. According to Manea (2011) all subgame perfect equilibria of the bargaining game are payoff equivalent. Players are supposed to anticipate these outcomes during the preceding network formation game and to choose their actions accordingly. We examine players' strategic behavior regarding network formation, characterize stable and efficient network structures, and determine induced bargaining outcomes.

After giving a description of the model including a summary of the underlying Manea (2011) bargaining game and his decisive results, we consider the seminal concept of pairwise stability established by Jackson and Wolinsky (1996). In the bargaining game, we assume players to be infinitely patient. For all levels of linking costs we state and prove sufficient conditions for a network to be pairwise stable (Theorem 3.1). While costs are relatively high, the only structures we find to be pairwise stable are specific unions of separated pairs and isolated players. When costs decrease, odd circles of increasing size can additionally emerge. At a transition point also lines of length three can be contained in a pairwise stable network. This result also establishes existence of pairwise stable networks at each level of linking costs. For each combination of the above subnetworks we establish precisely for which cost range it is pairwise stable and for which it is not (Corollary 3.1). Furthermore, we provide a complete characterization of pairwise stable equitable networks, i.e. of structures inducing homogeneous payoffs among players, by showing that in such a network any non-isolated player has to be contained in a separated pair or in an odd circle (Theorem 3.2). Then we focus on the remaining networks which must induce heterogeneous payoffs within a component and establish that any of these can at most be singularly pairwise stable, that is at most at a single cost level (Theorem 3.3). This concludes the complete characterization of non-singularly pairwise stable networks, which is a principal achievement of this chapter (Corollary 3.3). All non-singularly pairwise stable networks even prove to be pairwise Nash stable (Corollary 3.5).

As a second main result, we deduce that pairwise stability narrows down the diversity of induced bargaining outcomes among players substantially. However, though players are *ex ante* homogeneous, they do not have to be completely equal in this respect (Corollary 3.6). Beyond that, we find that singularly pairwise stable networks other than the few ones we identify might only occur at linking costs below a certain threshold (Corollary 3.4). Also, we reveal that networks containing a tree (with more than three players) or a certain kind of "cut-player" cannot even be singularly pair-

wise stable (Propositions 3.2 and 3.3). Moreover, we establish that, for sufficiently high linking costs, the networks being efficient in terms of a utilitarian welfare criterion coincide with the pairwise stable ones. As long as costs are low, however, the former networks constitute a proper subset of the latter while there also exists an intermediate cost range which does not even yield such a subset relation (Theorem 3.4 and Corollary 3.7). As a robustness check, we finally relax the assumption that players are infinitely patient and show that pairwise stability in this framework does not necessarily imply pairwise stability for the original case and vice versa (Examples 3.1 and 3.2).

For a concrete economic application which is captured by our model and which might contribute to a better understanding of the framework one can have the following in mind. Consider a number of similar firms beginning operation at the same time. They can mutually generate an (additional) surplus within bilateral projects by exploiting synergy potentials. For instance, this possibility might be based on capacity constraints or cost-saving opportunities. However, the underlying cooperation network is not existent yet and will therefore be the outcome of strategic interaction between firms. In charge of that are project managers who receive bonus payments proportional to their employer's profit from the project. Here, one-time costs might arise to prepare each two firms for mutual projects (adjustment of IT, joint training for workers etc.). We assume that each project manager keeps her job until she finalizes a joint project successfully by finding an agreement with the corresponding counterpart and leaves or is promoted afterwards and then gets replaced by a successor. Thus, the network remains unchanged after it has initially been established by the first project managers.

To take the suitable framework and convenient results established by Manea (2011) as a starting point in this context is fairly obvious.³⁷ To my best knowledge, it is the only work which purely focuses on the impact of explicit network structures on players' bargaining power and outcomes in a setting of decentralized bilateral bargaining without ex ante imposing any restrictions to the class of networks considered. Therefore, there are no distorting effects present in this setting as they might arise

³⁷Thus, the fact that O'Donnell (2011) chooses the same approach and pursues objectives similar to ours is not very surprising. Note, however, that the work at hand has been set up autonomously and independently from this (not publicly available) "honours thesis". Moreover, the two works differ mainly in three respects. First, the ways we choose to derive a complete characterization of (non-singularly) pairwise stable networks are distinct in large parts. Second, there are substantial shortcomings in O'Donnell's (2011) line of argumentation as we point out in Appendix 3.C. This goes as far as, based on O'Donnell (2011), our main results can only be considered as conjectures. And third, our analysis is in some parts more advanced as, for instance, we additionally consider the case of less than infinitely patient players (see Section 3.5).

from ex-ante heterogeneity among players and it is more general than buyer-seller scenarios which impose bipartite network structures. Moreover, Manea's (2011) network bargaining game remains analytically tractable and has some important properties. For any level of time discount there may exist several subgame perfect equilibria but he shows that all of these are payoff equivalent. Further, he develops an equally convenient and sophisticated algorithm determining the limit equilibrium payoffs for a given network of infinitely patient players. We make extensive use of this algorithm and contribute to a profound understanding of its features throughout this chapter.

The analysis of bargaining problems has a long tradition in the economic literature and dates back to the work of Nash (1950a, 1953). A Nash bargaining solution is based on factors like players' bargaining power and outside options, whereas their origin is not part of the theory. This also applies for Rubinstein (1982), who analyzes perfect equilibrium partitions in a two-player framework of sequential bargaining in discrete time with an infinite horizon, as well as for Rubinstein and Wolinsky (1985) who set up a model of bargaining in stationary markets with two populations. The work of Manea (2011), to which we add a preceding stage of strategic interaction, can be regarded as an extension or microfoundation of these seminal papers. Here, bargaining power is endogenized in a natural and well-defined manner as an outcome of the given network structure and the respective player's position herein. Further important contributions to the literature on decentralized bilateral bargaining in exogenously given networks have been made by Abreu and Manea (2012) and Corominas-Bosch (2004) where the latter considers the special case of buyer-seller networks.

Second, this chapter contributes to the more recently emerging literature on strategic network formation which was mainly aroused by the seminal paper of Jackson and Wolinsky (1996). Further prominent works which have been carried out since then, however not in a bargaining framework, are the ones by Bala and Goyal (2000); Calvó-Armengol (2004); Galeotti et al. (2006); Goyal and Joshi (2003, 2006); Watts (2001), just to name a few. Besides, some effort has been dedicated to gaining rather general insights regarding the existence, uniqueness and structure of stable networks. Hellmann (2013) and Hellmann and Landwehr (2014) are examples for this.³⁸

So far there exist only few papers combining these two fields of research. Calvó-Armengol (2003) studies a bargaining framework à la Rubinstein (1982) embedded in a network context and considers stability and efficiency issues. However, the mecha-

³⁸In Appendix 3.B we demonstrate that crucial conditions are not met in our model which implies that the results of Hellmann (2013) and Hellmann and Landwehr (2014) are in general not applicable to our framework.

nism determining bargaining partners is different from Manea (2011) and the network bargaining game ends after the first agreement has been found. As a consequence, in Calvó-Armengol's (2003) model a player's network position does not affect her bargaining power as such but only the probability that she is selected as proposer or responder. This leads to a characterization of pairwise stable networks in which the players' neighborhood size is the only relevant feature of the network structure. Thus, it differs substantially from our results though both works have in common the assumption that links are costly. In contrast, Manea (2011, Online Appendix) abstracts from explicit linking costs when approaching the issue of network formation as an extension of his model. He shows that for zero linking costs a network is pairwise stable if and only if it is equitable. Though results differ and get more complex for positive linking costs, we will see that the work at hand is in line with this finding such that both works complement one another.³⁹ Our additional considerations regarding efficiency and time discount further contribute to this. Most other papers studying strategic network formation in a bargaining context focus on buyer-seller networks, which is as well complementary to our more general approach. Kranton and Minehart (2001) and Polanski and Vega-Redondo (2013) are examples for this. Again, the latter does not involve explicit linking costs.

The rest of the chapter is organized as follows. In Section 3.2 we introduce the model including the decisive results of Manea (2011). The main results are developed in Section 3.3 which focuses on the structure of stable networks and induced bargaining outcomes in the case that players are infinitely patient. In Section 3.4 we examine networks with regard to efficiency. In Section 3.5 we state commonalities and differences regarding stability if players discount time to some degree. Finally, Section 3.6 concludes. Rather complex and lengthy proofs as well as closer considerations of certain related papers are presented in the appendix.

3.2 The Model

Let time be discrete and denoted by $t = 0, 1, 2, \dots$. For the initial period $t = 0$ consider a finite set of players $N = \{1, 2, \dots, n\}$. A *connection* or (*undirected*) *link* between two players $i, j \in N$, $i \neq j$, is denoted by $\{i, j\}$ which we abbreviate for

³⁹In fact, we show that only “skeletons” of equitable networks, that is certain unions of separated pairs and odd circles survive if costs are positive. However, non-equitable networks such as, for instance, unions of odd circles and an isolated player can additionally be pairwise stable in our setting.

simplicity by $ij = ji := \{i, j\}$. A collection of such links is an *undirected graph* or *network* $g \subseteq g^N := \{ij \mid i, j \in N, i \neq j\}$ where g^N is called the *complete network*. Let $N_i(g) := \{j \in N \mid ij \in g\}$ denote the set of player i 's *neighbors* in g and let $\eta_i(g) := |N_i(g)|$ be its cardinality which is also referred to as the *degree* of player i .

Furthermore, for a network g , a set $C \subseteq N$ is said to be a *component* if there exists a path between any two players in C and it is $N_j(g) \cap C = \emptyset$ for all $j \notin C$.^{40,41} The set of all components of g then gives a partition of the player set N . Moreover, a *subnetwork* $g' \subseteq g$ is said to be *component-induced* if there exists a component C of g such that $g' = g|_C$. In general, for any set $K \subseteq N$, we denote $g|_K := \{ij \in g \mid i, j \in K\}$ and we commonly consider such a subnetwork as being defined on the player set K instead of N (thus, disregarding isolated players in K^c). Besides, for two networks $g, g' \subseteq g^N$ let $g - g' := g \setminus g'$ ($g + g' := g \cup g'$, respectively) denote the network obtained by deleting the set of links $g' \cap g$ from (adding the set of links $g' \setminus g$ to) the network g .

In our model, *ex ante*, i.e. apart from their potentially differing network positions, players are assumed to be identical.⁴² These players are then assumed to strategically form links in period $t = 0$. The outcome of this network formation game is a network g . The interpretation of a link $ij \in g$ is that players $i, j \in N$ are able to mutually generate a unit surplus. On the contrary, each link causes constant costs of link formation $c > 0$ for both players involved. Thus, player i has to bear total costs of $\eta_i(g)c$ in $t = 0$.

We take this as a starting point for an infinite horizon bargaining game à la Manea (2011). In each period $t = 0, 1, 2, \dots$ nature randomly chooses one link $ij \in g$ which means that i and j are matched to bargain for a mutually generated unit surplus. One of the two players is randomly assigned the role of the proposer while the other one is selected as responder. Then the proposer makes an offer how to distribute the unit surplus and the responder has the choice: If she rejects, then both receive a payoff of zero and stay in the game whereas if she accepts, then both leave with the shares agreed on. In the latter case both players get replaced one-to-one in the next

⁴⁰We say that there exists a path between two players $i', i'' \in N$ in g if there exist players $i_1, i_2, \dots, i_{\bar{m}} \in N$, $\bar{m} \in \mathbb{N}$, such that $i_1 = i', i_{\bar{m}} = i''$ and $i_m i_{m+1} \in g$ for $m = 1, 2, \dots, \bar{m} - 1$.

⁴¹One can alternatively define the component $C_i(g) \subseteq N$ of player $i \in N$ in g as the minimal set of players such that both $i \in C_i(g)$ and $N_{i'}(g) \subseteq C_i(g)$ for all $i' \in C_i(g)$.

⁴²In the literature, this is sometimes referred to as a “homogeneous society” (see e.g. Hellmann and Landwehr, 2014).

period such that the initially formed network remains unchanged.^{43,44} This implies that each of the (initial) players $1, 2, \dots, n$ will bargain successfully one time at most. A player's strategy in this setting pins down the offer she makes as proposer and the answer she gives as responder after each possible history of the game. Based on this, a player's payoff is then specified as her discounted expected agreement gains. A strategy profile is said to be a "subgame perfect equilibrium" of the bargaining game if it induces Nash equilibria in subgames following every history (see Manea, 2011). Players are assumed to discount time by a uniform discount factor $\delta \in (0, 1)$.⁴⁵

The key result from Manea is that all subgame perfect equilibria are payoff equivalent and that each player's equilibrium payoff exclusively depends on her network position and the discount factor δ (see Manea, 2011, Theorem 1). Moreover, the *equilibrium payoff vector* which we denote as $v^{*\delta}(g) = (v_i^{*\delta}(g))_{i \in N}$ is the unique solution to the equation system

$$v_i = \left(1 - \sum_{j \in N_i(g)} \frac{1}{2d^{\#}(g)}\right) \delta v_i + \sum_{j \in N_i(g)} \frac{1}{2d^{\#}(g)} \max\{1 - \delta v_j, \delta v_i\}, \quad i \in N, \quad (3.1)$$

where $d^{\#}(g)$ denotes the total number of links in the network g . If we have $\delta(v_i^{*\delta}(g) + v_j^{*\delta}(g)) < 1$ for $ij \in g$, then this means that player i and j find an agreement when their mutual link is selected whereas $\delta(v_i^{*\delta}(g) + v_j^{*\delta}(g)) > 1$ means that each of them prefers to wait for a potentially better deal with a weaker partner.⁴⁶ This gives rise to the definition of the so called *equilibrium agreement network* $g^{*\delta} := \{ij \in g \mid \delta(v_i^{*\delta}(g) + v_j^{*\delta}(g)) \leq 1\}$.

We assume that players $1, 2, \dots, n$ know the whole structure of the network g . They are therefore able to anticipate equilibrium payoffs and are assumed to play a subgame perfect equilibrium strategy profile. Given a network g and a discount factor δ , we for simplicity refer to $v_i^{*\delta}(g)$ as player i 's *payoff*. Throughout this chapter it is important to distinguish this precisely from a player's *profit* which is given by

⁴³This replacement is primarily due to technical reasons. The implication that the network structure does not change over time makes the model analytically tractable. However, recalling the motivating example on bilateral project cooperation from Section 3.1 gives a hint that there are indeed situations in reality which are roughly captured by that.

⁴⁴This is why Manea carefully distinguishes between network positions and (potentially) different players being in one and the same position in different periods. However, as we examine solely the stage of network formation at time $t = 0$ here, we can neglect this distinction.

⁴⁵One might argue that players should be allowed to form (or delete) links in periods $t = 1, 2, \dots$ as well. However, as the game has an infinite horizon, in any period any player faces just the same situation as (the player who was in her network position) in the previous period. Therefore, there do not arise additional or altered incentives regarding link formation over time.

⁴⁶In the case $\delta(v_i^{*\delta}(g) + v_j^{*\delta}(g)) = 1$ both players are indifferent.

her payoff minus linking costs. Thus, given g and δ the profit of player $i \in N$ is

$$u_i^{*\delta}(g) := v_i^{*\delta}(g) - \eta_i(g)c.$$

Note that a non-isolated player's profit is therefore always strictly smaller than her payoff as we assume linking costs to be positive.

For our main results we focus on the limit case of $\delta \rightarrow 1$ which means that players are infinitely patient. For this case Manea (2011, Theorem 2) finds that for all δ being greater than some bound the corresponding equilibrium agreement networks are equal. This network g^* is then defined as the *limit equilibrium agreement network*. Moreover, we again take from Manea (2011, Theorem 2) that the so called *limit equilibrium payoff vector* $v^*(g) = \lim_{\delta \rightarrow 1} v^{*\delta}(g)$ is well-defined, i.e. it always exists. Beyond that, Manea (2011, Proposition 2) shows that the sum of the payoffs of two players being linked cannot be smaller than one, i.e. $v_i^*(g) + v_j^*(g) \geq 1$ for all $ij \in g$, and they must sum up to one if the link is contained in the limit equilibrium agreement network, i.e. we have $v_i^*(g) + v_j^*(g) = 1$ if $ij \in g^*$. We utilize this during our analysis as well.

Manea develops a smart algorithm to determine the limit equilibrium payoff vector $v^*(g)$ and we make heavily use of this computation method. To prepare for the implementation of the algorithm we need to introduce some additional notation. For any set of players $M \subseteq N$ and any network g let $L^g(M) := \{j \in N \mid ij \in g, i \in M\}$ be the corresponding *partner set* in g , that is the set of players having a link in g to a player in M .⁴⁷ Further, a set $M \subseteq N$ is called *g -independent* if we have $g|_M := \{ij \in g \mid i, j \in M\} = \emptyset$, i.e. if no two players contained in M are linked in g . Moreover, let $\mathcal{I}(g) \subseteq \mathcal{P}(N)$ denote the set of all non-empty g -independent subsets of N . Then the algorithm determining the payoff vector $v^*(g)$ is the following.

Definition 3.1 (Manea (2011)). *For a given network g on the player set N , the algorithm $\mathcal{A}(g)$ provides a sequence $(r_s, x_s, M_s, L_s, N_s, g_s)_{s=1, \dots, \bar{s}}$ which is defined recursively as follows. Let $N_1 := N$ and $g_1 := g$. For $s \geq 1$, if $N_s = \emptyset$ then stop and set $\bar{s} = s$. Otherwise, let*

$$r_s = \min_{M \subseteq N_s, M \in \mathcal{I}(g)} \frac{|L^{g_s}(M)|}{|M|}. \quad (3.2)$$

If $r_s \geq 1$ then stop and set $\bar{s} = s$. Otherwise, set $x_s = \frac{r_s}{1+r_s}$. Let M_s be the union of all minimizers M in (3.2). Denote $L_s := L^{g_s}(M_s)$. Let $N_{s+1} := N_s \setminus (M_s \cup L_s)$ and

⁴⁷Although it does not make a big difference, do not confuse with the notation of Manea who refers to $L^g(M)$ instead.

$g_{s+1} := g|_{N_{s+1}}$.

Given such a sequence $(r_s, x_s, M_s, L_s, N_s, g_s)_{s=1, \dots, \bar{s}}$ being the outcome of the described algorithm $\mathcal{A}(g)$, the limit equilibrium payoff vector for this network can be determined by applying a simple rule. Note that this rather sophisticated result of Manea (2011, Theorem 4) is absolutely fundamental for our work.

Payoff Computation (Manea (2011)). *Let $(r_s, x_s, M_s, L_s, N_s, g_s)_{s=1, \dots, \bar{s}}$ be the outcome of $\mathcal{A}(g)$ for a given network g . Then the limit equilibrium payoffs are given by*

$$\begin{aligned} v_i^*(g) &= x_s && \text{for all } i \in M_s, s < \bar{s}, \\ v_j^*(g) &= 1 - x_s && \text{for all } j \in L_s, s < \bar{s}, \\ v_k^*(g) &= \frac{1}{2} && \text{for all } k \in N_{\bar{s}}. \end{aligned} \tag{3.3}$$

Let us figure out what the algorithm $\mathcal{A}(g)$ in combination with the payoff calculation rule actually does. Starting with the network g and player set N , at each step s it identifies the so called *minimal shortage ratio* r_s among the remaining players N_s in the network $g_s = g|_{N_s}$. There is a largest g -independent set M_s which minimizes this shortage ratio such that

$$r_s = \frac{|L_s|}{|M_s|},$$

where L_s is the partner set of M_s . The limit equilibrium payoff of the players in M_s is then given by $x_s = \frac{r_s}{1+r_s} = \frac{|L_s|}{|M_s|+|L_s|} < \frac{1}{2}$ while their partners in L_s receive $1 - x_s = \frac{|M_s|}{|M_s|+|L_s|} > \frac{1}{2}$. These players and their links are then deleted from the network and the algorithm moves forward to the next step. It stops when there are either no more players left or if the minimal shortage ratio is greater than or equal to one. In the latter case the limit equilibrium payoff of all remaining players is $\frac{1}{2}$. Manea (2011, Proposition 3) shows that the sequence of minimal shortage ratios $(r_s)_s$ and therefore also $(x_s)_s$ are strictly increasing.

In the framework with $\delta \rightarrow 1$ the described algorithm $\mathcal{A}(g)$ together with the previous considerations then determines the profit $u_i^*(g) = v_i^*(g) - \eta_i(g)c$ of each player $i \in N$. Broadly speaking, the algorithm quantifies the main forces that, in terms of payoffs, each player benefits from being linked to other players while preferring neighbors who are sparsely connected or rather who are only connected to other players who are in stronger positions than oneself. It is important to note that the profile of payoffs and therefore also the profile of profits $u^* = (u_i^*)_{i \in N}$ is *component-decomposable*, that is $u_i^*(g) = u_i^*(g|_{C_i(g)})$ for all players $i \in N$ and networks g where $C_i(g) \subseteq N$ is the component of player i in g . Thus, a player's profit is not affected by

the existence or structure of subnetworks induced by components she is not contained in.

Beyond that, note that Manea develops the algorithm $\mathcal{A}(g)$ under the assumption that there are no isolated players in the underlying network g . However, it is easy to see that equations (3.3) are still fulfilled if one relaxes this restriction. It is clear that isolated players have a limit equilibrium payoff of zero since they have no bargaining partner they could generate a unit surplus with. At the same time the algorithm $\mathcal{A}(g)$ provides $r_1 = 0$ such that $x_1 = 0$. In this case, M_1 is the set of all isolated players in the network and we have $L_1 = \emptyset$. Then according to (3.3) and as required, all players in M_1 are assigned a limit equilibrium payoff of $x_1 = 0$.

Throughout the next section we assume that each player can influence the network structure by altering own links in $t = 0$, i.e. before the bargaining game starts. This means that the network is no longer exogenously given as in the work of Manea but the outcome of strategic interaction between players. This gives rise to questions regarding the stability of networks and leads to the main results of this chapter. Our analysis is mainly based on the seminal concept of pairwise stability which has been introduced by Jackson and Wolinsky (1996).

Definition 3.2 (Pairwise Stability, Jackson and Wolinsky (1996)). *Consider the player set N and a profile of network utility or profit functions $(u_i)_{i \in N}$. Then a network g is said to be pairwise stable if both*

(i) *for all $ij \in g$: $u_i(g) \geq u_i(g - ij)$ and*

(ii) *for all $ij \notin g$: if $u_i(g + ij) > u_i(g)$, then $u_j(g + ij) < u_j(g)$.*

So, according to this definition, a network is pairwise stable if no player can improve by deleting a single link and also no two players can both individually benefit from adding a mutual link. The analysis of our model demands to distinguish between networks being pairwise stable only at a single cost level and those fulfilling the conditions for two or more values of linking costs. For this purpose we introduce the following terminology.

Definition 3.3 (Singular and Non-Singular Pairwise Stability). *In the considered framework with network profit function $u = u^*$ and linking costs $c > 0$, a network g is called*

- singularly pairwise stable if g is pairwise stable at cost level $c > 0$ but nowhere else,

- non-singularly pairwise stable *if there exists another cost level $c' > 0$, $c' \neq c$, such that g is pairwise stable at c and c' .*

For a singularly pairwise stable network it is obviously very rare (even a singularity) to encounter precisely the parametrization where it is pairwise stable. For instance, this notion is therefore not robust with respect to any slight change of the cost level. Thus, we are predominantly interested in networks being non-singularly pairwise stable. In what follows, we therefore focus on this latter subclass of pairwise stable networks. Note that, in general, networks being pairwise stable for a continuum of cost levels constitute a subset of this subclass.⁴⁸ As an outcome of our analysis, however, it turns out that, in our model, all non-singularly pairwise stable networks are even pairwise stable on a cost interval of positive length.

Observe that the notions of stability considered so far focus exclusively on one link deviations. One might argue that in many contexts at least severing links does not cost any effort as it does not require coordination between players. Thus, one should allow for the possibility of multiple link deletion. This gives rise to the concept of *pairwise Nash stability* where in Definition 3.2 condition (i) is simply replaced by

$$(i)' \text{ for all } i \in N, l_i \subseteq \{ij \in g \mid j \in N\}: u_i(g) \geq u_i(g - l_i)$$

(see e.g. Bloch and Jackson, 2006). As opposed to stability issues, in Section 3.4 we focus on networks being efficient, that is on structures which yield maximal utilitarian welfare.

Definition 3.4 (Utilitarian Welfare and Efficiency, Jackson and Wolinsky (1996)). *Consider the player set N and a profile of network utility or profit functions $(u_i)_{i \in N}$.*

- *The utilitarian welfare yielded by a network g is defined as*

$$U(g) := \sum_{i \in N} u_i(g).$$

- *A network g is said to be efficient if for all $g' \subseteq g^N$ we have*

$$U(g) \geq U(g').$$

So we say that a network is efficient if the unweighted sum of individual profits

⁴⁸One might consider to call a network “generically pairwise stable” if it is pairwise stable for a continuum of cost levels. See e.g. Baetz (2015) who refers to a “generic equilibrium”.

cannot be further increased by altering links in any way.⁴⁹ Based on these fundamental concepts and considerations we are now able to establish our main results in the following Section 3.3 as well as, subsequently, further interesting findings.

3.3 Characterization of Stable Networks

In the main part of this section we derive a complete characterization of non-singularly pairwise stable networks for the case that players are infinitely patient. Further, we examine the implications for possible bargaining outcomes resulting thereof (see Subsection 3.3.1) as well as the possibility of networks to be singularly pairwise stable (see Subsection 3.3.2).

We consider period $t = 0$ and suppose that players, who anticipate the infinite horizon network bargaining game they will be involved in, individually intend to maximize their expected profits. Given the framework with $\delta \rightarrow 1$, we aim for a characterization of stable network structures in the sense that no player has incentives to add or delete links. As a first step, we identify pairwise stable structures for all levels of linking costs $c > 0$. Afterwards, we gradually rule out the possibility to be pairwise stable for a broad range of networks until we arrive at a complete characterization of non-singularly pairwise stable networks. It turns out that these structures are even pairwise Nash stable.

Note first that there exist some general results about existence, uniqueness and the structure of pairwise stable networks in the literature which we should take into account. One might suppose that the works of Hellmann (2013) and Hellmann and Landwehr (2014), which are perhaps closest within this strand of literature, could simplify the analysis of our model. In Appendix 3.B, however, we demonstrate that crucial conditions are not satisfied here which implies that their results are in general not applicable.

As already mentioned, we first focus on one link deviations which is captured by the notion of pairwise stability (see Definition 3.2). To get a first impression of the problem let us have a look at the situation for three players, i.e. for $N = \{1, 2, 3\}$. It turns out that this case already covers many important aspects of the network formation game. Figure 3.1 illustrates the four types of networks which might appear including the induced profits u_i^* for each player $i \in N$. To comprehend these profits, consider the algorithm introduced in Definition 3.1 and the subsequent payoff

⁴⁹This definition also goes back to Jackson and Wolinsky (1996) who call such a network “strongly efficient”. However, in the literature this property is usually simply referred to as efficiency (see e.g. Jackson, 2008b, p. 157).

computation rule. Besides, note that all other possible networks can be derived by permuting players which would not provide additional insights as players are assumed to be ex ante homogeneous.

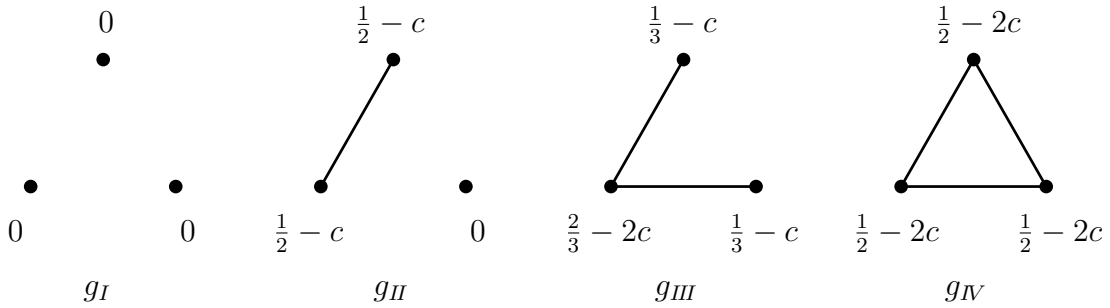


Figure 3.1: A sketch of the four network structures which can arise in the case $n = 3$ with induced profits

Let us consider these networks in detail. We see immediately that the network g_I is pairwise stable if and only if the linking costs c are greater than or equal to $\frac{1}{2}$. Otherwise any two players could increase their profit from zero to $\frac{1}{2} - c > 0$ by creating a mutual link. However, for $c = \frac{1}{2}$ also no player wants to delete this link and indeed, the cost range $c \in (\frac{1}{6}, \frac{1}{2}]$ is the one for which g_{II} is pairwise stable. Here, link deletion is obviously not beneficial and if one of the two connected players creates a link to the third player, then she would end up with a profit of $\frac{2}{3} - 2c$ which is strictly smaller than $\frac{1}{2} - c$ for this cost range. These latter two terms are equal for $c = \frac{1}{6}$ but the third player would improve from zero to $\frac{1}{6}$ in this case. Therefore, at this or an even smaller cost level, g_{II} cannot be pairwise stable. But so is g_{III} for $c = \frac{1}{6}$. This is because here no player has incentives to delete a link and the two players who are not connected are indifferent between creating a mutual link and not creating it as for this cost level we have $\frac{1}{3} - c = \frac{1}{6} = \frac{1}{2} - 2c$. However, if linking costs are even smaller, then both would profit from this link. Thus, g_{III} is pairwise stable if and only if $c = \frac{1}{6}$. Finally, the network g_{IV} is pairwise stable for $c \in (0, \frac{1}{6}]$ but obviously not at higher cost levels.

It turns out that the observed mechanisms being crucial in the three-player case hold similarly also in general. Our first theorem reveals sufficient conditions for networks to be pairwise stable. More precisely, it identifies, for all cost levels, concrete network structures being pairwise stable.

Theorem 3.1 (Sufficient Conditions for Networks to Be Pairwise Stable). *In the framework as introduced in Section 3.2 with $\delta \rightarrow 1$ and player set N the following holds:*

- (i) *The empty network is pairwise stable if $c \geq \frac{1}{2}$.*
- (ii) *A network which is a union of separated pairs and at most one isolated player is pairwise stable if $c \in (\frac{1}{6}, \frac{1}{2}]$.⁵⁰ Additionally, if $c = \frac{1}{2}$, then several isolated players can coexist in a pairwise stable network.*
- (iii) *A network which is a union of odd circles with at most $\frac{1}{2c}$ players and either separated pairs or at most one isolated player is pairwise stable if $c \in (0, \frac{1}{6}]$.⁵¹ Additionally, if $c = \frac{1}{6}$ and given that there is no isolated player, then there can also exist lines of length three in a pairwise stable network.⁵²*

The formal proof of this theorem is, like all other more complex or lengthy proofs, provided in Appendix 3.A. It is important to note that, when considering the above mentioned unions of subnetworks, we do not mean that the respective network necessarily has to be composed of all of the stated subnetworks to be pairwise stable. For instance, if we consider costs $c \in (0, \frac{1}{6}]$, then a network consisting only of separated pairs or only of (some of the permissible) odd circles is pairwise stable as well. Furthermore, note that all of these subnetworks are component-induced which implies that unions are disjoint with respect to contained links and players.

A byproduct of Theorem 3.1 is that it guarantees existence of a pairwise stable network at each level of linking costs. Furthermore, we have given a characterization of at least some pairwise stable networks for each level of costs. However, it is not clear at all that the types of networks mentioned in the theorem are in each case the only pairwise stable ones. Anyway, we can already state some consequences from our observations in the three-player case considered in Figure 3.1 and the proof of Theorem 3.1. This is done in the following corollary.

Corollary 3.1. *In the model with $\delta \rightarrow 1$ a network cannot be pairwise stable if it contains*

- (i) *more than one isolated player while $c < \frac{1}{2}$,*
- (ii) *a separated pair while $c > \frac{1}{2}$,*
- (iii) *a line of length three while $c \neq \frac{1}{6}$,*

⁵⁰A separated pair denotes a subnetwork induced by a two-player component.

⁵¹A circle denotes a component-induced subnetwork which is regular of degree two. A circle with m players or a m -player circle is induced by a component with cardinality $m \geq 3$ and it is called odd if this cardinality is an odd number.

⁵²A line of length $m \geq 3$ denotes a subnetwork induced by a m -player component which can be transformed to a m -player circle by adding one link.

(iv) an odd circle with more than $\frac{1}{2c}$ members,⁵³

(v) an isolated player combined with a separated pair or a line of length three while $c \leq \frac{1}{6}$.

Statements (i)–(iv) as well as the first part of Statement (v) of Corollary 3.1 follow immediately from what we learned in the three-player case and the proof of Theorem 3.1. To see that an isolated player and a line of length three cannot coexist in a network being pairwise stable for some $c \leq \frac{1}{6}$ is, however, also quite obvious. An isolated player's profit is always zero while each of the two players in a line of length three having one link receives $\frac{1}{3} - c$ as we know from the three-player case. If one of these players links to an isolated player, then the algorithm $\mathcal{A}(\cdot)$ yields that all players in the new component receive a payoff of $\frac{1}{2}$. Thus, it is beneficial for both players to build this mutual link as for $c \leq \frac{1}{6}$ we have $\frac{1}{2} - 2c \geq \frac{1}{3} - c$ and $\frac{1}{2} - c > 0$. One should perhaps mention that according to (iii) it is clear anyway that we cannot have a line of length three in a pairwise stable network if $c < \frac{1}{6}$. So the above additional consideration is actually only relevant for $c = \frac{1}{6}$.

The results we establish in the further course of this section together with the above corollary reveal that the conditions stated in Theorem 3.1 are not only sufficient but also necessary for a network to be non-singularly pairwise stable. Note however that a network which is composed of several isolated players and at least one separated pair is only singularly pairwise stable (at $c = \frac{1}{2}$, see Theorem 3.1(ii)). Similarly, networks containing a line of length three can at most be singularly pairwise stable (at $c = \frac{1}{6}$, see Theorem 3.1(iii)).

In general, it is clear that a network can only be pairwise stable if any link is profitable for both players involved or linking costs $c > 0$ are at least covered by the additional payoff. Therefore, the intuition says that pairwise stable networks cannot have so called *disagreement links*, that is links which are not contained in the corresponding limit equilibrium agreement network. One might argue that such a link leads to higher costs for both players connected through it whereas it seems to be irrelevant regarding payoffs. If it is selected by nature at some point in time, the two players will not find an agreement in the network bargaining game. So why should they connect? From another perspective, however, things seem to be a bit more complicated. With regard to the mechanism of the algorithm $\mathcal{A}(g)$ which determines the payoff of each player in a given network, a disagreement link could have a rather global effect. It might be conceivable that deleting such a link can change the whole

⁵³In particular, this means that there can be no odd circles at all in pairwise stable networks as long as $c > \frac{1}{6}$.

payoff structure induced by the network, which then might also affect the two edge players. For instance, the presence of a link, though giving rise to a disagreement, might prevent one of the players it connects and who receives a payoff of at least $\frac{1}{2}$ from being deleted during the algorithm as part of a g -independent set, which would induce a lower payoff for this player. However, the following proposition establishes that our first intuition is indeed correct.

Proposition 3.1 (Disagreement Links). *Consider the framework with $\delta \rightarrow 1$. If a network g is pairwise stable for some cost level $c > 0$, then we must have that the network itself and the corresponding limit equilibrium agreement network coincide, i.e. $g = g^*$, meaning that g does not contain disagreement links. In particular, this implies that we have $v_i^*(g) + v_j^*(g) = 1$ for all $ij \in g$.*

This is a valuable insight which we repeatedly make use of in what follows. The proof of the proposition can be found in the appendix. Basically, we adapt the proof of Manea (2011, Theorem 4) in such a way as to show that for any pairwise stable network g and any disagreement link $ij \in g$ it would have to be $v_k^*(g) = v_k^*(g - ij)$ for all $k \in N$. Thus, players i and j would obviously want to delete their mutual link which contradicts pairwise stability.

We now first consider networks inducing a homogeneous payoff structure. In line with Manea (2011) we call a network *equitable* if every player receives a payoff of $\frac{1}{2}$. For a given network g with player set N we define the subset $\tilde{N}(g) := \{i \in N \mid v_i^*(g) = \frac{1}{2}\}$. We utilize this notation in the following theorem. In combination with Proposition 3.1, it reveals that a network can only be pairwise stable if any player receiving a payoff of $\frac{1}{2}$ is contained in a component which either induces a separated pair or an odd circle.

Theorem 3.2 (Equitability and Pairwise Stability). *Consider the model with $\delta \rightarrow 1$. If a network g is pairwise stable for some cost level $c > 0$, then $g|_{\tilde{N}(g)}$ must be a union of separated pairs and odd circles.⁵⁴*

The proof, which is again given in Appendix 3.A, is by contradiction. The idea is to assume that g is pairwise stable but $g|_{\tilde{N}(g)}$ is not a union of separated pairs and odd circles. Note that by Proposition 3.1 a link from a player in $\tilde{N}(g)$ to a player outside this set cannot exist which implies that we have $v_i^*(g) = v_i^*(g|_{\tilde{N}(g)})$ for all $i \in \tilde{N}(g)$ as payoffs are component-decomposable. Further, we make use of both directions of Manea (2011, Theorem 5) who establishes that a network is equitable if and only if

⁵⁴As usual, $g|_{\tilde{N}(g)}$ here is considered as being defined on the player set $\tilde{N}(g)$ instead of N .

it has a so called “edge cover” g' composed of separated pairs and odd circles. A network g' is said to be an *edge cover* of $g|_{\tilde{N}(g)}$ if it fulfills $g' \subseteq g|_{\tilde{N}(g)}$ and no player in $\tilde{N}(g)$ is isolated in g' . This implies that any player in $\tilde{N}(g)$ has an incentive to delete each of her links not contained in g' .

Though statements differ, notice that Theorem 3.2 is in line with Manea (2011, Theorem 1(ii) of the Online Appendix). The latter establishes that for zero linking costs a network is pairwise stable if and only if it is equitable. Of course, in this case no player can gain anything from deleting redundant links from an equitable network. This then gives rise to a larger class of pairwise stable (equitable) networks. For instance, any even circle or line of even length is equitable and therefore pairwise stable as long as there are no linking costs whereas Theorem 3.2 rules out this possibility for $c > 0$. On the contrary, as we have seen in Figure 3.1 and Theorem 3.1, for positive linking costs there additionally exist non-equitable structures such as networks composed of an isolated player combined with separated pairs or odd circles which can be pairwise stable. Another example for this is the line of length three though such a component-induced subnetwork can only occur in a singularly pairwise stable network. In what follows, this kind of singularity is central to our investigation.

Summing up our results so far, for all levels of positive linking costs, we achieved a complete characterization of networks which are pairwise stable and induce homogeneous payoffs within each of its components. In these networks, all payoffs must be equal to either $\frac{1}{2}$ or zero by Proposition 3.1. According to Theorem 3.1, Corollary 3.1 and Theorem 3.2 certain unions of separated pairs, odd circles and isolated players constitute this class of networks.

Thus, it remains to consider structures which induce heterogeneous payoffs within a component. Most of the rest of the section is devoted to the examination of such networks and the question whether and in which cases they can potentially be pairwise stable. To begin with, let us make sure to be aware of the following property of pairwise stable non-equitable networks. Taking into account the payoff computation rule it is an immediate consequence of Proposition 3.1.

Corollary 3.2. *Consider the framework with $\delta \rightarrow 1$. Let $g \neq \emptyset$ be a non-equitable network having only one component and assume that it is pairwise stable for some cost level $c > 0$. Then there exists a unique partition $M \dot{\cup} L = N$ with $|M| > |L|$ and*

$g|_M = g|_L = \emptyset$, meaning that g is bipartite.⁵⁵ Payoffs are then given by

$$\begin{aligned} v_i^*(g) &= x && \text{for all } i \in M \quad \text{and} \\ v_j^*(g) &= 1 - x && \text{for all } j \in L, \end{aligned}$$

where $x = \frac{|L|}{|M|+|L|}$.

Note here that according to Manea (2011, Proposition 3) the sequence of minimal shortage ratios provided by the algorithm in Definition 3.1 is strictly increasing for any network. Thus, Corollary 3.2 implies that for any non-equitable pairwise stable network g consisting of only one component the algorithm $\mathcal{A}(g)$ has to stop after removing all players during the first step. This then leads to the heterogeneous payoff distribution with two different payoffs, one below and one above $\frac{1}{2}$.

Based on Corollary 3.2 the following theorem concludes the complete characterization of non-singularly pairwise stable networks. It establishes that any network in which players belonging to one component receive different payoffs can at most be singularly pairwise stable.

Theorem 3.3 (Payoff Heterogeneity and Pairwise Stability). *Consider the framework with $\delta \rightarrow 1$. If a network is pairwise stable for some cost level $c > 0$ and there is a component in which players receive heterogeneous payoffs, then in any such component there must occur exactly two different payoffs $x \in (0, \frac{1}{2})$ and $1 - x \in (\frac{1}{2}, 1)$ with*

$$x + c = \frac{1}{2}. \quad (3.4)$$

The proof rests on two lemmas which are of some independent interest. We shall now state these lemmas, one after the other, and then show how they combine to establish the theorem.

We first show that if any two players, whose payoffs in a pairwise stable network are strictly smaller than $\frac{1}{2}$, link to each other, then both receive a payoff of $\frac{1}{2}$ in the resulting network.

Lemma 3.1. *In the framework with $\delta \rightarrow 1$ consider a pairwise stable network g for which the algorithm $\mathcal{A}(g)$ provides $(r_1, x_1, M_1, L_1, N_1, g_1)$, i.e. $\bar{s} = 1$, such that $r_1 \in (0, 1)$. Then for all $i, j \in M_1$ it is*

$$v_i^*(g + ij) = v_j^*(g + ij) = \frac{1}{2}.$$

⁵⁵If we write $M \dot{\cup} L$, this simply denotes the union of two sets M and L being disjoint. We use this notation whenever disjointness is of importance.

Further, if the player set $N = \{1, \dots, n\}$ is extended by a player $n+1$ while the network g remains unchanged, it similarly is $v_i^*(g + i(n+1)) = v_{n+1}^*(g + i(n+1)) = \frac{1}{2}$.

The second lemma, in contrast, considers link deletion and players who receive a payoff being strictly greater than $\frac{1}{2}$ in a pairwise stable network. It establishes that one link deletion cannot effect these players' payoffs to fall below $\frac{1}{2}$.

Lemma 3.2. *In the framework with $\delta \rightarrow 1$ consider a pairwise stable network g for which the algorithm $\mathcal{A}(g)$ provides $(r_1, x_1, M_1, L_1, N_1, g_1)$, i.e. $\bar{s} = 1$, such that $r_1 \in (0, 1)$. Then for all $j \in L_1$, $i \in M_1$ and $kl \in g$ it is*

$$v_j^*(g - kl) \geq \frac{1}{2} \geq v_i^*(g - kl).$$

The proofs of these lemmas are somewhat lengthy and as usual provided in the appendix. In both cases we show that if the respective statement were not true, then this would imply that the player set is infinite. To arrive at this contradiction we make use of an additional, rather technical lemma which we also provide in the appendix (see Lemma 3.3). Based on these lemmas, the proof of the theorem is straightforward.

Proof of Theorem 3.3. Let g be a pairwise stable network inducing heterogeneous payoffs within a component $C \subseteq N$. Let $g' := g|_C$. According to Corollary 3.2, the algorithm $\mathcal{A}(g')$ (with $N_1 = C$) has to stop after the first step, i.e. $\bar{s}' = 1$.⁵⁶ Let $(r'_1, x, M'_1, L'_1, N'_1, g'_1)$ be the outcome of $\mathcal{A}(g')$ and $i \in M'_1$, $j \in L'_1$. Then any player in C must either receive a payoff of $x = \frac{|L'_1|}{|M'_1| + |L'_1|} \in (0, \frac{1}{2})$ or $1 - x = \frac{|M'_1|}{|M'_1| + |L'_1|} \in (\frac{1}{2}, 1)$. Then Lemma 3.1 provides the stability condition

$$x - \eta_i(g')c \geq \frac{1}{2} - (\eta_i(g') + 1)c \quad \Leftrightarrow \quad x + c \geq \frac{1}{2}.$$

Similarly, according to Lemma 3.2 we must have

$$(1 - x) - \eta_j(g')c \geq \frac{1}{2} - (\eta_j(g') - 1)c \quad \Leftrightarrow \quad x + c \leq \frac{1}{2}.$$

So payoffs must be $x = \frac{1}{2} - c$ and $1 - x = \frac{1}{2} + c$. Obviously, this has to hold for all components of g in which players receive heterogeneous payoffs. \square

Notice by considering the limit case $c \rightarrow 0$ that Theorem 3.3 is in line with Manea's (2011, Online Appendix) result that for zero linking costs any pairwise stable network

⁵⁶Disregarding isolated players here by considering the restricted player set is w.l.o.g. as the profile of payoffs respectively profits is component-decomposable.

must be equitable. As an immediate consequence of Theorem 3.3 and the previous findings we arrive at the main result of this chapter which we write down as a corollary.

Corollary 3.3 (Complete Characterization). *In the framework with $\delta \rightarrow 1$ the class of non-singularly pairwise stable networks is completely characterized by Theorem 3.1 for each level of linking costs $c > 0$.⁵⁷ Thus, specific unions of isolated players, separated pairs and odd circles constitute this class.*

To see this, recall first that according to Theorem 3.2 any network g not mentioned in Theorem 3.1 can only be pairwise stable if it induces heterogeneous payoffs within at least one component. Each player contained in such a component must either receive a payoff of $x = \frac{1}{2} - c$ or $1 - x = \frac{1}{2} + c$ by Theorem 3.3.⁵⁸ Be aware that these equations do not represent calculation rules determining payoffs in g but necessary conditions for a network to (possibly) be pairwise stable. Recall that x is in fact determined by the algorithm $\mathcal{A}(g)$, meaning that it solely depends on the structure of g and that $c > 0$ is an independent parameter of the model. Therefore, such a network g can only be pairwise stable at the single cost level $c = \frac{1}{2} - x$. Together with Corollary 3.1 this establishes Corollary 3.3.

Besides, given this cost level, it is of course not at all clear that a network in which each player either receives a payoff of $x \in (0, \frac{1}{2})$ or $1 - x$ is actually pairwise stable. However, even if this is the case, i.e. if such a network is singularly pairwise stable, then any two players with a payoff of x are indifferent between leaving the network unchanged and adding a mutual link (see Lemma 3.1). Also, any player receiving a payoff of $1 - x$ must be indifferent between keeping all of her links and deleting any of them (see Lemma 3.2). In this sense, the case that a network which induces heterogeneous payoffs within a component is pairwise stable and does indeed form is really special and insofar a singularity. We are able, however, to specify one such network (and variations respectively generalizations of it as a component-induced subnetwork), namely the line of length three with payoffs $x = \frac{1}{3}$ and $1 - x = \frac{2}{3}$ which is pairwise stable if and only if $c = \frac{1}{6}$.⁵⁹ As opposed to this, we even rule out the possibility to be singularly pairwise stable for a broad range of network structures in the further course of this section (see Subsection 3.3.2).

A first step in this direction can be stated as a further corollary to our results so far, in particular to Theorem 3.3.

⁵⁷Of course, singularly pairwise stable networks mentioned in Theorem 3.1 are to be ignored here.

⁵⁸By the way, recalling Corollary 3.2, the induced subnetwork must be bipartite.

⁵⁹In Section 3.5, we additionally reveal that the stability of this particular network is not robust in another respect either.

Corollary 3.4. *In the framework with $\delta \rightarrow 1$ a network cannot be pairwise stable*

- (i) *at any cost level $c > \frac{1}{4}$ if it is not mentioned in Theorem 3.1,*
- (ii) *at any cost level $c > 0$ if it has a component which induces a bipartite subnetwork with $m \in \mathbb{N}$ players on one side and less than $\frac{m}{3}$ on the other.*

Let us comprehend this. According to Corollary 3.3 any network not mentioned in Theorem 3.1 can at most be singularly pairwise stable. Moreover, there must be a player $i \in N$ who receives a payoff of $x = \frac{1}{2} - c$. This player can save costs of c when deleting a link while falling back to a payoff of zero in the worst case. This leads to the stability condition

$$x - \eta_i(g)c \geq 0 - (\eta_i(g) - 1)c \Leftrightarrow x \geq c \Leftrightarrow \frac{1}{2} - c \geq c \Leftrightarrow c \leq \frac{1}{4}.$$

Beyond that, note that a component-induced subnetwork g as mentioned in Part (ii) can only be contained in a pairwise stable network if the algorithm $\mathcal{A}(g)$ stops after the first step.⁶⁰ Let $(r_1, x, M_1, L_1, N_1, g_1)$ denote its outcome. By assumption we obviously have $|M_1| = m$ and $|L_1| < \frac{m}{3}$ but on the contrary we get

$$\frac{1}{4} \geq c = \frac{1}{2} - x = \frac{1}{2} - \frac{|L_1|}{|M_1| + |L_1|} = \frac{1}{2} \frac{|M_1| - |L_1|}{|M_1| + |L_1|} \Leftrightarrow 3|L_1| \geq |M_1|.$$

Arriving at a contradiction this gives Part (ii). Recall that, with regard to networks inducing heterogeneous payoffs within a component, the assumption of bipartiteness is not an additional restriction here according to Corollary 3.2. However, the main insight we gain from Corollary 3.4 is that Theorem 3.1 does not solely give a complete characterization of non-singularly pairwise stable networks for all $c > 0$ (recall Corollary 3.3) but also a complete characterization of pairwise stable networks for all $c > \frac{1}{4}$.

So far we considered pure one link deviations and the concept of pairwise stability. To conclude the main part of this section, we now relax this assumption to some extent and allow for multiple link deletion as it is captured by the notion of pairwise Nash stability (see Section 3.2). It is clear that every pairwise Nash stable network is also pairwise stable whereas the reverse is in general not true. This gives rise to the question whether in our model there exist pairwise stable networks which are not pairwise Nash stable. However, it is quite obvious that, at least for networks being non-singularly pairwise stable, this is not the case.

⁶⁰See again Footnote 56.

Corollary 3.5 (Pairwise Nash Stability). *In the framework with $\delta \rightarrow 1$ consider a non-singularly pairwise stable network g . Then g is pairwise Nash stable at each cost level for which it is pairwise stable.*

Note that, according to our previous results, in a non-singularly pairwise stable network, a player can only have more than one link if she is contained in an odd circle. Obviously, the definitions of pairwise Nash stability and pairwise stability differ only for such players. Recall that odd circles can only occur in networks being pairwise stable at linking costs $c \leq \frac{1}{6}$ and that each player contained in such a circle component receives a payoff of $\frac{1}{2}$. Hence, any such player must receive a profit of at least $\frac{1}{6}$. On the contrary, multiple link deletion would lead to a profit of zero since the player would be isolated afterwards.

Combining Corollaries 3.4 and 3.5 then gives that we even have a general equivalence between pairwise stability and pairwise Nash stability for linking costs $c > \frac{1}{4}$. Finally, observe that the line of length three is also even pairwise Nash stable at $c = \frac{1}{6}$. This is because the central player who receives a profit of $\frac{1}{3}$ would get isolated when deleting both of her links. This then would again induce a lower profit of zero.

3.3.1 Stability and Bargaining Outcomes

After characterizing (non-singularly) pairwise stable networks we now turn to see what our findings imply for outcomes of the infinite horizon network bargaining game. As a second main result we show that payoffs and profits induced by (non-singularly) pairwise stable networks are in general highly but not completely homogeneous. However, given our previous results of this section this can be stated as a corollary.

Corollary 3.6 (Limited Outcome Diversity). *In the framework with $\delta \rightarrow 1$ consider a network g which is pairwise stable at a given level of linking costs $c > 0$. Then players' payoffs must be such that either $v_i^*(g) \in \{\frac{1}{2} - c, \frac{1}{2}, \frac{1}{2} + c\}$ with $c \in (0, \frac{1}{4}]$ or $v_i^*(g) \in \{0, \frac{1}{2}\}$ for all $i \in N$. Moreover, if g is non-singularly pairwise stable, then only the latter of these two cases can occur and there exists a set $P(g) \subset \{0, \frac{1}{2} - 2c, \frac{1}{2} - c\}$ with $|P(g)| \leq 2$ such that for players' profits it holds that $u_i^*(g) \in P(g)$ for all $i \in N$.*

This result is basically an immediate consequence of Theorems 3.2 and 3.3, Corollaries 3.1 and 3.4, and Lemma 3.1. To see this, recall that in pairwise stable networks, in terms of payoffs, there can only occur four kinds of players. Namely, these are isolated players receiving zero, players belonging to a separated pair or an odd circle with a payoff of $\frac{1}{2}$ and, for $c \in (0, \frac{1}{4}]$, players contained in a component with heterogeneous payoffs who receive $\frac{1}{2} + c$ or $\frac{1}{2} - c$. However, the second part of Lemma 3.1

implies that the former and the latter player type cannot coexist in a pairwise stable network. Further, as every non-singularly pairwise stable network is a union of isolated players, odd circles and separated pairs, the only possible profits are zero, $\frac{1}{2} - 2c$ and $\frac{1}{2} - c$ in this case. However, for any cost level $c > 0$, only two of these three kinds of component-induced subnetworks can coexist in a pairwise stable network according to Corollary 3.1(iv) and (v).

Taken together, we have that the diversity of possible bargaining outcomes gets narrowed down substantially compared to the work of Manea (2011) if one considers a preceding stage of strategic network formation. To this end observe that in Manea's (2011) basic framework with $\delta \rightarrow 1$ one can obtain any rational number from the interval $[0, 1)$ as a payoff induced by an appropriate network on a sufficiently large player set.⁶¹

3.3.2 Singular Pairwise Stability

As already announced, we conclude this section by ruling out the possibility to be pairwise stable at all for certain network structures not considered yet. According to our previous results any remaining network can at most be singularly pairwise stable and there must be a component in which players receive heterogeneous payoffs (recall Theorem 3.3). In the following propositions, similarly as in Corollary 3.4, we consider specific classes of networks of that kind. The main idea of the proofs, which are rather lengthy and therefore again given in the appendix, is to identify generic network positions in which the respective player receives a payoff strictly greater than $\frac{1}{2}$ and still does so after deleting a certain link. Applying the notation of Theorem 3.3 we then must have $x + c < \frac{1}{2}$ for the corresponding stability condition to be fulfilled. Thus, arriving at a contradiction, such a network cannot be pairwise stable. Another approach we use focuses on players who are in a weak bargaining position but whose loss in payoff from dropping a certain own link is too small to be compatible with the condition $x + c = \frac{1}{2}$.

We show first that all networks not considered in Theorem 3.1 which contain a tree cannot even be singularly pairwise stable.⁶²

Proposition 3.2 (Trees). *Consider the framework with $\delta \rightarrow 1$. If a network g is pairwise stable, then it cannot have a component of more than three players which*

⁶¹For the rational number $\frac{n'}{n''} \in [0, 1)$ with $n', n'' \in \mathbb{N}$, consider the player set N with $n = n''$ and the complete bipartite network with n' players on one side and $n'' - n'$ players on the other side. Then the algorithm $\mathcal{A}(\cdot)$ yields payoffs $\frac{n'' - n'}{n''}$ and $\frac{n'}{n''}$.

⁶²A tree denotes a component-induced subnetwork which is minimally connected.

induces a tree.

Proposition 3.2 further reduces the class of potentially pairwise stable networks. It implies that any component of a pairwise stable network either contains at most three players or induces a subnetwork which has a cycle.⁶³ The former case has been analyzed exhaustively in Theorem 3.1 and Corollary 3.1. Thus, the only structures which are not captured by our analysis yet are networks which have a cycle and in which players receive heterogeneous payoffs. A subclass of such networks is considered in the following proposition. To state it we require an additional piece of notation. For a given network g , a player $k \in N$ is called *cut-player* if $g|_{N \setminus \{k\}}$ has more components than g .⁶⁴

Proposition 3.3 (Cycles and Cut-Players). *Consider the framework with $\delta \rightarrow 1$. If a network g is pairwise stable, then there cannot be a cut-player who is part of a cycle and receives a payoff strictly greater than $\frac{1}{2}$.*

This statement might seem somewhat artificial but it rules out the possibility to be pairwise stable for several generic kinds of networks. For instance, many networks containing a component-induced subnetwork which has a cycle and a loose-end player, i.e. there is a player who has one link, are excluded. See Figure 3.2 for an illustration of exemplary subnetworks which cannot even be contained in singularly pairwise stable networks according to Proposition 3.3.

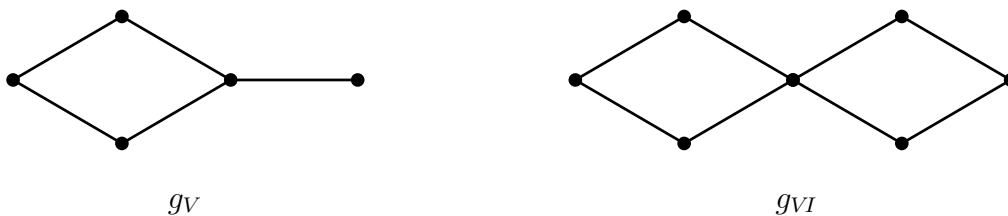


Figure 3.2: A sketch of networks which cannot be pairwise stable according to Proposition 3.3

However, there exist other networks not captured by our (explicitly stated) results which could potentially still be singularly pairwise stable at some cost level $c \in (0, \frac{1}{4}]$. Two examples for this are given in Figure 3.3.

⁶³A network g is said to have a cycle if there exist distinct players $i_1, i_2, \dots, i_{\bar{m}} \in N$, $\bar{m} \geq 3$, such that $i_1 i_{\bar{m}} \in g$ and $i_m i_{m+1} \in g$ for $m = 1, 2, \dots, \bar{m} - 1$. For instance, this implies that any network containing a circle has a cycle.

⁶⁴This notation comes from graph theory where vertices of that kind are typically called “cut-vertices” (see e.g. West, 2001). For instance, each player contained in a component which induces a tree and having more than one link is a cut-player.

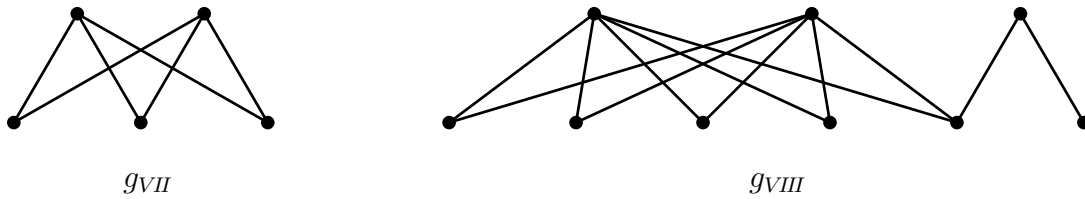


Figure 3.3: A sketch of networks which, based on our general results in Section 3.3, might still be singularly pairwise stable

Though a further generalization is not reached here, it is easy to check that the concrete networks illustrated in the figure cannot be pairwise stable. In g_{VII} one can obviously delete any link without changing payoffs and, for instance as an immediate consequence of Manea (2011, Theorem 6), a network like g_{VIII} is not pairwise stable at any cost level either. It remains as a conjecture that, in our framework with $\delta \rightarrow 1$, the only singularly pairwise stable networks inducing heterogeneous payoffs within a component are the ones containing a line of length three at cost level $c = \frac{1}{6}$.

3.4 Efficiency

Beside the issue of stability, it is of interest to analyze our model of network formation with regard to efficiency. From the perspective of a social planner it is important to understand the connection between pairwise stable network structures on the one hand and efficient ones on the other. In this light, Polanski and Vega-Redondo (2013) argue that the discrepancy between pairwise stability and efficiency in their model is due to the ex-ante heterogeneity between players. Throughout this section we establish that, in general, the two classes of pairwise stable and efficient networks do not coincide either in our model though players are assumed to be ex ante homogeneous. Our analysis is based on the concept of utilitarian welfare which postulates that a society's (or player set's) welfare is simply given by the sum of players' individual profits (recall Definition 3.4).⁶⁵ The following theorem gives a complete characterization of efficient networks for all levels of positive linking costs.

Theorem 3.4 (Efficiency). *Consider the framework with $\delta \rightarrow 1$ and player set N . For linking costs $c > \frac{1}{2}$, the empty network is the only efficient one. For $c = \frac{1}{2}$, a*

⁶⁵Note that we solely consider profits of the initial players here. One might argue that this is somewhat short-sighted but it is these players who are present today, i.e. at time $t = 0$ which we focus on throughout this work. Also, these are the players who are in charge of forming the network and who are therefore in a crucial position. Moreover, recall that it is in general uncertain whether or when players will get replaced during the subsequent network bargaining game. For these reasons, a social planner might also restrict attention to the initial period.

network is efficient if and only if it is a union of isolated players and separated pairs. And for $c \in (0, \frac{1}{2})$, a network is efficient if and only if it is a union of separated pairs, in case that n is odd supplemented by

- an isolated player if $c \in [\frac{1}{6}, \frac{1}{2})$,
- a line of length three if $c \in [\frac{1}{12}, \frac{1}{6}]$, or
- a three-player circle if $c \in (0, \frac{1}{12}]$.

The key insight of this theorem is that the empty network is the only efficient one while costs are high whereas for low costs any efficient network is (essentially) composed of separated pairs. In what follows, we therefore direct attention to the situation in which we have an even number of players in the player set. Nevertheless, a proof for the case n odd is as usual provided in the appendix. For both cases we require an additional piece of notation. Given a network g , we define the set $\bar{N}(g) := \{i \in N \mid \eta_i(g) \geq 1\}$, that is the set of players who are not isolated in g . Moreover, for a set of players $N' \subseteq N$ with $|N'|$ even, let $g_{N'}^{SP}$ denote a network composed of $\frac{|N'|}{2}$ separated pairs and $|N| - |N'|$ isolated players.⁶⁶

Proof of Theorem 3.4 (for n even). Consider a network g and furthermore let $(r_s, x_s, M_s, L_s, N_s, g_s)_{s=1, \dots, \bar{s}}$ be the outcome of the algorithm $\mathcal{A}(g)$ with $N_1 = \bar{N}(g)$. This means that we consider $\bar{N}(g)$ instead of N as player set here (thus, again disregarding isolated players which is w.l.o.g. as their profit is zero anyway).⁶⁷ Moreover, note that for any $y, z \in \mathbb{R}$ we have $y \cdot z \leq \frac{1}{4}(y+z)^2$ and that this holds strictly as long as $y \neq z$. Using this we calculate

$$\begin{aligned}
 U^*(g) &= \sum_{i \in \bar{N}(g)} (v_i^*(g) - \eta_i(g)c) = \sum_{i \in \bar{N}(g)} v_i^*(g) - 2d^\#(g)c \\
 &= \sum_{s=1}^{\bar{s}-1} (x_s|M_s| + (1-x_s)|L_s|) + \frac{1}{2}|N_{\bar{s}}| - 2d^\#(g)c \\
 &= 2 \sum_{s=1}^{\bar{s}-1} \frac{|M_s||L_s|}{|M_s| + |L_s|} + \frac{1}{2}|N_{\bar{s}}| - 2d^\#(g)c \\
 &\leq \frac{1}{2} \sum_{s=1}^{\bar{s}-1} (|M_s| + |L_s|) + \frac{1}{2}|N_{\bar{s}}| - 2d^\#(g)c \\
 &= \frac{1}{2}|\bar{N}(g)| - 2d^\#(g)c. \tag{3.5}
 \end{aligned}$$

⁶⁶Strictly speaking, there exist of course many networks of this kind. However, since any two of these can be converted into each other by permuting the ex ante homogeneous players, they are all payoff respectively welfare equivalent.

⁶⁷Alternatively, one might think of this as considering the network $g|_{\bar{N}(g)}$ with the usual restriction on the player set.

Since it is $\eta_i(g) \geq 1$ for all $i \in \bar{N}(g)$, we have that $d^\#(g) \geq \frac{1}{2}|\bar{N}(g)|$. Moreover, if $d^\#(g) = \frac{1}{2}|\bar{N}(g)|$, then this obviously implies that $|\bar{N}(g)|$ is even and we must have $g = g_{\bar{N}(g)}^{SP}$. Hence, according to (3.5), for a network $g \neq g_{\bar{N}(g)}^{SP}$ it holds that

$$U^*(g) < \frac{1}{2}|\bar{N}(g)| - |\bar{N}(g)|c = |\bar{N}(g)|\left(\frac{1}{2} - c\right) \leq \begin{cases} 0 = U^*(g_\emptyset^{SP}) = U^*(\emptyset) & \text{if } c \geq \frac{1}{2} \\ n(\frac{1}{2} - c) = U^*(g_N^{SP}) & \text{if } c \leq \frac{1}{2} \end{cases}.$$

This means that only networks which are unions of separated pairs and isolated players can be efficient. Among these candidates, it remains to identify the networks which induce maximal utilitarian welfare. Obviously, for $c > \frac{1}{2}$ this is solely the network g with minimal $|\bar{N}(g)|$, namely the empty network, whereas for $c \in (0, \frac{1}{2})$ we have to choose the one with maximal $|\bar{N}(g)|$, namely g_N^{SP} . For $c = \frac{1}{2}$ all candidates yield the same welfare of zero. \square

A comparison of Theorem 3.4 with the results established in Section 3.3 reveals some interesting insights concerning the relationship between efficient and pairwise stable networks. We summarize these in the following corollary.

Corollary 3.7 (Efficiency vs. Pairwise Stability). *In the model with $\delta \rightarrow 1$ it applies*

- (i) *for $c > \frac{1}{4}$ that a network is efficient if and only if it is pairwise stable,*
- (ii) *for $c \in (\frac{1}{6}, \frac{1}{4}]$ that a network is efficient if and only if it is non-singularly pairwise stable,*
- (iii) *for $c \in [\frac{1}{12}, \frac{1}{6}]$ that there exists both efficient networks being not pairwise stable and pairwise stable networks being not efficient and*
- (iv) *for $c \in (0, \frac{1}{12})$ that every efficient network is also pairwise stable but there exist pairwise stable networks being not efficient.*

We can constitute that as long as linking costs are high enough, efficient and pairwise stable networks coincide. However, there is an intermediate cost level for which a statement is not possible at all whereas for low costs the efficient networks constitute a proper subset of the pairwise stable ones. This confirms the intuition that, as long as linking costs are relatively low, there might be incentives for players to implement individually beneficial but non-efficient outside options. For an illustration consider the exemplary networks sketched in Figure 3.4 which are efficient for certain cost ranges but not pairwise stable or vice versa.

Observe that for $c = \frac{1}{6}$ the network g_{IX} is efficient according to Theorem 3.4 but not pairwise stable (see Corollary 3.1(v) or Figure 3.1). The same is true for g_X and

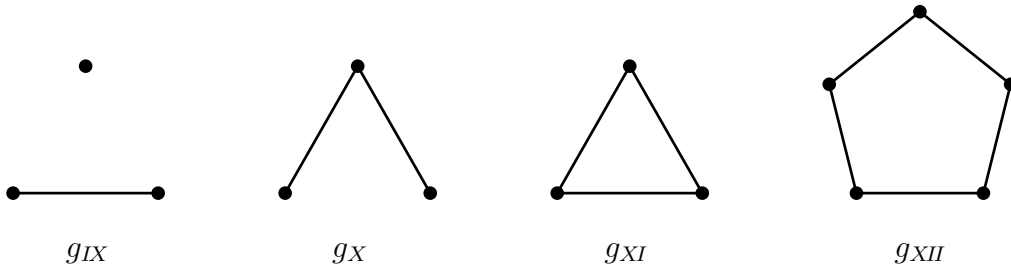


Figure 3.4: A sketch of networks revealing that efficiency and pairwise stability are in general not equivalent

the cost range $c \in [\frac{1}{12}, \frac{1}{6})$. On the other hand, g_{XI} is pairwise stable but not efficient for $c \in (\frac{1}{12}, \frac{1}{6}]$. And finally, for $c \in (0, \frac{1}{12}]$ the network g_{XII} is pairwise stable but circles containing more than three players are never efficient.

In summary, we find that efficiency does not in general coincide with pairwise stability although we deal with a setting of ex ante homogeneous players. Note, however, that the efficient networks are a subset of the pairwise stable ones at each level of linking costs $c > 0$ if we restrict our attention to player sets with an even number of players.

3.5 Effects of Time Discount

Our analysis in Sections 3.3 and 3.4 focuses on the limit case $\delta \rightarrow 1$ where players are assumed to be infinitely patient. However, in many situations it might be reasonable to consider players who are less than infinitely patient, meaning that in the network bargaining game they discount time at least to some degree. In the underlying model this is captured by a parametrization with $\delta \in (0, 1)$. In this section we take the limit case as a benchmark and reveal some important commonalities and differences between both cases with regard to strategic network formation and stability.

In Proposition 3.1 we establish that there are no disagreement links in pairwise stable networks if we have $\delta \rightarrow 1$. For two reasons it should be intuitively clear that this must still hold if we consider $\delta \in (0, 1)$ instead. On the one hand, if $ik \in g$ is a disagreement link, then we have $\delta v_i^{*\delta}(g) > 1 - \delta v_k^{*\delta}(g)$ by definition. Therefore, the i th equation of the system (3.1) determining the equilibrium payoffs is equivalent to

$$v_i = \left(1 - \sum_{j \in N_i(g-ik)} \frac{1}{2d^\#(g)} \right) \delta v_i + \sum_{j \in N_i(g-ik)} \frac{1}{2d^\#(g)} \max\{1 - \delta v_j, \delta v_i\}.$$

This means that from player i 's perspective it does not make a difference whether

she can get selected to bargain with player k or not since they will either way not find an agreement. This is of course similarly true from player k 's and also any other player's point of view. On the other hand, an additional amplifying effect comes into play when players are less than infinitely patient. In this case, players care about the time they have to wait until a certain outcome of a bargain is achieved as they discount these payments by δ when calculating their expected payoffs. The existence of a disagreement link prolongs the expected time until any other link gets selected in the bargaining game and, therefore, must have a negative impact on any player's payoff.

Next, we demonstrate that there exist networks which are pairwise stable for a certain level of linking costs if players are infinitely patient while this possibility can be ruled out if there is time discount. The converse turns out to be true as well.

Example 3.1. *Consider a player set with three players. Then a line of length three is pairwise stable if we have $\delta \rightarrow 1$ and $c = \frac{1}{6}$. In the framework with $\delta \in (0, 1)$, however, it is not pairwise stable for any $c > 0$.*

The first part of Example 3.1 is established by Theorem 3.1(iii). So let us examine why a line of length three cannot be pairwise stable if players are impatient to some degree. So let $N = \{1, 2, 3\}$ and let g be a line of length three. By applying the equation system (3.1) to g we find that the payoff of player 1, who is assumed to be the player having two links, is $v_1^{*\delta}(g) = \frac{2}{4-\delta}$. Further, the two loose-end players 2 and 3 receive $v_2^{*\delta}(g) = v_3^{*\delta}(g) = \frac{1}{4-\delta}$. Similarly, for the networks $g - 12$ and $g + 23$ we calculate $v_1^{*\delta}(g - 12) = \frac{1}{2}$, $v_2^{*\delta}(g - 12) = 0$ and $v_2^{*\delta}(g + 23) = v_3^{*\delta}(g + 23) = \frac{1}{3-\delta}$. Hence, for g to be pairwise stable, the following three conditions would have to be satisfied simultaneously.

$$u_2^{*\delta}(g) \geq u_2^{*\delta}(g - 12) \Leftrightarrow v_2^{*\delta}(g) \geq c \Leftrightarrow \frac{1}{4-\delta} \geq c, \quad (3.6a)$$

$$u_1^{*\delta}(g) \geq u_1^{*\delta}(g - 12) \Leftrightarrow v_1^{*\delta}(g) - v_1^{*\delta}(g - 12) \geq c \Leftrightarrow \frac{\delta}{2(4-\delta)} \geq c, \quad (3.6b)$$

$$u_2^{*\delta}(g) \geq u_2^{*\delta}(g + 23) \Leftrightarrow v_2^{*\delta}(g + 23) - v_2^{*\delta}(g) \leq c \Leftrightarrow \frac{1}{(3-\delta)(4-\delta)} \leq c. \quad (3.6c)$$

However, one can show by simple transformations that conditions (3.6b) and (3.6c) cannot be fulfilled at the same time. Figure 3.5 illustrates this. According to condition (3.6b) the level of costs must be below the blue line and (3.6c) requires that c is above the red line, which is obviously not possible simultaneously for $\delta < 1$.

Thus, either linking costs are so low that players 2 and 3 want to add a mutual link or they are so high that player 1 has incentives to delete one of her links. To

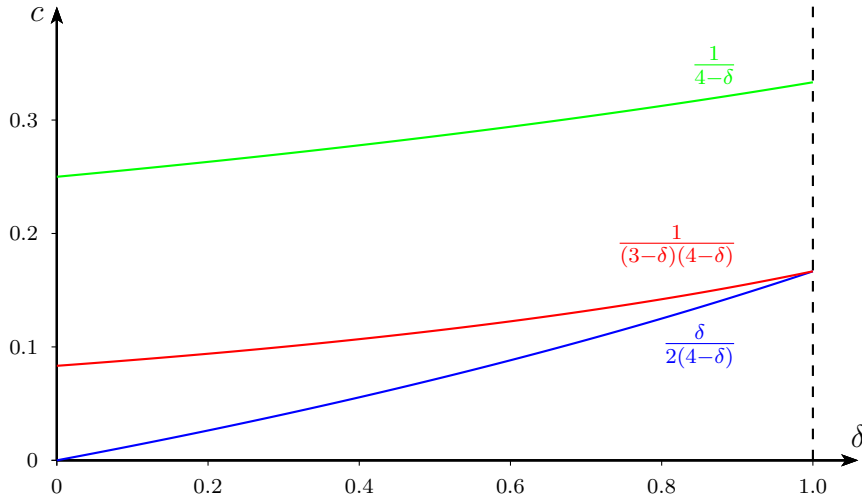


Figure 3.5: An illustration of the stability conditions (3.6) which arise in Example 3.1

sum up, we find that, in the framework with players being infinitely patient, the existence of lines of length three in pairwise stable networks is not robust in two respects. We already know that networks containing these subnetworks can at most be singularly pairwise stable. Additionally, we now have that a marginal decrease of δ , meaning however that players can still be almost infinitely patient, already causes general instability for this kind of networks.

On the contrary, there exist networks which, given any $\delta \in (0, 1)$, are pairwise stable at some level of linking costs whereas such a $c > 0$ does not exist if players are infinitely patient, that is if we consider $\delta \rightarrow 1$ instead.

Example 3.2. *Consider a player set with at least four players. Then for all $\delta \in (0, 1)$ there exists $\bar{c} > 0$ such that the complete network is pairwise stable for all $c \in (0, \bar{c}]$. In the framework with $\delta \rightarrow 1$, however, complete networks with more than three players are not pairwise stable for any $c > 0$.*

As usual, the proof is provided in the appendix. However, the second part should be clear, for instance by Theorem 3.2. To establish the first part we basically solve the equation system (3.1) for g^N and $g^N - ij$ and show that for sufficiently small costs it is profitable for any two players $i, j \in N$ to keep their mutual link.

In this context, note that the cost range for which the complete network of less than infinitely patient players is pairwise stable gets arbitrarily small and close to zero as δ approaches one. In this sense, the limit case $\delta \rightarrow 1$ does not constitute a discontinuity regarding our previous results as it might seem at first sight in the light of Example 3.2. Also, note that, in the framework of Manea (2011, Online Appendix)

with infinitely patient players and zero linking costs, complete networks are always pairwise stable as they are equitable.

3.6 Conclusion

In this chapter we develop a well-founded and analytically tractable model of strategic network formation in the context of decentralized bilateral bargaining involving ex ante homogeneous players and explicit linking costs. One reasonable application of our model is constituted by the stylized example of project collaboration between firms which we introduce at the beginning.

In the case that players are infinitely patient, we derive a complete characterization of non-singularly pairwise (Nash) stable networks. Depending on the level of linking costs, specific unions of separated pairs, odd circles and isolated players constitute this class. For a sufficiently high cost level our result even yields a complete characterization of pairwise (Nash) stable networks. The induced bargaining outcomes are mostly homogeneous but a certain level of diversity regarding players' payoffs and profits can still occur. Besides, we study the remaining networks which could possibly be singularly pairwise stable and succeed in ruling out this possibility for a broad range of structures. These results are complementary to Manea (2011, Online Appendix). Furthermore, we provide a complete characterization of networks being efficient in terms of a utilitarian welfare criterion and reveal that these coincide only partially, that is only at some cost levels, with the (non-singularly) pairwise stable ones. As a robustness check we also relax the assumption that players are infinitely patient and gain insights regarding commonalities and differences between the two cases.

Altogether, our work contributes to a better understanding of the behavior of players in a non-cooperative setting of decentralized bilateral bargaining when the underlying network is not exogenously given but the outcome of preceding strategic interaction. We gain insights concerning the structure of resulting networks, induced bargaining outcomes and regarding the effects which influence players when aiming for an optimization of their bargaining position in the network.

Regarding future research, it would be a reasonable next step, in our framework with infinitely patient players, to approach a complete characterization of pairwise stable networks in general, that is beyond the case of non-singular pairwise stability and for all levels of linking costs. This would call for a further discussion of networks which, according to our results, might be singularly pairwise stable for small

costs. For this purpose, it might be a promising approach to strive for a generalization of Manea's (2011) Theorem 6 to the case where the "buyer-seller ratio" is not (necessarily) an integer. If such an extension or generalization of our results is not possible, one could alternatively work towards a generalization of Example 3.1 as an additional robustness check. Anyway, it could be enriching to thoroughly analyze the class of stable and efficient networks when allowing players to discount time to some degree. A consideration of alternative stability concepts such as pairwise stability with transfers, which seems quite natural in a bargaining context, could generate further important insights. Beyond that, it would surely be interesting to set up an analytically tractable model of network formation in a bargaining framework in which players do not get replaced one-to-one after dropping out. Due to the resulting stochastic change of the network structure over time, this would certainly constitute a challenging research topic.

Appendix 3.A Proofs

3.A.1 Proof of Theorem 3.1

Consider a component $C \subseteq N$ of some network g which induces a circle or a separated pair. Then in both cases it is impossible to find a g -independent subset $M \subseteq C$ such that for the corresponding partner set we have $|L^g(M)| < |M|$. Hence, the algorithm $\mathcal{A}(g)$ yields a payoff of $\frac{1}{2}$ for each player contained in C in both cases (recall Definition 3.1 and the subsequent payoff computation rule). Now consider two players $i, j \in N$ with $ij \notin g$ where

- (a) both are contained in the same component inducing an odd circle,
- (b) they are contained in different components each inducing an odd circle,
- (c) they are contained in different components each inducing a separated pair,
- (d) one is contained in a component inducing an odd circle and the other one is contained in a component inducing a separated pair, or
- (e) one is contained in a component inducing an odd circle and the other one is an isolated player.

Then in each of these cases the algorithm $\mathcal{A}(g + ij)$ again yields a payoff of $\frac{1}{2}$ for all players contained in the new component $C_i(g + ij) = C_j(g + ij)$. The best way to see

this is again to consider Definition 3.1 and the subsequent payoff computation rule.⁶⁸ Therefore, at least one of the two players i and j (in Cases (a)–(d) even both) will receive an unchanged payoff after having established this link. Regarding profits this means, however, that this player is worse off as she has to bear additional linking costs of $c > 0$. Thus, the respective link will never be formed.

Next, recall the three-player case. From this it is straightforward to see that Part (i) of the theorem is indeed true. Also, we can deduce that a pairwise stable network can contain both an isolated player and a separated pair if we have linking costs $c \in (\frac{1}{6}, \frac{1}{2}]$. Together with the above Case (c) this establishes Part (ii) of the theorem.

Consider again a network g and now two players $i', j' \in N$ with $i'j' \in g$. Moreover, assume that these players are contained in a component C which induces an odd circle with $m \geq 3$ players. We already know that g induces a payoff of $\frac{1}{2}$ for both players. Now consider the network $g' := g - i'j'$ which is obviously a line of length m . Let $(r'_s, x'_s, M'_s, L'_s, N'_s, g'_s)_{s=1, \dots, \bar{s}'}$ be the outcome of $\mathcal{A}(g')$ (with $N_1 = C$).⁶⁹ As m is an odd number, we have that $\bar{s}' = 1$ and $i', j' \in M'_1$. Further, it is $|M'_1| = \frac{m+1}{2}$ and $|L'_1| = \frac{m-1}{2}$ which implies that $v_{i'}^*(g') = v_{j'}^*(g') = x'_1 = \frac{m-1}{2m}$. As a stability condition this gives

$$u_{i'}^*(g) - u_{i'}^*(g') = \frac{1}{2} - 2c - \left(\frac{m-1}{2m} - c \right) \geq 0 \Leftrightarrow \frac{1}{2m} \geq c \Leftrightarrow m \leq \frac{1}{2c}.$$

Of course, the same holds for player j' . Together with the above Cases (a) and (b) this means that a network which is composed of odd circles is pairwise stable if and only if each circle has at most $\frac{1}{2c}$ members. Note that a pairwise stable network can therefore contain odd circles only if we have $c \leq \frac{1}{6}$ since a circle must have at least three members by definition.

Furthermore, observe that for the cost range $c \in (0, \frac{1}{6}]$ we have $\frac{1}{2} - c > 0$ which means that no player contained in a component inducing a separated pair has incentives to delete her link. This together with the above Cases (c)–(e) gives that, potentially besides one or several odd circles with a permissible number of players, a network being pairwise stable at $c \in (0, \frac{1}{6}]$ can contain separated pairs or one isolated player. As we know from the three-player case, however, an isolated player and a separated pair cannot coexist in a pairwise stable network at these levels of linking costs. This proves the first statement in Part (iii).

⁶⁸However, a shortcut would be to consider Manea (2011, Theorem 5) which we make use of when proving our Theorem 3.2.

⁶⁹Disregarding players in C^c is w.l.o.g. as the profile of payoffs respectively profits is component-decomposable.

Finally, consider the transition point $c = \frac{1}{6}$. In what follows, let the network g be composed of two lines of length three, an odd circle, and a separated pair as sketched in Figure 3.6.

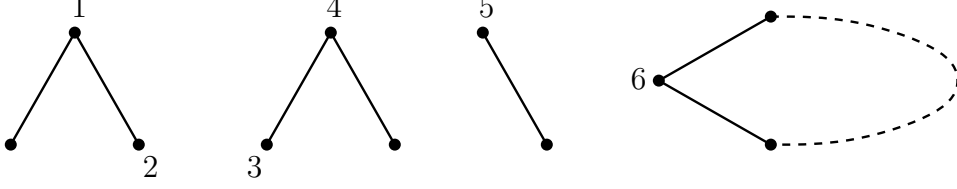


Figure 3.6: A sketch of a network g containing lines of length three

W.l.o.g. we focus on the labeled players 1, 2, ..., 6. At $c = \frac{1}{6}$ the algorithm $\mathcal{A}(g)$ yields the following profits:

$$u_1^*(g) = \frac{2}{3} - 2c = \frac{1}{3}, \quad u_2^*(g) = u_3^*(g) = \frac{1}{3} - c = \frac{1}{6}, \quad u_6^*(g) = \frac{1}{2} - 2c = \frac{1}{6}$$

Based on this, we are able to establish that link addition either leads to a worsening for at least one of the two players or both are indifferent. Specifically, applying the respective algorithm gives

$$\begin{aligned} u_2^*(g + 23) &= u_3^*(g + 23) = \frac{1}{2} - 2c = \frac{1}{6} = u_2^*(g) = u_3^*(g), \\ u_1^*(g + 13) &= u_1^*(g + 14) = u_1^*(g + 15) = u_1^*(g + 16) = \frac{2}{3} - 3c = \frac{1}{6} < \frac{1}{3} = u_1^*(g), \\ u_2^*(g + 25) &= \frac{2}{5} - 2c = \frac{1}{15} < \frac{1}{6} = u_2^*(g), \\ u_6^*(g + 26) &= \frac{1}{2} - 3c = 0 < \frac{1}{6} = u_6^*(g). \end{aligned}$$

Since we know from the three-player case that within the component of a line of length three there are no incentives to add or delete a link at this cost level, this concludes the proof of Part (iii) and of the whole theorem. \square

3.A.2 Proof of Proposition 3.1

For ease of notation consider a network g' and assume that it is pairwise stable. Moreover, assume that there is a disagreement link in the network, that is $g' \setminus g'^* \neq \emptyset$. Let w.l.o.g. $12 \in g' \setminus g'^*$ be such a link and define $g := g' - 12$. This implies $g'^* \subseteq g$. Furthermore, assume w.l.o.g. that every player has at least one link in g' (otherwise disregard isolated players which is permissible since the profile of payoffs respectively profits is component-decomposable). According to Manea (2011, Lemma 1) every player has at least one link in g'^* and therefore also in g .

Take the network g as a basis and let $(r_s, x_s, M_s, L_s, N_s, g_s)_{s=1,2,\dots,\bar{s}}$ be the outcome of $\mathcal{A}(g)$ (recall again Definition 3.1). Then by equations (3.3) the limit equilibrium payoff vector $v^*(g)$ is given by

$$\begin{aligned} v_i^*(g) &= x_s & \forall i \in M_s \forall s < \bar{s}, \\ v_j^*(g) &= 1 - x_s & \forall j \in L_s \forall s < \bar{s}, \\ v_k^*(g) &= \frac{1}{2} & \forall k \in N_{\bar{s}}. \end{aligned}$$

Now consider g^* . The following findings being equivalent to Manea (2011, Proposition 2, Theorem 3) are important:

- From Manea (2011, Proposition 2) we have that if $ij \in g$, then $v_i^*(g') + v_j^*(g') \geq 1$ and if $ij \in g^*$, then $v_i^*(g') + v_j^*(g') = 1$.
- By Manea (2011, Theorem 3) for all $M \in \mathcal{I}(g^*)$ the following bounds on limit equilibrium payoffs hold:

$$\begin{aligned} \min_{i \in M} v_i^*(g') &\leq \frac{|L^{g^*}(M)|}{|M| + |L^{g^*}(M)|} \\ \max_{j \in L^{g^*}(M)} v_j^*(g') &\geq \frac{|M|}{|M| + |L^{g^*}(M)|} \end{aligned}$$

If in Manea's (2011, Theorem 4) proof of the payoff computation rule (3.3) one replaces g^* by g' , v_i^* by $v_i^*(g')$, v_j^* by $v_j^*(g')$, v_k^* by $v_k^*(g')$, and Proposition 2, Lemma 1 and Theorem 3 (Manea, 2011) by the corresponding statements from above, then this leads to the result that also

$$\begin{aligned} v_i^*(g') &= x_s & \forall i \in M_s \forall s < \bar{s}, \\ v_j^*(g') &= 1 - x_s & \forall j \in L_s \forall s < \bar{s}, \\ v_k^*(g') &= \frac{1}{2} & \forall k \in N_{\bar{s}}. \end{aligned}$$

Thus, it is $v^*(g') = v^*(g)$ and hence

$$u_1^*(g') = v_1^*(g') - \eta_1(g')c = v_1^*(g) - (\eta_1(g) + 1)c < v_1^*(g) - \eta_1(g)c = u_1^*(g) - 12).$$

Arriving at a contradiction this proves that a pairwise stable network cannot contain a disagreement link.

Finally, notice that for any network g we have from Manea (2011, Proposition 2) that $v_i^*(g) + v_j^*(g) = 1$ for all $ij \in g^*$. Thus, the above result implies that $v_i^*(g) + v_j^*(g) = 1$ for all $ij \in g$ if g is pairwise stable. \square

3.A.3 Proof of Theorem 3.2

Consider a pairwise stable network g and assume that $g|_{\tilde{N}(g)}$ is not a union of separated pairs and odd circles. Notice that due to Proposition 3.1 for any component $C \subseteq N$ of g it must either be $C \subseteq \tilde{N}(g)$ or $C \subseteq \tilde{N}(g)^c$. Furthermore, recall that the profile of payoffs is component-decomposable, meaning that $v_i^*(g) = v_i^*(g|_{\tilde{N}(g)})$ for all $i \in \tilde{N}(g)$. Thus, the network $g|_{\tilde{N}(g)}$ is equitable such that by Manea (2011, Theorem 5) respectively by Berge (1981) it has a so called edge cover composed of separated pairs and odd circles. This means that there exists a union of separated pairs and odd circles $g' \subseteq g|_{\tilde{N}(g)}$ such that no player $i \in \tilde{N}(g)$ is isolated in g' . By assumption there must exist a link $ij \in g|_{\tilde{N}(g)} \setminus g'$. Obviously, the network g' is also an edge cover of the network $g|_{\tilde{N}(g)} - ij$. Again from Manea (2011, Theorem 5) respectively from Berge (1981) it then follows that $g|_{\tilde{N}(g)} - ij$ is still equitable. Hence, recalling the implication of Proposition 3.1 mentioned above, this gives

$$\begin{aligned} u_i^*(g) = v_i^*(g|_{\tilde{N}(g)}) - \eta_i(g|_{\tilde{N}(g)})c &= \frac{1}{2} - \eta_i(g|_{\tilde{N}(g)})c < \frac{1}{2} - (\eta_i(g|_{\tilde{N}(g)}) - 1)c \\ &= v_i^*(g|_{\tilde{N}(g)} - ij) - \eta_i(g|_{\tilde{N}(g)} - ij)c \\ &= u_i^*(g - ij). \end{aligned}$$

Thus, arriving at a contradiction, this concludes the proof. \square

3.A.4 Proof of Lemma 3.1 and Lemma 3.2

As announced in Section 3.3, the proofs of both lemmas rest on another rather technical lemma which we provide and prove first.

Lemma 3.3. *Let \tilde{g} be a network with $\mathcal{A}(\tilde{g})$ providing $(\tilde{r}_s, \tilde{x}_s, \tilde{M}_s, \tilde{L}_s, \tilde{N}_s, \tilde{g}_s)_s$. For any step $s < \bar{s}$ and any set $I \subseteq N$ the following implications must apply:*

$$\begin{aligned} (i) \quad 1 \leq |\tilde{M}_s \cap I| \leq |\tilde{L}_s \cap I| &\Rightarrow |L^{\tilde{g}_s}(\tilde{L}_s \cap I) \cap \tilde{M}_s \cap I^c| \geq 1, \\ (ii) \quad 1 \leq |\tilde{M}_s \cap I| < |\tilde{L}_s \cap I| &\Rightarrow |L^{\tilde{g}_s}(\tilde{L}_s \cap I) \cap \tilde{M}_s \cap I^c| \geq 2. \end{aligned}$$

Proof of Lemma 3.3. We prove the two parts of the lemma one after the other.

Part (i):

Assume that we have $1 \leq |\tilde{M}_s \cap I| \leq |\tilde{L}_s \cap I|$ and $L^{\tilde{g}_s}(\tilde{L}_s \cap I) \cap \tilde{M}_s \cap I^c = \emptyset$ for some

step $s < \bar{s}$ and some set $I \subseteq N$. Recalling Definition 3.1, this obviously implies

$$\frac{|\tilde{L}_s|}{|\tilde{M}_s|} = \tilde{r}_s < 1 \leq \frac{|\tilde{L}_s \cap I|}{|\tilde{M}_s \cap I|}.$$

Additionally, we have that $\tilde{M}_s = (\tilde{M}_s \cap I) \dot{\cup} (\tilde{M}_s \setminus I)$ and $\tilde{L}_s = (\tilde{L}_s \cap I) \dot{\cup} (\tilde{L}_s \setminus I)$. This induces that $\tilde{M}_s \setminus I \neq \emptyset$ since it is $|\tilde{M}_s \cap I| \leq |\tilde{L}_s \cap I| \leq |\tilde{L}_s|$ but $|\tilde{M}_s| > |\tilde{L}_s|$. It follows that

$$\frac{|\tilde{L}_s \setminus I|}{|\tilde{M}_s \setminus I|} < \frac{|\tilde{L}_s|}{|\tilde{M}_s|}.$$

Moreover, it is $L^{\tilde{g}_s}(\tilde{M}_s \setminus I) \subseteq \tilde{L}_s \setminus I$ since by assumption $L^{\tilde{g}_s}(\tilde{L}_s \cap I) \cap \tilde{M}_s \subseteq I$. Taken together, this then gives

$$\frac{|L^{\tilde{g}_s}(\tilde{M}_s \setminus I)|}{|\tilde{M}_s \setminus I|} \leq \frac{|\tilde{L}_s \setminus I|}{|\tilde{M}_s \setminus I|} < \frac{|\tilde{L}_s|}{|\tilde{M}_s|} = \tilde{r}_s,$$

which contradicts the minimality of \tilde{r}_s .

Part (ii):

It remains to show that having $1 \leq |\tilde{M}_s \cap I| < |\tilde{L}_s \cap I|$ and $|L^{\tilde{g}_s}(\tilde{L}_s \cap I) \cap \tilde{M}_s \cap I^c| = 1$ in some step $s < \bar{s}$ and for some set $I \subseteq N$ leads to a contradiction as well. In such a situation, slightly different from Part (i), we have

$$\frac{|\tilde{L}_s|}{|\tilde{M}_s|} = \tilde{r}_s < 1 \leq \frac{|\tilde{L}_s \cap I|}{|\tilde{M}_s \cap I| + 1}.$$

Again, it holds that $\tilde{M}_s = (\tilde{M}_s \cap I) \dot{\cup} (\tilde{M}_s \setminus I)$ and $\tilde{L}_s = (\tilde{L}_s \cap I) \dot{\cup} (\tilde{L}_s \setminus I)$, which in this case even guarantees that $|\tilde{M}_s \setminus I| \geq 2$ since it is $|\tilde{M}_s \cap I| < |\tilde{L}_s \cap I| \leq |\tilde{L}_s|$, but $|\tilde{M}_s| > |\tilde{L}_s|$. This gives

$$\frac{|\tilde{L}_s \setminus I|}{|\tilde{M}_s \setminus I| - 1} < \frac{|\tilde{L}_s|}{|\tilde{M}_s|}.$$

Moreover, we have that there exists exactly one player $\tilde{i} \in L^{\tilde{g}_s}(\tilde{L}_s \cap I) \cap \tilde{M}_s \cap I^c$. Similarly to Part (i) this implies that it is $L^{\tilde{g}_s}(\tilde{M}_s \setminus (I \dot{\cup} \{\tilde{i}\})) \subseteq \tilde{L}_s \setminus I$, which combined with the above leads to

$$\frac{|L^{\tilde{g}_s}(\tilde{M}_s \setminus (I \dot{\cup} \{\tilde{i}\}))|}{|\tilde{M}_s \setminus (I \dot{\cup} \{\tilde{i}\})|} \leq \frac{|\tilde{L}_s \setminus I|}{|\tilde{M}_s \setminus I| - 1} < \frac{|\tilde{L}_s|}{|\tilde{M}_s|} = \tilde{r}_s,$$

which again contradicts the minimality of \tilde{r}_s . \square

Now, we can turn to the proof of the first of the two lemmas which are stated in

Section 3.3.

Proof of Lemma 3.1. For $i, j \in M_1$ consider the network $g' := g + ij$. Let $(r'_s, x'_s, M'_s, L'_s, N'_s, g'_s)_{s=1, \dots, \bar{s}'}$ be the outcome of $\mathcal{A}(g')$. Assume for contradiction that there exists a step $\hat{s} \in \{1, \dots, \bar{s}' - 1\}$ such that $L_1 \cap M'_s = M_1 \cap L'_s = \emptyset$ for all $s \in \{1, \dots, \hat{s} - 1\}$ but $M_1 \cap L'_{\hat{s}} \neq \emptyset$. Note that $L_1 \cap M'_{\hat{s}} \neq \emptyset$ would also entail $M_1 \cap L'_{\hat{s}} \neq \emptyset$. This is because any player contained in $L_1 \cap M'_{\hat{s}}$ must have a neighbor $k \in M_1$ in g due to the minimality of $r_1 < 1$ and it can obviously neither be $k \in L'_s$ nor $k \in M'_s$ for any $s \in \{1, \dots, \hat{s} - 1\}$. In what follows, we construct a sequence of players (i_0, i_1, i_2, \dots) and show by induction that the underlying procedure which sequentially adds players to it can never break up so that we get a contradiction to the finiteness of the player set N . For $m \in \mathbb{N}$ let $I_m := \{i_0, i_1, \dots, i_m\} \subseteq N$ denote the players of the sequence up to the m th one. We need to distinguish two cases.

Case 1: $i \in L'_{\hat{s}}$

Set $i_0 = i$. It then must be $|N_{i_0}(g'_{\hat{s}}) \cap M'_{\hat{s}}| \geq 2$ since otherwise one could reduce $r'_{\hat{s}}$ by not including i_0 and possibly her one contact belonging to $M'_{\hat{s}}$. This guarantees that there exists $i_1 \in N_{i_0}(g'_{\hat{s}}) \cap M'_{\hat{s}} \setminus \{j\}$. So it is $i_0 \in M_1 \cap L'_{\hat{s}}$ and $i_1 \in L_1 \cap M'_{\hat{s}}$. Let $I_1 = \{i_0, i_1\}$. Now consider some odd number $m \in \mathbb{N}$. Assume that $L_1 \cap I_m = M'_{\hat{s}} \cap I_m$, $M_1 \cap I_m = L'_{\hat{s}} \cap I_m$ and that the cardinalities of these two sets are equal. We then have:

- It is $1 \leq |M_1 \cap I_m| = |L_1 \cap I_m|$ and therefore by Lemma 3.3(i) there exists a player $i_{m+1} \in L^g(L_1 \cap I_m) \cap M_1 \cap I_m^c$. For this player it must hold that $i_{m+1} \in M_1 \cap L'_{\hat{s}} \setminus I_m$ since $L_1 \cap I_m \subseteq M'_{\hat{s}}$ and $M_1 \cap L'_s = \emptyset$ for all $s \in \{1, \dots, \hat{s} - 1\}$.
- It then is $1 \leq |M'_{\hat{s}} \cap I_{m+1}| < |L'_{\hat{s}} \cap I_{m+1}|$ and therefore by Lemma 3.3(ii) there exists a player $i_{m+2} \in L^{g'_{\hat{s}}}(L'_{\hat{s}} \cap I_{m+1}) \cap M'_{\hat{s}} \cap I_{m+1}^c \setminus \{j\}$. For this player it must hold that $i_{m+2} \in L_1 \cap M'_{\hat{s}} \setminus I_{m+1}$ since $L'_{\hat{s}} \cap I_{m+1} \subseteq M_1$ and $i_{m+2} \neq j$.

Thus, it is $L_1 \cap I_{m+2} = M'_{\hat{s}} \cap I_{m+2}$, $M_1 \cap I_{m+2} = L'_{\hat{s}} \cap I_{m+2}$ and also the cardinalities of these two sets are equal. Moreover, it is $|I_{m+2}| = |I_m| + 2$. By induction it follows that the player set N must be infinitely large. Thus, we arrive at a contradiction.

Case 2: $i \notin L'_{\hat{s}}$

In this case we must have $j \notin M'_{\hat{s}}$ since by assumption $M_1 \cap L'_s = \emptyset$ for all $s \in \{1, \dots, \hat{s} - 1\}$. For the same reason, $i \in M'_{\hat{s}}$ would imply $j \in L'_{\hat{s}}$ which is equivalent to Case 1. This is also true for $i \notin M'_{\hat{s}}$ and $j \in L'_{\hat{s}}$. So it remains to consider the case that $i, j \notin (M'_{\hat{s}} \cup L'_{\hat{s}})$. However, by assumption there must be a player $i_0 \in M_1 \cap L'_{\hat{s}}$. As

in the previous case, existence of another player $i_1 \in N_{i_0}(g'_s) \cap M'_s$ is then guaranteed and it must be $i_1 \notin \{i, j\}$ since $i, j \notin M'_s$. Therefore it is $i_1 \in L_1 \cap M'_s$. Let again $I_1 = \{i_0, i_1\}$ and assume for some odd number $m \in \mathbb{N}$ that $L_1 \cap I_m = M'_s \cap I_m$, $M_1 \cap I_m = L'_s \cap I_m$ and that the cardinalities of these two sets are equal. Furthermore, assume that $i, j \notin I_m$. Similarly to the first case we have:

- There exists $i_{m+1} \in M_1 \cap L'_s \setminus I_m$ for the stated reasons.
- By Lemma 3.3(ii) there then exists a player $i_{m+2} \in L^{g'_s}(L'_s \cap I_{m+1}) \cap M'_s \cap I_{m+1}^c$. For this player it must hold that $i_{m+2} \in L_1 \cap M'_s \setminus I_{m+1}$ since $L'_s \cap I_{m+1} \subseteq M_1 \setminus \{i, j\}$.

Thus, it is again $L_1 \cap I_{m+2} = M'_s \cap I_{m+2}$, $M_1 \cap I_{m+2} = L'_s \cap I_{m+2}$ and also the cardinalities of these two sets are equal. Beyond that, we have $i, j \notin I_{m+2}$. By induction this leads again to a contradiction to the finiteness of the player set N .

Summing up, we have that $L_1 \cap M'_s = M_1 \cap L'_s = \emptyset$ for all $s < \bar{s}'$. Therefore, it must be $v_i^*(g'), v_j^*(g') \leq \frac{1}{2}$. On the contrary, we know by Manea (2011, Proposition 2) that $v_i^*(g') + v_j^*(g') \geq 1$. Taken together, this implies $v_i^*(g') = v_j^*(g') = \frac{1}{2}$.

With regard to the second part of the lemma consider the network $g' := g + i(n+1)$ and let $(r'_s, x'_s, M'_s, L'_s, N'_s, g'_s)_{s=1, \dots, \bar{s}'}$ be the outcome of $\mathcal{A}(g')$. It is clear that $n+1 \notin L'_s$ for all $s < \bar{s}'$ since otherwise one could simply reduce r'_s by deleting $n+1$ from L'_s and possibly her one neighbor i from M'_s . The possibility that $i \in L'_s$ for some $s < \bar{s}'$ can be ruled out by a line of argumentation which is equivalent to the proof of the first part if one substitutes $n+1$ for j , M_2 for M_1 and L_2 for L_1 (while taking into account that $\mathcal{A}(g)$ provides $M_1 = \{n+1\}$ and $L_1 = \emptyset$ in this case). \square

And finally we establish the second of the two lemmas.

Proof of Lemma 3.2. W.l.o.g. assume that g has only one component.⁷⁰ Beside g consider the network $g' := g - kl$ and let $(r'_s, x'_s, M'_s, L'_s, N'_s, g'_s)_{s=1, \dots, \bar{s}'}$ be the outcome of $\mathcal{A}(g')$. Similarly to the proof of Lemma 3.1 assume for contradiction that there exists a step $\hat{s} \in \{1, \dots, \bar{s}' - 1\}$ such that $L_1 \cap M'_s = M_1 \cap L'_s = \emptyset$ for all $s \in \{1, \dots, \hat{s} - 1\}$ but $L_1 \cap M'_s \neq \emptyset$. Observe that $M_1 \cap L'_s \neq \emptyset$ would also entail $L_1 \cap M'_s \neq \emptyset$ since due to the minimality of r'_s any player in $M_1 \cap L'_s$ needs to have a g' -neighbor in M'_s who then must have been a neighbor in g as well. We again construct a sequence of players (i_0, i_1, i_2, \dots) and show by induction that the underlying procedure which sequentially adds players to it can never break up, meaning that we get a contradiction to the

⁷⁰Again, this is w.l.o.g. as the profile of payoffs respectively profits is component-decomposable.

finiteness of the player set N . For $m \in \mathbb{N}$ let $I_m := \{i_0, i_1, \dots, i_m\} \subseteq N$ denote the players of the sequence up to the m th one.

Initially, select some player $i_0 \in L_1 \cap M'_s$. i_0 cannot be isolated or a loose-end player, i.e. she must have more than one link in g , since otherwise one could reduce r_1 by not including i_0 in L_1 and possibly her one contact in M_1 . This guarantees that there exists $i_1 \in N_{i_0}(g')$. It must be $i_1 \in M_1 \cap L'_s$ since by assumption $M_1 \cap L'_s = \emptyset$ for all $s \in \{1, \dots, \hat{s} - 1\}$. Let $I_1 = \{i_0, i_1\}$. Now consider some odd number $m \in \mathbb{N}$. Assume that $L_1 \cap I_m = M'_s \cap I_m$, $M_1 \cap I_m = L'_s \cap I_m$ and that the cardinalities of these two sets are equal. We then have:

- It is $1 \leq |M'_s \cap I_m| = |L'_s \cap I_m|$ and therefore by Lemma 3.3(i) there exists a player $i_{m+1} \in L^{g'_s}(L'_s \cap I_m) \cap M'_s \cap I_m^c$. For this player it must hold that $i_{m+1} \in L_1 \cap M'_s \setminus I_m$ since it is $L'_s \cap I_m \subseteq M_1$.
- Then it is $1 \leq |M_1 \cap I_{m+1}| < |L_1 \cap I_{m+1}|$ and therefore by Lemma 3.3(ii) there exists a player $i_{m+2} \in L^g(L_1 \cap I_{m+1}) \cap M_1 \cap I_{m+1}^c \cap L^{g'_s}(L_1 \cap I_{m+1})$ since g' arose from g by a single link deletion and, additionally, $M_1 \cap L'_s = \emptyset$ for all $s \in \{1, \dots, \hat{s} - 1\}$ and $L_1 \cap I_{m+1} \subseteq M'_s$. This reasoning then also implies that $i_{m+2} \in M_1 \cap L'_s \setminus I_{m+1}$.

Thus it is $L_1 \cap I_{m+2} = M'_s \cap I_{m+2}$, $M_1 \cap I_{m+2} = L'_s \cap I_{m+2}$ and also the cardinalities of these two sets are equal. Moreover, it is $|I_{m+2}| = |I_m| + 2$. Again, by induction this leads to a contradiction to the finiteness of the player set N . Consequently, it must be $L_1 \cap M'_s = M_1 \cap L'_s = \emptyset$ for all $s < \bar{s}$. \square

3.A.5 Proof of Proposition 3.2

Consider a network g which is a tree with $n > 3$ players and assume that it is pairwise stable.⁷¹ By Theorem 3.2 it cannot be the case that all players receive a payoff of $\frac{1}{2}$ in g . According to Proposition 3.1 and Corollary 3.2, the algorithm $\mathcal{A}(g)$ therefore has to stop after the first step providing an outcome $(r_1, x_1, M_1, L_1, N_1, g_1)$ with $M_1 \dot{\cup} L_1 = N$, $|M_1| > |L_1|$ and $g|_{M_1} = g|_{L_1} = \emptyset$. So we have $r_1 \in (0, 1)$ and $v_i^*(g) = 1 - v_j^*(g) = x_1 \in (0, \frac{1}{2})$ for all $i \in M_1$, $j \in L_1$. Theorem 3.3 then implies that

$$x_1 + c = \frac{1}{2}. \quad (3.7)$$

⁷¹Again, it is w.l.o.g. to assume that g consists of only one component as the profile of payoffs respectively profits is component-decomposable.

The class of tree networks we consider here can be divided into the following subclasses:

- (a) No player has more than two links in g , meaning that g is a line network.
- (b) There is a player who has more than two links in g such that at least two of her neighbors are loose-end players.⁷²
- (c) There is a player who has more than two links in g but no player has more than one loose-end contact.

In the following, we distinguish between these three subclasses and show separately that there arises a contradiction to pairwise stability.

Subclass (a):

W.l.o.g. let $g := \{12, 23, \dots, (n-1)n\}$. Here n must be odd since otherwise it would obviously be $\frac{|L^g(M)|}{|M|} \geq 1$ for all g -independent sets $M \subseteq N$ inducing a payoff of $\frac{1}{2}$ for every player. So by assumption it must be $n \geq 5$. Considering the algorithm $\mathcal{A}(g)$, we find that the shortage ratio is minimized by the g -independent set which contains the players $1, 3, \dots, n-2, n$. Therefore, it is $r_1 = \frac{n-1}{n+1}$ and $x = \frac{n-1}{2n}$. Hence, here equation (3.7) is equivalent to

$$c = \frac{1}{2n}. \quad (3.8)$$

Now, if player 3 deletes her link to player 2, then she becomes a loose-end player. Moreover, in the network $g-23$ she is contained in a component with an odd number of players which induces a line of length $n-2$. Hence, it is $v_3^*(g-23) = \frac{n-3}{2(n-2)}$. Taking into account equation (3.8), the corresponding stability condition yields

$$\begin{aligned} u_3^*(g) - u_3^*(g-23) \geq 0 &\Leftrightarrow v_3^*(g) - v_3^*(g-23) - c \geq 0 \\ &\Leftrightarrow \frac{n-1}{2n} - \frac{n-3}{2(n-2)} - \frac{1}{2n} \geq 0 \\ &\Leftrightarrow \frac{4-n}{2n(n-2)} \geq 0. \end{aligned}$$

Obviously, this is never fulfilled for $n \geq 5$, meaning that a line network cannot be pairwise stable.

⁷²Recall that some $i \in N$ is said to be a loose-end player if it is $\eta_i(g) = 1$, that is if she has exactly one link in g .

Subclass (b):

Let $k \in N$ be a player with at least three neighbors including two or more loose-end players. Then Manea (2011, Theorem 3) implies that it is $v_k^*(g) \geq \frac{2}{3}$. So it must be $k \in L_1$. Select a player $i \in N_k(g)$ such that $\eta_i(g) \geq \eta_{i'}(g)$ for all $i' \in N_k(g)$. Note that in the network $g - ki$, player k still has at least two loose-end contacts such that again according to Manea (2011, Theorem 3) we have $v_k^*(g - ki) \geq \frac{2}{3}$. The corresponding stability condition then gives

$$u_k^*(g) \geq u_k^*(g - ki) \Leftrightarrow v_k^*(g) - c \geq v_k^*(g - ki) \Rightarrow 1 - x_1 - c \geq \frac{2}{3} \Leftrightarrow x_1 + c \leq \frac{1}{3}.$$

This obviously contradicts equation (3.7). Thus, a network g belonging to Subclass (b) cannot be pairwise stable.

Subclass (c):

First deliberate the following: For any tree network \tilde{g} and any player $k \in N$ there exists a unique partition $(Br_\nu^k)_{\nu \in N_k(\tilde{g})}$ of $N \setminus \{k\}$ such that for all $\nu \in N_k(\tilde{g})$ it is $\nu \in Br_\nu^k$ and $\tilde{g}|_{Br_\nu^k}$ is connected, i.e. $\tilde{g}|_{Br_\nu^k}$ has only one component (if one restricts the player set to Br_ν^k). Based on this observation, we define the subnetworks $(\tilde{g}|_{Br_\nu^k})_{\nu \in N_k(\tilde{g})}$ to be the *branches* of player k in \tilde{g} and $\nu \in N_k(\tilde{g})$ is said to be the *fork player* of $\tilde{g}|_{Br_\nu^k}$.

Note that if g belongs to Subclass (c), then there exists a player $k \in N$ who has more than two links such that for at least all but one of her branches, all players contained in these have at most two links in g . If this would not be the case, the following procedure would never stop, meaning that there would have to be infinitely many players in N : Initially, select a player k_0 having more than two links and one of her branches containing another player k_1 with more than two links. Then by assumption player k_1 must have a branch in g which does not contain player k_0 but a player k_2 who also has more than two links. For this player k_2 there must again be a branch in g not containing k_0 and k_1 but a player k_3 having more than two links. Continuing this way, for any $m \in \mathbb{N}$ there is a player $k_{m+1} \in N \setminus \{k_0, \dots, k_m\}$, which then gives a contradiction by induction. Thus, a player k as mentioned above must indeed exist.

In the following we distinguish two cases.

Case (c).1: $k \in L_1$

If there are other players having more than two links, then let $i \in N$ be the fork player of player k 's branch which contains all of them. Otherwise, arbitrarily pick some $i \in N_k(g)$. In both cases consider the network $g - ki$ and the component $C \subset N$ which player k is contained in. In the network $g|_C$, there is only player k who might

have more than two links. Furthermore, every branch of player k in $g|_C$ must be a line of odd length as Manea (2011, Theorem 3) implies that any loose-end player in g is contained in M_1 . In turn, this implies that for any $g|_C$ -independent set M with $\frac{|L^{g|_C}(M)|}{|M|} < 1$ it is $k \in L^{g|_C}(M)$. One example for such a set is $M_1 \cap C$ with partner set $L_1 \cap C$. Hence, it must be $v_k^*(g - ki) > \frac{1}{2}$. The corresponding stability condition then gives

$$u_k^*(g) \geq u_k^*(g - ki) \Leftrightarrow v_k^*(g) - c \geq v_k^*(g - ki) \Rightarrow 1 - x_1 - c > \frac{1}{2} \Leftrightarrow x_1 + c < \frac{1}{2}.$$

This obviously again contradicts equation (3.7). Consequently, a network g belonging to Subclass (c) with $k \in L_1$ cannot be pairwise stable.

Case (c).2:

We need to introduce some additional notation here. Identify a branch of player k which is a line network with minimal length among all of these line branches. We denote the set of players in this branch by $B^1 \subset N$. Note that any branch of player k which is a line must be of even length. Let $\hat{M}^1 := M_1 \cap B^1$ and $\hat{L}^1 := L_1 \cap B^1$. Then it is $|\hat{M}^1| = |\hat{L}^1|$. Let j denote the fork player of this branch. In addition, let $B^2 \subset N$ denote the set of all players contained in the other line branch(es) of player k . Let similarly $\hat{M}^2 := M_1 \cap B^2$ and $\hat{L}^2 := L_1 \cap B^2$. Then we have $|\hat{M}^2| = |\hat{L}^2| \geq |\hat{M}^1|$. Finally, let $B^3 := N \setminus (B^1 \cup B^2 \cup \{k\})$ and $\hat{M}^3 := M_1 \cap B^3$, $\hat{L}^3 := L_1 \cap B^3$. Then it must be $|\hat{M}^3| \geq |\hat{L}^3|$ as we have $|M_1| > |L_1|$.

Note that we must have $r_1 = \frac{|L_1|}{|M_1|} \leq \frac{|\hat{L}^3|}{|\hat{M}^3|}$ since r_1 is the minimal shortage ratio for g and obviously $L^g(\hat{M}^3) = \hat{L}^3$. Thus, applying the above notation gives

$$x_1 = \frac{|L_1|}{|M_1| + |L_1|} = \frac{|\hat{M}^1| + |\hat{M}^2| + |\hat{L}^3|}{2|\hat{M}^1| + 2|\hat{M}^2| + |\hat{M}^3| + |\hat{L}^3| + 1}. \quad (3.9)$$

Now consider the network $g' := g - kj$ and let $(r'_s, x'_s, M'_s, L'_s, N'_s, g'_s)_{s=1, \dots, \bar{s}}$ be the outcome of the algorithm $\mathcal{A}(g')$. Notice first that the set $\hat{M}^2 \cup \hat{M}^3 \cup \{k\} \subset M_1$ is g' -independent and $\hat{L}^2 \cup \hat{L}^3$ is the corresponding partner set in g' . Furthermore, we have

$$\frac{|\hat{L}^2| + |\hat{L}^3|}{|\hat{M}^2| + |\hat{M}^3| + 1} = \frac{|\hat{M}^2| + |\hat{L}^3|}{|\hat{M}^2| + |\hat{M}^3| + 1} < 1.$$

Assume for contradiction that there is another g' -independent set $M' \subseteq N$ with partner set $L' = L^{g'}(M') \subseteq N$ which is shortage ratio minimizing in step $s = 1$ of $\mathcal{A}(g')$. Since the set B^1 is a component of g' and it induces a line network of even length where every player receives a payoff of $\frac{1}{2}$, we must have $(M' \cup L') \cap B^1 = \emptyset$

and $\bar{s}' \geq 2$. Moreover, Lemma 3.2 yields that $M_1 \cap L'_s = L_1 \cap M'_s = \emptyset$ for all $s < \bar{s}'$. Hence, we must have $M' \subset \hat{M}^2 \dot{\cup} \hat{M}^3 \dot{\cup} \{k\}$ and $L' \subset \hat{L}^2 \dot{\cup} \hat{L}^3$ such that

$$\frac{|L'|}{|M'|} < \frac{|\hat{M}^2| + |\hat{L}^3|}{|\hat{M}^2| + |\hat{M}^3| + 1} < 1.$$

On the contrary, $M' \dot{\cup} \hat{M}^1 \subset M_1$ is g -independent and we have $L^g(M' \dot{\cup} \hat{M}^1) = L' \dot{\cup} \hat{L}^1$. The minimality of $r_1 = \frac{|L_1|}{|M_1|}$ in $\mathcal{A}(g)$ then implies

$$r_1 = \frac{|\hat{M}^2| + |\hat{L}^3| + |\hat{M}^1|}{|\hat{M}^2| + |\hat{M}^3| + 1 + |\hat{M}^1|} \leq \frac{|L'| + |\hat{M}^1|}{|M'| + |\hat{M}^1|} < 1 \quad \Rightarrow \quad \frac{|\hat{M}^2| + |\hat{L}^3|}{|\hat{M}^2| + |\hat{M}^3| + 1} \leq \frac{|L'|}{|M'|}.$$

Thus, arriving at a contradiction, this implies that

$$v_k^*(g') = \frac{|\hat{M}^2| + |\hat{L}^3|}{2|\hat{M}^2| + |\hat{M}^3| + |\hat{L}^3| + 1}. \quad (3.10)$$

Taking into account equation (3.8), the corresponding stability condition demands

$$\begin{aligned} u_k^*(g) \geq u_k^*(g - kj) &\Leftrightarrow v_k^*(g) - \eta_k(g)c \geq v_k^*(g') - \eta_k(g')c \\ &\Leftrightarrow x_1 \geq v_k^*(g') + \frac{1}{2} - x_1 \\ &\Leftrightarrow 2x_1 - v_k^*(g') \geq \frac{1}{2} \end{aligned} \quad (3.11)$$

However, we now establish that it must be $2x_1 - v_k^*(g') < \frac{1}{2}$. Recalling equations (3.9) and (3.10), some calculation yields

$$\begin{aligned} 2x_1 - v_k^*(g') &= \frac{2|\hat{M}^1| + 2(|\hat{M}^2| + |\hat{L}^3|)}{2|\hat{M}^1| + (2|\hat{M}^2| + |\hat{M}^3| + |\hat{L}^3| + 1)} - \frac{(|\hat{M}^2| + |\hat{L}^3|)}{(2|\hat{M}^2| + |\hat{M}^3| + |\hat{L}^3| + 1)} \\ &= \frac{2|\hat{M}^1|(|\hat{M}^2| + |\hat{M}^3| + 1) + (|\hat{M}^2| + |\hat{L}^3|)(2|\hat{M}^2| + |\hat{M}^3| + |\hat{L}^3| + 1)}{2|\hat{M}^1|(2|\hat{M}^2| + |\hat{M}^3| + |\hat{L}^3| + 1) + (2|\hat{M}^2| + |\hat{M}^3| + |\hat{L}^3| + 1)^2} \\ &= \frac{D - R}{2D}, \end{aligned}$$

where

$$D = 2|\hat{M}^1|(2|\hat{M}^2| + |\hat{M}^3| + |\hat{L}^3| + 1) + (2|\hat{M}^2| + |\hat{M}^3| + |\hat{L}^3| + 1)^2 > 0$$

and

$$\begin{aligned}
 R &= -2|\hat{M}^1||\hat{M}^3| + 2|\hat{M}^1||\hat{L}^3| - 2|\hat{M}^1| + 2|\hat{M}^2||\hat{M}^3| - 2|\hat{M}^2||\hat{L}^3| + 2|\hat{M}^2| + |\hat{M}^3|^2 \\
 &\quad + 2|\hat{M}^3| - |\hat{L}^3|^2 + 1 \\
 &= 2(\underbrace{|\hat{M}^2| - |\hat{M}^1|}_{\geq 0}) + 2(\underbrace{|\hat{M}^3| - |\hat{L}^3|}_{\geq 0})(\underbrace{|\hat{M}^2| - |\hat{M}^1|}_{\geq 0}) + (\underbrace{|\hat{M}^3|^2 - |\hat{L}^3|^2}_{\geq 0}) + 2|\hat{M}^3| + 1 \\
 &\geq 2|\hat{M}^3| + 1 \\
 &> 0.
 \end{aligned}$$

Hence, we indeed have

$$2x_1 - v_k^*(g - kj) = \frac{D - R}{2D} < \frac{1}{2}.$$

This concludes the proof for Subclass (c) and of the whole proposition. \square

3.A.6 Proof of Proposition 3.3

Consider a pairwise stable network g and assume that there is a cut-player $k \in N$ who is part of a cycle and receives a payoff $v_k^*(g) > \frac{1}{2}$. Assume w.l.o.g. that g has only one component. According to Proposition 3.1 and Corollary 3.2, the algorithm $\mathcal{A}(g)$ must stop after the first step providing an outcome $(r_1, x_1, M_1, L_1, N_1, g_1)$ with $M_1 \dot{\cup} L_1 = N$, $|M_1| > |L_1|$ and $g|_{M_1} = g|_{L_1} = \emptyset$. So we have $r_1 = \frac{|L_1|}{|M_1|} \in (0, 1)$, $k \in L_1$ and $v_k^*(g) = 1 - x_1$. Further, by Theorem 3.3 it is $x_1 + c = \frac{1}{2}$.

In what follows, we prove that player k can delete a certain link such that in the resulting network she still receives a payoff greater than $\frac{1}{2}$. To start with, note that by assumption there must be a set $K \subset N$ with $k \in K$ such that

- $L^g(K \setminus \{k\}) \subseteq K$,
- k is contained in a cycle in $g|_{K \cup \{k\}}$ and
- $g|_{K^c}$ has only one component (as usual, considering K^c as player set).

As g has only one component, it must be $k \in L^g(K \setminus \{k\})$, meaning that $N_k(g) \cap K \neq \emptyset$. Moreover, there exists $i' \in N_k(g) \setminus K$ such that k and i' belong to the same cycle in g . Now consider the network $g' := g - ki'$ and let $(r'_s, x'_s, M'_s, L'_s, N'_s, g'_s)_{s=1, \dots, \bar{s}'}$ be the outcome of $\mathcal{A}(g')$. Lemma 3.2 yields that $v_k^*(g') \geq \frac{1}{2}$. Assume for contradiction that we have $v_k^*(g') = \frac{1}{2}$, meaning that $k \in N'_{\bar{s}'}$.

Consider the set $C'_k := C_k(g'_{\bar{s}'|K}) = C_k(g|_{N'_{\bar{s}' \cap K})$, that is the component of player k in the network g restricted to the set $N'_{\bar{s}' \cap K}$. As a first step, we establish that it is

$$\frac{|L_1 \cap C'_k|}{|M_1 \cap C'_k|} = 1. \quad (3.12)$$

Note first that we have $N_k(g'|_K) \neq \emptyset$. Furthermore, it must be $N_k(g') \subseteq M_1 \cap N'_{\bar{s}'}$ as Lemma 3.2 yields $M_1 \cap L'_s = \emptyset$ for all $s < \bar{s}'$. This guarantees that $M_1 \cap C'_k \neq \emptyset$. Based on this, we can immediately rule out the possibility that the left-hand side of (3.12) is strictly smaller than one since $M_1 \cap C'_k$ is g' -independent and clearly $L^{g'_{\bar{s}'}}(M_1 \cap C'_k) \subseteq L_1 \cap C'_k$. So assume that the left-hand side of (3.12) is strictly greater than one. We make use of the following implication which we verify at the end of the proof:

$$\begin{aligned} |\hat{L}| = |\hat{M}| \geq 1 \text{ for } \hat{L} \subseteq L_1 \cap C'_k \setminus \{k\}, N_k(g) \cap K \subseteq \hat{M} \subseteq M_1 \cap C'_k \\ \Rightarrow L^{g'_{\bar{s}'}}(\hat{L}) \setminus \hat{M} \neq \emptyset \end{aligned} \quad (3.13)$$

We know that it is $\emptyset \neq N_k(g) \cap K \subseteq N_k(g') \subseteq N'_{\bar{s}'}$. Let $\hat{M}^0 := N_k(g) \cap K$ such that $\hat{M}^0 \subseteq M_1 \cap C'_k$. Hence, it must be $|L_1 \cap C'_k \setminus \{k\}| \geq |\hat{M}^0|$ since otherwise we would get

$$\frac{|L_1 \cap C'_k|}{|M_1 \cap C'_k|} \leq \frac{|L_1 \cap C'_k|}{|\hat{M}^0|} \leq 1,$$

that is a contradiction to our assumption. So select a set of players $\hat{L}^0 \subseteq L_1 \cap C'_k \setminus \{k\}$ with $|\hat{L}^0| = |\hat{M}^0|$. Note that \hat{M}^0 and \hat{L}^0 satisfy the conditions of implication (3.13).

Based on this, we can construct a sequence of players (j_1, j_2, j_3, \dots) in a certain way such that according to the previous considerations, the underlying procedure which sequentially adds players to the sequence can never break up. As in the proofs of Lemma 3.1 and 3.2, this leads to a contradiction to the finiteness of the player set N . Given such a sequence, let $\hat{M}^m := \{j_l \mid 1 \leq l \leq m, l \text{ odd}\} \cup \hat{M}^0$ and $\hat{L}^m := \{j_l \mid 1 \leq l \leq m, l \text{ even}\} \cup \hat{L}^0$ for $m \in \mathbb{N}$. Now consider some even number $m \in \mathbb{N} \cup \{0\}$. Assume that $\hat{L}^m \subseteq L_1 \cap C'_k \setminus \{k\}$, $N_k(g) \cap K \subseteq \hat{M}^m \subseteq M_1 \cap C'_k$ and $|\hat{L}^m| = |\hat{M}^m| \geq 1$. We then have:

- By implication (3.13) there exists $j_{m+1} \in L^{g'_{\bar{s}'}}(\hat{L}^m) \setminus \hat{M}^m$. For this player it must hold that $j_{m+1} \in M_1 \cap C'_k \setminus \hat{M}^m$ since $\hat{L}^m \subseteq L_1 \cap C'_k \setminus \{k\}$.
- Then there must exist $j_{m+2} \in L_1 \cap C'_k \setminus (\hat{L}^{m+1} \cup \{k\})$ since otherwise we would

have

$$1 < \frac{|L_1 \cap C'_k|}{|M_1 \cap C'_k|} \leq \frac{|\hat{L}^{m+1} \dot{\cup} \{k\}|}{|\hat{M}^{m+1}|} = 1.$$

Thus it is $\hat{L}^{m+2} \subseteq L_1 \cap C'_k \setminus \{k\}$, $N_k(g) \cap K \subseteq \hat{M}^{m+2} \subseteq M_1 \cap C'_k$ and $|\hat{L}^{m+2}| = |\hat{M}^{m+2}| = |\hat{L}^m| + 1 \geq 1$. By induction this leads to a contradiction to the finiteness of the player set N . This establishes equation (3.12), however, under the assumption of having $v_k^*(g') = \frac{1}{2}$.

During the second step we now use this to construct a concluding contradiction of similar kind arising from the assumption that $v_k^*(g') = \frac{1}{2}$. Here, we make use of the following implication:

$$|\tilde{L}| = |\tilde{M}| \geq 1 \text{ for } \tilde{L} \subseteq L_1 \cap N'_{\bar{s}'} \setminus K, \tilde{M} \subseteq M_1 \cap N'_{\bar{s}'} \setminus K \Rightarrow L^{g'_{\bar{s}'}}(\tilde{L}) \setminus (\tilde{M} \dot{\cup} K) \neq \emptyset \quad (3.14)$$

Its verification is postponed to the end of this proof as well. Note that by definition it is $\frac{|L^{g'_{\bar{s}'}}(\tilde{M})|}{|\tilde{M}|} \geq 1$ for all g' -independent sets $\tilde{M} \subseteq N'_{\bar{s}'}$. Based on this, we can again construct a sequence of players (i_1, i_2, i_3, \dots) such that, according to the previous considerations, the sequential adding of new players can never break up. Thus, we again get a contradiction to the finiteness of the player set N . For this purpose, we define the sets $\tilde{M}^m := \{i_l \mid 1 \leq l \leq m, l \text{ odd}\}$ and $\tilde{L}^m := \{i_l \mid 1 \leq l \leq m, l \text{ even}\}$ for $m \in \mathbb{N}$.

Initially, select a player $i_1 \in M_1 \cap N'_{\bar{s}'} \setminus K$. Such a player must exist as $k \in L_1 \cap N'_{\bar{s}'}$ is part of a cycle in $g|_{N \setminus K \cup \{k\}}$ and, according to Lemma 3.2, we have $M_1 \cap L'_s = \emptyset$ for all $s < \bar{s}'$. Now consider some odd number $m \in \mathbb{N}$. Assume that $\tilde{M}^m \subseteq M_1 \cap N'_{\bar{s}'} \setminus K$, $\tilde{L}^m \subseteq L_1 \cap N'_{\bar{s}'} \setminus K$ and that $|\tilde{M}^m| = \frac{m+1}{2} > \frac{m-1}{2} = |\tilde{L}^m|$. We then have:

- $\tilde{M}^m \dot{\cup} (M_1 \cap C'_k) \subseteq N'_{\bar{s}'}$ is g' -independent and

$$\frac{|\tilde{L}^m \dot{\cup} (L_1 \cap C'_k)|}{|\tilde{M}^m \dot{\cup} (M_1 \cap C'_k)|} < 1$$

since it is $|L_1 \cap C'_k| = |M_1 \cap C'_k|$ as we know from equation (3.12). As we have $k \in L^{g'_{\bar{s}'}}(M_1 \cap C'_k) \subseteq L_1 \cap C'_k$, this implies that there must exist a player $i_{m+1} \in L^{g'_{\bar{s}'}}(\tilde{M}^m) \setminus (\tilde{L}^m \dot{\cup} K)$. It is $i_{m+1} \in L_1 \cap N'_{\bar{s}'} \setminus (\tilde{L}^m \dot{\cup} K)$ since $\tilde{M}^m \subseteq M_1$.

- We then have $|\tilde{L}^{m+1}| = |\tilde{M}^{m+1}| = \frac{m+1}{2} \geq 1$ and $\tilde{L}^{m+1} \subseteq L_1 \cap N'_{\bar{s}'} \setminus K$, $\tilde{M}^{m+1} \subseteq M_1 \cap N'_{\bar{s}'} \setminus K$. Hence, by implication (3.14) there exists $i_{m+2} \in L^{g'_{\bar{s}'}}(\tilde{L}^{m+1}) \setminus (\tilde{M}^{m+1} \dot{\cup} K)$. It is $i_{m+2} \in M_1 \cap N'_{\bar{s}'} \setminus (\tilde{M}^{m+1} \dot{\cup} K)$ since $\tilde{L}^{m+1} \subseteq L_1$.

Thus, we have $\tilde{M}^{m+2} \subseteq M_1 \cap N'_{\bar{s}'} \setminus K$, $\tilde{L}^{m+2} \subseteq L_1 \cap N'_{\bar{s}'} \setminus K$ and $|\tilde{M}^{m+2}| = \frac{(m+2)+1}{2} > \frac{(m+2)-1}{2} = |\tilde{L}^{m+2}|$. Again, by induction this leads to a contradiction to the finiteness of the player set N . This proves that player k 's payoff must indeed be strictly greater than $\frac{1}{2}$. The corresponding stability condition then yields

$$u_k^*(g) \geq u_k^*(g - ki') \Leftrightarrow v_k^*(g) - c \geq v_k^*(g') \Rightarrow 1 - x_1 - c > \frac{1}{2} \Leftrightarrow x_1 + c < \frac{1}{2},$$

which is a contradiction to Theorem 3.3. Hence, such a network g cannot be pairwise stable.

It remains to prove implications (3.13) and (3.14). We start with the first one. Given the two sets $\hat{L} \subseteq L_1 \cap C'_k \setminus \{k\}$ and $\hat{M} \subseteq M_1 \cap C'_k$ with $N_k(g) \cap K \subseteq \hat{M}$ and $|\hat{L}| = |\hat{M}| \geq 1$ assume for contradiction that $L^{g'}(\hat{L}) \subseteq \hat{M}$. Note that we have $N_j(g'_{\bar{s}'}) = N_j(g)$ for all $j \in \hat{L}$ since it is $\hat{L} \subseteq L_1 \cap N'_{\bar{s}'} \setminus \{k\}$ and, according to Lemma 3.2, $M_1 \cap L'_s = \emptyset$ for all $s < \bar{s}'$. Together with the assumption this implies that $L^g(M_1 \cap K \setminus \hat{M}) \subseteq L_1 \cap K \setminus \hat{L}$. Moreover, since $N_k(g) \cap K \subseteq \hat{M}$, it even is $L^g(M_1 \cap K \setminus \hat{M}) \subseteq L_1 \cap K \setminus (\hat{L} \cup \{k\})$.

Additionally, we need the following inequalities:

$$\frac{|L_1 \cap K| - 1}{|M_1 \cap K|} \leq r_1 \leq \frac{|L_1 \cap K|}{|M_1 \cap K|} \leq 1 \quad (3.15)$$

To see that these are correct, note first that it is $L^g(M_1 \cap K) \subseteq L_1 \cap K$ and similarly $L^g(M_1 \setminus K) \subseteq L_1 \setminus K \cup \{k\}$. So we must have $r_1 \leq \frac{|L_1 \cap K|}{|M_1 \cap K|}$ and $r_1 \leq \frac{|L_1 \setminus K| + 1}{|M_1 \setminus K|}$ as r_1 is the minimal shortage ratio. Moreover, it is $r_1 = \frac{|L_1|}{|M_1|} < 1$, $M_1 = (M_1 \cap K) \cup (M_1 \setminus K)$ and $L_1 = (L_1 \cap K) \cup (L_1 \setminus K)$. Together this implies that $\frac{|L_1 \cap K| - 1}{|M_1 \cap K|} = \frac{|L_1| - (|L_1 \setminus K| + 1)}{|M_1| - |M_1 \setminus K|} \leq r_1$. In particular, this means that $|L_1 \cap K| - 1 < |M_1 \cap K|$ which in turn implies $\frac{|L_1 \cap K|}{|M_1 \cap K|} \leq 1$.

According to the third inequality in (3.15) we must have $M_1 \cap K \setminus \hat{M} \neq \emptyset$ since otherwise it would be $|L_1 \cap K| \leq |M_1 \cap K| = |\hat{M}| = |\hat{L}| < |\hat{L} \cup \{k\}| \leq |L_1 \cap K|$. Taken together, this leads to the following contradiction:

$$\begin{aligned} r_1 &\leq \frac{|L^g(M_1 \cap K \setminus \hat{M})|}{|M_1 \cap K \setminus \hat{M}|} \leq \frac{|L_1 \cap K \setminus (\hat{L} \cup \{k\})|}{|M_1 \cap K \setminus \hat{M}|} \\ &= \frac{|L_1 \cap K| - |\hat{L}| - 1}{|M_1 \cap K| - |\hat{M}|} \\ &= \frac{|L_1 \cap K| - 1 - |\hat{L}|}{|M_1 \cap K| - |\hat{L}|} < \frac{|L_1 \cap K| - 1}{|M_1 \cap K|} \leq r_1, \end{aligned}$$

where the last two inequalities are due to (3.15) and the fact that $r_1 < 1$.

Similarly, to prove implication (3.14), we consider the two sets $\tilde{L} \subseteq L_1 \cap N'_{\tilde{s}} \setminus K$ and $\tilde{M} \subseteq M_1 \cap N'_{\tilde{s}} \setminus K$ with $|\tilde{L}| = |\tilde{M}| \geq 1$ and assume for contradiction that we have $L^{g'_{\tilde{s}}}(\tilde{L}) \subseteq \tilde{M}$. Again according to Lemma 3.2, it must be $N_j(g'_{\tilde{s}}) = N_j(g)$ for all $j \in \tilde{L}$. Hence, we have that $L^g(M_1 \setminus \tilde{M}) \subseteq L_1 \setminus \tilde{L}$. Also, it is clear that $M_1 \setminus \tilde{M} \neq \emptyset$ since otherwise we would have $|L_1| < |M_1| = |\tilde{M}| = |\tilde{L}| \leq |L_1|$. Summing up, this implies

$$r_1 \leq \frac{|L^g(M_1 \setminus \tilde{M})|}{|M_1 \setminus \tilde{M}|} \leq \frac{|L_1 \setminus \tilde{L}|}{|M_1 \setminus \tilde{M}|} = \frac{|L_1| - |\tilde{L}|}{|M_1| - |\tilde{M}|} = \frac{|L_1| - |\tilde{L}|}{|M_1| - |\tilde{L}|} < \frac{|L_1|}{|M_1|} = r_1,$$

which is obviously again a contradiction. So we have that $L^{g'_{\tilde{s}}}(\tilde{L}) \setminus (\tilde{M} \cup K) \neq \emptyset$ since it is $L^{g'_{\tilde{s}}}(\tilde{L}) \subseteq K^c$. This concludes the proof. \square

3.A.7 Proof of Theorem 3.4 (for n odd)

Consider a network g and let $n = |N|$ be odd (as only this case is remaining). Considering again $\bar{N}(g)$ instead of N as player set, let $g_{\bar{N}(g)}^{SPL}$ denote a representative of the networks consisting of $\frac{|\bar{N}(g)|-3}{2}$ separated pairs and one line of length three. Similarly, let $g_{\bar{N}(g)}^{SPC}$ be a network consisting of $\frac{|\bar{N}(g)|-3}{2}$ separated pairs and one three-player circle. Since we did not utilize that n was even to derive inequality (3.5), we again have

$$U^*(g) \leq \frac{1}{2}|\bar{N}(g)| - 2d^\#(g)c.$$

Again, since $\eta_i(g) \geq 1$ for all $i \in \bar{N}(g)$, we must have $d^\#(g) \geq \frac{1}{2}|\bar{N}(g)|$. We now distinguish two cases.

Case 1: $|\bar{N}(g)|$ even

If we have $d^\#(g) = \frac{1}{2}|\bar{N}(g)|$ here, then this implies again that $g = g_{\bar{N}(g)}^{SP}$. So conversely, for a network $g \neq g_{\bar{N}(g)}^{SP}$ this means that we have $d^\#(g) > \frac{1}{2}|\bar{N}(g)|$ and therefore, according to (3.5),

$$U^*(g) < \frac{1}{2}|\bar{N}(g)| - |\bar{N}(g)|c = \sum_{i \in \bar{N}(g)} \left(\frac{1}{2} - c \right) = U^*(g_{\bar{N}(g)}^{SP}).$$

Case 2: $|\bar{N}(g)|$ odd

Note first that it must be $d^\#(g) \geq \frac{1}{2}(|\bar{N}(g)|+1)$ as the interval $[\frac{1}{2}|\bar{N}(g)|, \frac{1}{2}(|\bar{N}(g)|+1))$ does not contain an integer here. Consider the following subcases.

Subcase 2(a): $d^\#(g) = \frac{1}{2}(|\bar{N}(g)| + 1)$

This subcase implies that $g = g_{\bar{N}(g)}^{SPL}$. This is because otherwise we would either have $\eta_j(g) \geq 3$ for at least one player $j \in \bar{N}(g)$ or $\eta_k(g), \eta_l(g) \geq 2$ for $k, l \in \bar{N}(g)$, $k \neq l$, which both would give $d^\#(g) \geq \frac{1}{2}(4 + (|\bar{N}(g)| - 2)) = \frac{1}{2}(|\bar{N}(g)| + 2) > \frac{1}{2}(|\bar{N}(g)| + 1)$.

Subcase 2(b): $d^\#(g) = \frac{1}{2}(|\bar{N}(g)| + 3)$

Here, g must either be

- (i) a network with three players having two links each and $|\bar{N}(g)| - 3$ players with one link,
- (ii) a network consisting of one player with three links, one player with two links and $|\bar{N}(g)| - 2$ players with one link or
- (iii) a network with one player having four links and $|\bar{N}(g)| - 1$ players with one link.

Note that the network $g_{\bar{N}(g)}^{SPC}$ is included here in Class (i). On closer examination, one finds that, for any other $g \neq g_{\bar{N}(g)}^{SPC}$ belonging to Class (i), (ii) or (iii), the algorithm $\mathcal{A}(g)$ (with $N_1 = \bar{N}(g)$) yields $|M_1| > |L_1|$. This implies a strict inequality in (3.5). Hence, for $g \neq g_{\bar{N}(g)}^{SPC}$ with $d^\#(g) = \frac{1}{2}(|\bar{N}(g)| + 3)$ it is

$$U^*(g) < \frac{1}{2}|\bar{N}(g)| - 2d^\#(g)c = \frac{1}{2}|\bar{N}(g)| - (|\bar{N}(g)| + 3)c = |\bar{N}(g)|\left(\frac{1}{2} - c\right) - 3c = U^*(g_{\bar{N}(g)}^{SPC}).$$

Subcase 2(c): $d^\#(g) > \frac{1}{2}(|\bar{N}(g)| + 3)$

In this subcase, again according to (3.5), we have

$$U^*(g) \leq \frac{1}{2}|\bar{N}(g)| - 2d^\#(g)c < \frac{1}{2}|\bar{N}(g)| - (|\bar{N}(g)| + 3)c = |\bar{N}(g)|\left(\frac{1}{2} - c\right) - 3c = U^*(g_{\bar{N}(g)}^{SPC}).$$

Summarizing this, we have shown that a network $g \notin \{g_{\bar{N}(g)}^{SP}, g_{\bar{N}(g)}^{SPL}, g_{\bar{N}(g)}^{SPC}\}$ cannot be efficient. To conclude the proof, we have to examine which of the remaining candidates is efficient depending on the level of linking costs. Note that for the set $\bar{N}(g)$ it must hold that $0 \leq |\bar{N}(g)| \leq n$ and $|\bar{N}(g)| \neq 1$. Moreover, recall that $g_{\bar{N}(g)}^{SP}$ is only well-defined for $|\bar{N}(g)|$ even while $g_{\bar{N}(g)}^{SPL}$ and $g_{\bar{N}(g)}^{SPC}$ only are so for $|\bar{N}(g)|$ odd. Thus, we have

$$\begin{aligned} \max_{\bar{N}(g) \text{ feasible}} U^*(g_{\bar{N}(g)}^{SP}) &= \max_{\bar{N}(g) \text{ feasible}} |\bar{N}(g)|\left(\frac{1}{2} - c\right) \\ &= \begin{cases} 0, & \text{for } c \geq \frac{1}{2} \\ (n-1)\left(\frac{1}{2} - c\right), & \text{for } c \in \left(0, \frac{1}{2}\right) \end{cases}, \end{aligned}$$

$$\begin{aligned}
 \max_{\bar{N}(g) \text{ feasible}} U^*(g_{\bar{N}(g)}^{SPL}) &= \max_{\bar{N}(g) \text{ feasible}} |\bar{N}(g)| \left(\frac{1}{2} - c \right) - \left(c + \frac{1}{6} \right) \\
 &= \begin{cases} \frac{4}{3} - 4c, & \text{for } c \geq \frac{1}{2} \\ n \left(\frac{1}{2} - c \right) - \left(c + \frac{1}{6} \right), & \text{for } c \in \left(0, \frac{1}{2} \right) \end{cases}, \\
 \max_{\bar{N}(g) \text{ feasible}} U^*(g_{\bar{N}(g)}^{SPC}) &= \max_{\bar{N}(g) \text{ feasible}} |\bar{N}(g)| \left(\frac{1}{2} - c \right) - 3c \\
 &= \begin{cases} \frac{3}{2} - 6c, & \text{for } c \geq \frac{1}{2} \\ n \left(\frac{1}{2} - c \right) - 3c, & \text{for } c \in \left(0, \frac{1}{2} \right) \end{cases}.
 \end{aligned}$$

Hence, for $c \geq \frac{1}{2}$ it is

$$\max_{\bar{N}(g) \text{ feasible}} U^*(g_{\bar{N}(g)}^{SPL}), \max_{\bar{N}(g) \text{ feasible}} U^*(g_{\bar{N}(g)}^{SPC}) < 0 = \max_{\bar{N}(g) \text{ feasible}} U^*(g_{\bar{N}(g)}^{SP}).$$

So in this case, a network $g \in \{g_{\bar{N}(g)}^{SPL}, g_{\bar{N}(g)}^{SPC}\}$ cannot be efficient. For $c > \frac{1}{2}$, g with $\bar{N}(g) = \emptyset$ is the unique maximizer of $U^*(g_{\bar{N}(g)}^{SP})$, meaning that the empty network is uniquely efficient. On the contrary, for $c = \frac{1}{2}$, any network $g = g_{\bar{N}(g)}^{SP}$ maximizes $U^*(g_{\bar{N}(g)}^{SP})$, meaning that a network is efficient if and only if it is a union of separated pairs and isolated players.

Concerning linking costs $c \in (0, \frac{1}{2})$ we calculate

$$\max \left\{ - \left(\frac{1}{2} - c \right), - \left(c + \frac{1}{6} \right), -3c \right\} = \begin{cases} - \left(\frac{1}{2} - c \right), & \text{for } c \geq \frac{1}{6} \\ - \left(c + \frac{1}{6} \right), & \text{for } c \in \left[\frac{1}{12}, \frac{1}{6} \right] \\ -3c, & \text{for } c \leq \frac{1}{12} \end{cases}.$$

This means that a network is efficient for linking costs

- $c \in [\frac{1}{6}, \frac{1}{2}]$ if and only if it is a union of $\frac{n-1}{2}$ separated pairs and one isolated player,
- $c \in [\frac{1}{12}, \frac{1}{6}]$ if and only if it is a union of $\frac{n-3}{2}$ separated pairs and a line of length three,
- $c \in (0, \frac{1}{12}]$ if and only if it is a union of $\frac{n-3}{2}$ separated pairs and a three-player circle.

This concludes the proof. □

3.A.8 Proof of Example 3.2

The second part of the example is an immediate consequence of Theorem 3.2. So consider the case $\delta \in (0, 1)$. To start with, we solve the equation

$$v_1 = \left(1 - \frac{n-1}{n(n-1)}\right)\delta v_1 + \frac{n-1}{n(n-1)}(1 - \delta v_1)$$

which gives $v_1 = \frac{1}{(1-\delta)n+2\delta}$. Obviously, it is $v_1 \in (0, \frac{1}{2})$ which implies $1 - \delta v_1 > \delta v_1$. Moreover, we have $d^\#(g^N) = \frac{n(n-1)}{2}$ and all players are in symmetric positions. This shortcut avoiding extensive calculations establishes that the n -tuple (v_1, v_1, \dots, v_1) solves the equation system (3.1) for the network g^N . Therefore, we have

$$v_i^{*\delta}(g^N) = \frac{1}{(1-\delta)n+2\delta} \quad (3.16)$$

for all $i \in N$. Next, consider the network $g^N - ij$ for some $i, j \in N$. For this purpose, let $\tilde{v} = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_n)$ be given by

$$\begin{aligned} \tilde{v}_i = \tilde{v}_j &= \frac{(1-\delta)n^2 + (2\delta-1)n - (\delta+2)}{(\delta^2 - 2\delta + 1)n^3 + (-3\delta^2 + 3\delta)n^2 + (2\delta^2 + 3\delta - 3)n - (2\delta^2 + 2\delta + 2)}, \\ \tilde{v}_k &= \frac{(1-\delta)n^2 + \delta n - (2\delta+1)}{(\delta^2 - 2\delta + 1)n^3 + (-3\delta^2 + 3\delta)n^2 + (2\delta^2 + 3\delta - 3)n - (2\delta^2 + 2\delta + 2)}, \end{aligned} \quad (3.17)$$

where $k \in N \setminus \{i, j\}$. By showing that the denominator of the terms in (3.17) is in both cases greater than the numerator and that both numerators are greater than zero, we establish first that $\tilde{v} \in (0, 1)^n$. For $\delta \in (0, 1)$ and $n \geq 4$ we have

$$\begin{aligned} & \left((\delta^2 - 2\delta + 1)n^3 + (-3\delta^2 + 3\delta)n^2 + (2\delta^2 + 3\delta - 3)n - (2\delta^2 + 2\delta + 2) \right) \\ & - \left((1-\delta)n^2 + (2\delta-1)n - (\delta+2) \right) \\ &= (1-\delta)^2 n^3 + (1-\delta)(3\delta-1)n^2 + (2\delta^2 + \delta - 2)n - \delta(2\delta+1) \\ &= (1-\delta)n \left[(1-\delta)n^2 + (3\delta-1)n - (2\delta+3) \right] + n - \delta(2\delta+1) \\ &> (1-\delta)n[2n - (2\delta+3)] + n - \delta(2\delta+1) \\ &> (1-\delta)n[2n-5] + (n-3) \\ &> 0 \end{aligned}$$

and

$$\begin{aligned}
& \left((\delta^2 - 2\delta + 1)n^3 + (-3\delta^2 + 3\delta)n^2 + (2\delta^2 + 3\delta - 3)n - (2\delta^2 + 2\delta + 2) \right) \\
& \quad - \left((1 - \delta)n^2 + \delta n - (2\delta + 1) \right) \\
&= (1 - \delta)^2 n^3 + (1 - \delta)(3\delta - 1)n^2 + (2\delta^2 + 2\delta - 3)n - (2\delta^2 + 1) \\
&= (1 - \delta)n \left[(1 - \delta)n^2 + (3\delta - 1)n - (2\delta + 4) \right] + n - (2\delta^2 + 1) \\
&> (1 - \delta)n[2n - (2\delta + 4)] + n - (2\delta^2 + 1) \\
&> (1 - \delta)n[2n - 6] + (n - 3) \\
&> 0,
\end{aligned}$$

and, moreover,

$$\begin{aligned}
(1 - \delta)n^2 + (2\delta - 1)n - (\delta + 2) &> n - (\delta + 2) > n - 3 > 0, \\
(1 - \delta)n^2 + \delta n - (2\delta + 1) &> n - (2\delta + 1) > n - 3 > 0.
\end{aligned}$$

Next, we show that it is $1 - \delta\tilde{v}_i - \delta\tilde{v}_k > 0$ and $1 - 2\delta\tilde{v}_k > 0$, which implies that $\max\{1 - \delta\tilde{v}_i, \delta\tilde{v}_k\} = 1 - \delta\tilde{v}_i$, $\max\{1 - \delta\tilde{v}_k, \delta\tilde{v}_i\} = 1 - \delta\tilde{v}_k$ and $\max\{1 - \delta\tilde{v}_k, \delta\tilde{v}_k\} = 1 - \delta\tilde{v}_k$. We calculate

$$\begin{aligned}
1 - \delta\tilde{v}_i - \delta\tilde{v}_k &= \frac{(\delta^2 - 2\delta + 1)n^3 + (-\delta^2 + \delta)n^2 + (-\delta^2 + 4\delta - 3)n + (\delta^2 + \delta - 2)}{(\delta^2 - 2\delta + 1)n^3 + (-3\delta^2 + 3\delta)n^2 + (2\delta^2 + 3\delta - 3)n - (2\delta^2 + 2\delta + 2)} \\
&= \frac{(1 - \delta) \left[(1 - \delta)n^3 + \delta n^2 + (\delta - 3)n - (\delta + 2) \right]}{(\delta^2 - 2\delta + 1)n^3 + (-3\delta^2 + 3\delta)n^2 + (2\delta^2 + 3\delta - 3)n - (2\delta^2 + 2\delta + 2)} \\
&> \frac{(1 - \delta)[n^2 + (\delta - 3)n - (\delta + 2)]}{(\delta^2 - 2\delta + 1)n^3 + (-3\delta^2 + 3\delta)n^2 + (2\delta^2 + 3\delta - 3)n - (2\delta^2 + 2\delta + 2)} \\
&> \frac{(1 - \delta) \overbrace{[n^2 - 3n - 3]}^{>0}}{(\delta^2 - 2\delta + 1)n^3 + (-3\delta^2 + 3\delta)n^2 + (2\delta^2 + 3\delta - 3)n - (2\delta^2 + 2\delta + 2)} \\
&> 0
\end{aligned}$$

and

$$\begin{aligned}
1 - 2\delta\tilde{v}_k &= \frac{(\delta^2 - 2\delta + 1)n^3 + (-\delta^2 + \delta)n^2 + (3\delta - 3)n + (2\delta^2 - 2)}{(\delta^2 - 2\delta + 1)n^3 + (-3\delta^2 + 3\delta)n^2 + (2\delta^2 + 3\delta - 3)n - (2\delta^2 + 2\delta + 2)} \\
&= \frac{(1 - \delta) \left[(1 - \delta)n^3 + \delta n^2 - 3n - 2(\delta + 1) \right]}{(\delta^2 - 2\delta + 1)n^3 + (-3\delta^2 + 3\delta)n^2 + (2\delta^2 + 3\delta - 3)n - (2\delta^2 + 2\delta + 2)} \\
&> \frac{(1 - \delta)[n^2 - 3n - 2(\delta + 1)]}{(\delta^2 - 2\delta + 1)n^3 + (-3\delta^2 + 3\delta)n^2 + (2\delta^2 + 3\delta - 3)n - (2\delta^2 + 2\delta + 2)}
\end{aligned}$$

$$\begin{aligned}
 &> \frac{(1-\delta)\overbrace{[n^2 - 3n - 4]}^{\geq 0}}{(\delta^2 - 2\delta + 1)n^3 + (-3\delta^2 + 3\delta)n^2 + (2\delta^2 + 3\delta - 3)n - (2\delta^2 + 2\delta + 2)} \\
 &\geq 0.
 \end{aligned}$$

Furthermore, note that $d^\#(g^N - ij) = \frac{n(n-1)-2}{2}$ and, hence, for the network $g^N - ij$ equation system (3.1) is equivalent to

$$\begin{aligned}
 v_l &= \left(1 - \frac{n-2}{n(n-1)-2}\right)\delta v_l + \frac{n-2}{n(n-1)-2} \max\{1 - \delta v_k, \delta v_l\}, \\
 v_k &= \left(1 - \frac{n-1}{n(n-1)-2}\right)\delta v_k + \frac{2}{n(n-1)-2} \max\{1 - \delta v_l, \delta v_k\} \\
 &\quad + \frac{n-3}{n(n-1)-2} \max\{1 - \delta v_k, \delta v_k\},
 \end{aligned} \tag{3.18}$$

for all $l \in \{i, j\}$, $k \in N \setminus \{i, j\}$. Using our preparatory work, one can show by straightforward calculations that \tilde{v} as given in (3.17) is a solution to the system (3.18). Hence, as we know from Manea (2011, Theorem 1) that this solution is unique, the equilibrium payoff vector is $v^{*\delta}(g^N - ij) = \tilde{v}$.

After deriving the payoffs in both networks g^N and $g^N - ij$, it remains to show that for all $\delta \in (0, 1)$ and $n \geq 4$ there exists $\bar{c} > 0$ such that for all $c \in (0, \bar{c}]$ and $i, j \in N$ it is

$$v_i^{*\delta}(g^N) - v_i^{*\delta}(g^N - ij) \geq c. \tag{3.19}$$

For this purpose let

$$\bar{c} := \frac{2(1-\delta)(n-1)}{((1-\delta)n+2\delta)((\delta^2-2\delta+1)n^3 + (-3\delta^2+3\delta)n^2 + (2\delta^2+3\delta-3)n - (2\delta^2+2\delta+2))}.$$

Note that the denominator is positive as it is the product of the denominators of the terms in (3.16) and (3.17). Hence, we have $\bar{c} > 0$ and calculate

$$\begin{aligned}
 &v_i^{*\delta}(g^N) - v_i^{*\delta}(g^N - ij) \\
 &= \frac{1}{(1-\delta)n+2\delta} - \frac{(1-\delta)n^2 + (2\delta-1)n - (\delta+2)}{(\delta^2-2\delta+1)n^3 + (-3\delta^2+3\delta)n^2 + (2\delta^2+3\delta-3)n - (2\delta^2+2\delta+2)} \\
 &= \frac{(\delta^2-2\delta+1)n^2 + (-3\delta^2+4\delta-1)n + (2\delta-2)}{((1-\delta)n+2\delta)((\delta^2-2\delta+1)n^3 + (-3\delta^2+3\delta)n^2 + (2\delta^2+3\delta-3)n - (2\delta^2+2\delta+2))} \\
 &= \frac{(1-\delta)[(1-\delta)n^2 + (3\delta-1)n - 2]}{((1-\delta)n+2\delta)((\delta^2-2\delta+1)n^3 + (-3\delta^2+3\delta)n^2 + (2\delta^2+3\delta-3)n - (2\delta^2+2\delta+2))}
 \end{aligned}$$

$$\begin{aligned}
&> \frac{(1 - \delta)[2n - 2]}{((1 - \delta)n + 2\delta)((\delta^2 - 2\delta + 1)n^3 + (-3\delta^2 + 3\delta)n^2 + (2\delta^2 + 3\delta - 3)n - (2\delta^2 + 2\delta + 2))} \\
&= \bar{c}.
\end{aligned}$$

This concludes the proof of the example and Appendix 3.A. □

Appendix 3.B Relation to the Works of Hellmann (2013) and Hellmann and Landwehr (2014)

As mentioned in Sections 3.1 and 3.3, there is a strand of literature which provides some general results about existence, uniqueness and the structure of (pairwise) stable networks. Prima facie, at least some findings of Hellmann (2013) and Hellmann and Landwehr (2014) are most likely in line for an application to our framework. According to Hellmann (2013) the existence of a pairwise stable network is, for instance, guaranteed if the profile of utility or profit functions is ordinal convex in own links and satisfies ordinal strategic complements. Other findings of Hellmann (2013) and Hellmann and Landwehr (2014) concerning these issues are further and among other properties based on concavity, anonymous convexity, the strategic substitutes property or strong preference for centrality. In this appendix, we first provide explanations and definitions of these concepts. Second, we provide appropriate counterexamples which demonstrate that, among these properties, the profit function considered in our model with $\delta \rightarrow 1$ does not satisfy some crucial ones at least for a broad range of cost levels. In this light, this confirms that our analysis is not a special case of questions which have already been answered before but of some independent interest.

The findings of Hellmann (2013) and Hellmann and Landwehr (2014) are mainly based on marginal effects of link creation. To be able to summarize certain properties in this context and to make use of the subsequent counterexamples we require some additional notation. For a given network g and $ij \notin g$ let $\Delta u_i(g+ij, ij) := u_i(g+ij) - u_i(g)$ denote the *marginal utility* of the link ij for player $i \in N$. Further, let the set of all own links of a player $i \in N$ in a network g be denoted by $L_i(g) := \{ij \in g \mid j \in N\}$ whereas $L_{-i}(g) := g - L_i(g)$ denotes all other links.

At least one of the following properties is part of the conditions of each relevant theorem, proposition or corollary of Hellmann (2013) and Hellmann and Landwehr (2014).

Definition 3.5 (Marginal Effects). *A profile of utility functions $(u_i)_{i \in N}$*

- *is concave in own links if for all $g \subseteq g^N$, $i \in N$, $l_i \subseteq L_i(g^N - g)$, $ij \notin g + l_i$ we have*

$$\Delta u_i(g + ij, ij) \geq \Delta u_i(g + l_i + ij, ij),$$

- *is ordinal convex in own links if for all $g \subseteq g^N$, $i \in N$, $l_i \subseteq L_i(g^N - g)$, $ij \notin g + l_i$ we have*

$$(i) \Delta u_i(g + ij, ij) \geq 0 \Rightarrow \Delta u_i(g + l_i + ij, ij) \geq 0 \text{ and}$$

$$(ii) \Delta u_i(g + ij, ij) > 0 \Rightarrow \Delta u_i(g + l_i + ij, ij) > 0,$$

- *satisfies anonymous convexity if for all $g \subseteq g^N$, $i, j, k \in N$ with $\eta_i(g) \leq \eta_j(g)$, $ik \in g$ and $jk \notin g$ we have*

$$\Delta u_i(g, ik) \geq 0 \Rightarrow \Delta u_j(g + jk, jk) \geq 0,$$

- *satisfies strong preference for centrality if for all $g \subseteq g^N$, $i, j, k \in N$ with $\eta_j(g) \leq \eta_k(g)$, $ij \in g$ and $ik \notin g$ we have*

$$\Delta u_i(g, ij) \geq 0 \Rightarrow \Delta u_i(g + ik, ik) > 0,$$

- *satisfies ordinal strategic complements (substitutes) if for all $g \subseteq g^N$, $i \in N$, $l_{-i} \subseteq L_{-i}(g^N - g)$, $ij \notin g$ we have*

$$(i) \Delta u_i(g + ij, ij) \geq 0 \Rightarrow (\Leftrightarrow) \Delta u_i(g + l_{-i} + ij, ij) \geq 0 \text{ and}$$

$$(ii) \Delta u_i(g + ij, ij) > 0 \Rightarrow (\Leftrightarrow) \Delta u_i(g + l_{-i} + ij, ij) > 0, \text{ and}$$

- *satisfies positive (negative) externalities if for all $g \subseteq g^N$, $jk \notin g$, $i \in N \setminus \{j, k\}$ we have*

$$u_i(g + jk) \geq (\leq) u_i(g).$$

We now provide counterexamples which establish that, at least for some cost levels, the profile of profit functions $(u_i^*)_{i \in N}$ is neither concave or ordinal convex nor does it satisfy anonymous convexity, strong preference for centrality, ordinal strategic complements/substitutes or positive/negative externalities.

Counterexample 3.1 (Concavity). *Consider the player set N with $n = 7$ and the network $g := \{14, 45, 56, 67\}$. Further let $l_1 := \{13\}$.*

Note first that it is $l_1 \subseteq L_1(g^N - g)$ as required. Applying the algorithm $\mathcal{A}(\cdot)$ to the different networks gives

$$\begin{aligned}\Delta u_1^*(g + 12, 12) &= u_1^*(g + 12) - u_1^*(g) = \frac{1}{2} - 2c - \left(\frac{2}{5} - c\right) = \frac{1}{10} - c, \quad \text{and} \\ \Delta u_1^*(g + l_1 + 12, 12) &= u_1^*(g + l_1 + 12) - u_1^*(g + l_1) = \frac{2}{3} - 3c - \left(\frac{1}{2} - 2c\right) = \frac{1}{6} - c.\end{aligned}$$

This yields $\Delta u_1^*(g + l_1 + 12, 12) > \Delta u_1^*(g + 12, 12)$ for all $c > 0$ which means that our profit function is not concave.

Counterexample 3.2 (Ordinal Convexity). *Consider the player set N with $n = 4$ and the network $g := \{24\}$. Further let $l_1 := \{13, 14\}$.*

Note again first that we have $l_1 \subseteq L_1(g^N - g)$ in this counterexample as well. For $c \in \left(0, \frac{1}{3}\right]$ we calculate by using the algorithm $\mathcal{A}(\cdot)$ that

$$\begin{aligned}\Delta u_1^*(g + 12, 12) &= u_1^*(g + 12) - u_1^*(g) = \frac{1}{3} - c - 0 \geq 0, \quad \text{but} \\ \Delta u_1^*(g + l_1 + 12, 12) &= u_1^*(g + l_1 + 12) - u_1^*(g + l_1) = \frac{1}{2} - 3c - \left(\frac{1}{2} - 2c\right) = -c < 0.\end{aligned}$$

Thus, our profit function is in general not convex either, even not in ordinal notion.

Counterexample 3.3 (Anonymous Convexity and Strong Preference for Centrality). *Consider the player set N with $n = 4$ and the network $g := \{13, 24\}$.*

Note that it is $\eta_i(g) = 1$ for all $i \in N$. For $c \in \left(0, \frac{1}{2}\right]$ we have according to the outcome of $\mathcal{A}(\cdot)$ that

$$\begin{aligned}\Delta u_1^*(g, 13) &= u_1^*(g) - u_1^*(g - 13) = \frac{1}{2} - c - 0 \geq 0, \quad \text{but both} \\ \Delta u_2^*(g + 23, 23) &= u_2^*(g + 23) - u_2^*(g) = \frac{1}{2} - 2c - \left(\frac{1}{2} - c\right) = -c < 0 \quad \text{and} \\ \Delta u_1^*(g + 12, 12) &= u_1^*(g + 12) - u_1^*(g) = \frac{1}{2} - 2c - \left(\frac{1}{2} - c\right) = -c < 0.\end{aligned}$$

Thus, our profit function does neither in general satisfy anonymous convexity nor strong preference for centrality.

Counterexample 3.4 (Ordinal Strategic Substitutes). *Consider the player set N with $n = 5$ and the network $g := \{14, 23\}$. Further let $l_{-1} := \{45\}$.*

Note that, as required, we have $l_{-1} \subseteq L_{-1}(g^N - g)$ in this case. For $c \in \left(0, \frac{1}{15}\right]$ we

get again from $\mathcal{A}(\cdot)$ that

$$\begin{aligned}\Delta u_1^*(g + l_{-1} + 12, 12) &= u_1^*(g + l_{-1} + 12) - u_1^*(g + l_{-1}) = \frac{2}{5} - 2c - \left(\frac{1}{3} - c\right) \\ &= \frac{1}{15} - c \geq 0, \quad \text{but} \\ \Delta u_1^*(g + 12, 12) &= u_1^*(g + 12) - u_1^*(g) = \frac{1}{2} - 2c - \left(\frac{1}{2} - c\right) = -c < 0.\end{aligned}$$

Hence, our profit function does not in general satisfy ordinal strategic substitutes.

Counterexample 3.5 (Ordinal Strategic Complements). *Consider the player set N with $n = 4$ and the network $g := \{14\}$. Further let $l_{-1} := \{23\}$.*

Note first, that it is again $l_{-1} \subseteq L_{-1}(g^N - g)$. Calculating payoffs in the usual way gives for $c \in \left(0, \frac{1}{6}\right]$ that

$$\begin{aligned}\Delta u_1^*(g + 12, 12) &= u_1^*(g + 12) - u_1^*(g) = \frac{2}{3} - 2c - \left(\frac{1}{2} - c\right) = \frac{1}{6} - c \geq 0, \quad \text{but} \\ \Delta u_1^*(g + l_{-1} + 12, 12) &= u_1^*(g + l_{-1} + 12) - u_1^*(g + l_{-1}) = \frac{1}{2} - 2c - \left(\frac{1}{2} - c\right) < 0.\end{aligned}$$

Hence, our profit function does not in general satisfy ordinal strategic complements either.

Counterexample 3.6 (Positive/Negative Externalities). *Consider the player set N with $n = 4$ and the network $g := \{12, 23\}$.*

For this counterexample we calculate

$$\begin{aligned}u_1^*(g + 34) &= \frac{1}{2} - c > \frac{1}{3} - c = u_1^*(g) \quad \text{and} \\ u_2^*(g + 34) &= \frac{1}{2} - 2c < \frac{2}{3} - 2c = u_2^*(g).\end{aligned}$$

Consequently, our profit function does neither satisfy positive nor negative externalities for all $c > 0$.

Thus, the question of existence and structure of stable networks cannot be answered by applying the results of Hellmann (2013) and Hellmann and Landwehr (2014). In this sense, our problem seems to be independent and indeed requires a detailed analysis as conducted in Section 3.3.

Appendix 3.C Relation to the Work of O’Donnell (2011)

In this appendix we discuss the closely related “honours thesis” of O’Donnell (2011). First, we give a brief overview of his work which reveals major commonalities and differences to our setting. In doing so, we point out some rather general issues concerning his approach. Second, we go more into detail and focus on the proofs of crucial results presented and applied in O’Donnell (2011, Chapter 4). Mainly by providing straightforward counterexamples we argue that there are several shortcomings in his line of argument. This unveils that his key result must be considered as being unproven which means that, contrary to what is suggested, O’Donnell does not provide a complete characterization of pairwise stable networks by far.⁷³

Though entitled “Preliminary Results” (O’Donnell, 2011, p. 41), Chapter 4 is rather fundamental for his work. The findings in this chapter are supposed to rule out the possibility to be pairwise stable for a broad range of networks. Though the line of argument is very different, regarding the underlying idea and significance for his work the chapter’s results are comparable to Theorem 3.2 and Theorem 3.3 in our work. O’Donnell states several lemmas whose proofs shall build on one another and which shall finally combine to the following main result.

“Theorem 4.1: *When $c > 0$, any link stable network G must be of degree two or less, meaning it is made up of circle segments, line segments, and disconnected agents.*” (O’Donnell, 2011, p. 44)

As we establish in Section 3.3, this is true for non-singularly pairwise stable networks. Though we do not prove this, our further results suggest that this is indeed even true for pairwise stable networks. However, there are various shortcomings within O’Donnell’s proofs of the lemmas as we point out in the second part of this appendix. Thus, the theorem might be true, however, based on the work of O’Donnell, it can at most be considered as a conjecture.

This is followed by a rather descriptive chapter characterizing pairwise stable networks (see O’Donnell, 2011, pp. 59–67). However, some questions remain open. For instance, it is not clarified why line networks of length greater than three are never pairwise stable. Also, the examination of odd circles is rather short and seems to be incomplete as only the special cases of three and five player circles are considered explicitly. Regarding its content and purpose, this chapter is comparable to our

⁷³O’Donnell uses the expression “link stable” synonymously for “pairwise stable”.

Theorem 3.1 and Corollary 3.1 where we establish sufficient conditions for networks to be pairwise stable.⁷⁴ There are no contradictions between both works here.

Next, O'Donnell (2011, pp. 69–78) focuses on “Nash stability” and “Pareto optimality”, however, while redefining these notions. Usually, a network is called Nash stable if it is supported by a Nash equilibrium in the non-cooperative network formation game à la Myerson (1991, p. 448).⁷⁵ This seems to be equivalent to what he calls “strong Nash stability” while his notion of Nash stability does not seem to be well-founded. Furthermore, in economics the notion of Pareto optimality or Pareto efficiency is commonly associated with a status quo where no player can improve without another one being worse off (see e.g. Pareto, 1964; Jackson, 2008b, p. 157). However, O'Donnell uses it as follows: “[...] examining the set of networks that is Pareto optimal, that is the set of networks that maximise the sum of the payoffs” (p. 69). This is rather a description of efficiency based on a utilitarian welfare notion that we consider in Section 3.4. However, different from O'Donnell we give a complete characterization of this class of networks.⁷⁶ Moreover, we consider the concept of pairwise Nash stability (see Corollary 3.5) which, different than one might suppose, does not coincide with any of his alternative notions of stability. Finally, note that, while examining the model for $\delta \rightarrow 1$, O'Donnell does not provide results for the case $\delta \in (0, 1)$ which we consider in Section 3.5.

In the following second part of this appendix, we state most of O'Donnell's lemmas which are supposed to combine to establish the key result in his Theorem 4.1. However, as already announced, we reveal substantial shortcomings in the corresponding proofs by providing appropriate counterexamples.

3.C.1 Lemma 4.1 of O'Donnell (2011)

“Lemma 4.1: *Let r'_s and r''_s be the minimum shortage ratios in the networks G' and G'' respectively where $G' \subset G''$. It must be that $r'_s \leq r''_s$.”* (O'Donnell, 2011, p. 45)

Note that O'Donnell uses capital letters when referring to networks. In what follows we adopt this notation. The following two counterexamples demonstrate that this statement is in general not true.

⁷⁴Note that, among other parts of the work at hand, I accomplished this (as well as the elaboration of the model) before I became aware of O'Donnell's thesis on March 06, 2014 and he kindly send it to me by e-mail on April 08, 2014 (as it is not publicly available).

⁷⁵Here, a Nash equilibrium denotes a strategy profile (s^1, s^2, \dots, s^n) where no player $i \in N$ wants to deviate from her strategy $s^i \in \{0, 1\}^{n-1}$ which, together with the other players' strategies, determines a network g with $jk \in g$ if and only if $s_k^j = s_j^k = 1$ (see also Bloch and Jackson, 2006).

⁷⁶See Footnote 74.

Counterexample 3.7. Consider the player set N with $n = 3$ and the networks $G' := \{12\}$ and $G'' := \{12, 13\}$.

Note that, as required, we have $G' \subset G''$. However, the algorithms $\mathcal{A}(G')$ and $\mathcal{A}(G'')$ yield $r'_1 = 0 < \frac{1}{2} = r''_1$ but $r'_2 = 1$ while r''_2 does not exist.

In the second counterexample, both algorithms stop after the same step but Lemma 4.1 is still violated.

Counterexample 3.8. Consider the player set N with $n = 8$ and the networks $G' := \{12, 23, 34, 45, 56, 67\}$ and $G'' := G' + 68 = \{12, 23, 34, 45, 56, 67, 68\}$ as sketched in Figure 3.7.

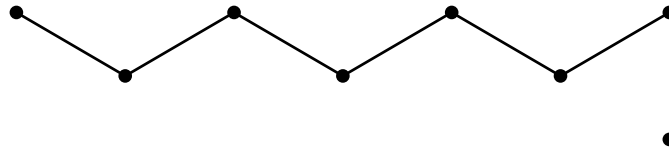


Figure 3.7: A sketch of the network G' considered in Counterexample 3.8

Again, we obviously have $G' \subset G''$. Now, the algorithms $\mathcal{A}(G')$ and $\mathcal{A}(G'')$ yield $r'_1 = 0 < \frac{1}{2} = r''_1$ but $r'_2 = \frac{3}{4} > \frac{2}{3} = r''_2$.

However, following O'Donnell's reasoning one can show that at least it holds that $r'_1 \leq r''_1$ for $G' \subset G''$.

3.C.2 Lemma 4.2 of O'Donnell (2011)

Next, we consider Lemma 4.2. First, one notices that the statements in the preceding explanation and in the lemma itself differ.

“**Lemma 4.2:** Let G be a network such that $ij \in G$ and $jk \notin G$. Then let $G' = \{ij | ij \in G\} \cup \{ik\}$. Then $v_i(G) \geq v_i(G')$.” (O'Donnell, 2011, p. 46)

Here, $v_i(\cdot)$ is equal to $v_i^*(\cdot)$ if one applies our notation. To avoid confusion, however, we again adopt O'Donnell's notation in what follows. The statement is in general not true as the next counterexample shows.

Counterexample 3.9. Consider the player set N with $n = 3$ and the network $G := \{12\}$.

Observe that for $i = 1, j = 2$ and $k = 3$ the network considered in the counterexample satisfies the conditions of Lemma 4.2. We then have $G' = \{12, 13\}$ and $v_i(G) = \frac{1}{2} < \frac{2}{3} = v_i(G')$ which contradicts the above statement.

If on the contrary we follow the explanation given as an introduction to the lemma, it should read as follows.

Lemma 4.2’: *Let G be a network such that $ij \in G$ and $jk \notin G$. Then let $G' = \{ij | ij \in G\} \cup \{jk\}$. Then $v_i(G) \geq v_i(G')$.*

This statement might be correct but there is a mistake in the proof. The author argues that in the first case, where it is $v_i(G) < \frac{1}{2}$, “player i ’s payoff does not change” (O’Donnell, 2011, p. 47) if $G' = G + jk$ is considered instead. This is not true as the following counterexample reveals.

Counterexample 3.10. *Consider the player set N with $n = 4$ and the network $G := \{12, 24\}$.*

Note that for $i = 1$, $j = 2$ and $k = 3$ the algorithm $\mathcal{A}(G)$ gives $v_i(G) = \frac{1}{3} < \frac{1}{2}$. Further we have $G' = G + jk = \{12, 23, 24\}$ and in this network the payoff of player i is changed to $v_i(G') = \frac{1}{4}$.

3.C.3 Lemma 4.3 of O’Donnell (2011)

To prepare for the following considerations, the set $L_i(G) \subset N$ “of players to whom player i is connected to in G plus himself” (O’Donnell, 2011, p. 45) is defined. Thus, we have $L_i(G) = N_i(G) \cup \{i\}$ here, which implies $|L_i(G)| = \eta_i(G) + 1$.

“**Lemma 4.3:** *In any link stable network there exists a maximum number of links any single player can have depending on c . This number is determined by the following inequality. It is possible for a player to have L_i links in a link stable network only if*

$$\frac{1}{(L_i + 1)L_i} - c \geq 0.”$$

(O’Donnell, 2011, p. 48)

Possibly, he refers to a (link stable) network G and a player i here and has $|L_i(G)|$ in mind when writing L_i . The corresponding proof, however, is not convincing. The author argues that Lemma 4.2 proves that it is sufficient to confine oneself to star networks “in which the potential partners of i under consideration have no other connections in the network” (O’Donnell, 2011, p. 48). Even if one assumes that this is justified and neglects the fact that Lemma 4.2 or Lemma 4.2’ have to be considered as unproven (see Appendix 3.C.2), a reasonable explanation is still missing here.

3.C.4 Lemmas 4.5 and 4.6 of O'Donnell (2011)

“**Lemma 4.5:** *It is not possible to have a link stable network where $0 < r_s < \frac{1}{2}$, regardless of $c > 0$.*” (O'Donnell, 2011, p. 51)

“**Lemma 4.6:** *It is not possible to have a link stable network where $\frac{1}{2} < r_s < 1$, regardless of $c > 0$.*” (O'Donnell, 2011, p. 53)

At first notice that it is not clear, which step s of the algorithm is considered here. As we know from Corollary 3.2, the algorithm $\mathcal{A}(G)$ has to break off after the first step $s = 1$ if G is pairwise stable and consisting of only one component. Therefore, let us assume that O'Donnell refers to r_1 in both lemmas. Under this assumption the statements in both lemmas might indeed be true. However, they are again not proven properly. Given partner set L , G -independent set M and players $j, k \in M$, the author argues within the proof of Lemma 4.5 that “if j and k were to link, then j would receive:

$$\frac{\frac{L+1}{M-1}}{1 + \frac{L+1}{M-1}} - L_j c ”$$

(O'Donnell, 2011, p. 51). Assuming that he means $|L|$, $|M|$ and $|L_j(G)|$ here, the following counterexample shows that this is not true.

Counterexample 3.11. *Consider the player set N with $n = 5$ and the star network $G := \{12, 13, 14, 15\}$.*

Note that for $j = 2$ and $k = 3$ it is $j, k \in M$ here. Now consider $G' := G + jk$. Then the algorithm $\mathcal{A}(G')$ gives

$$v_j(G + jk) = \frac{1}{2} > \frac{2}{5} = \frac{\frac{|L|+1}{|M|-1}}{1 + \frac{|L|+1}{|M|-1}}$$

which contradicts the above statement. Moreover, we know from Lemma 3.1 that for a pairwise stable network G with $r_1 < 1$ and $j, k \in M$ we always have $v_j(G + jk) = \frac{1}{2}$. O'Donnell uses this in the proof of Lemma 4.6, however, without a proof, when stating that “if j and k were to link, then j would receive: $\frac{1}{2} - L_j c$ ” (p. 54). Beyond that, in the proof of Lemma 4.5 he infers that “ $0 < r_s < \frac{1}{2}$ implies $\frac{L+1}{M-1} < 1$ ” (p. 51). As it can again easily be seen from the following counterexample this is not true.

Counterexample 3.12. *Consider the player set N with $n = 4$ and the star network $G := \{12, 13, 14\}$.*

For this network the algorithm $\mathcal{A}(G)$ yields $r_1 = \frac{1}{3} < \frac{1}{2}$ but we have $\frac{|L_1|+1}{|M_1|-1} = 1$.

3.C.5 Lemma 4.7 of O’Donnell (2011)

“**Lemma 4.7:** *If G is a link stable network where $r_s = \frac{1}{2}$, then $c = \frac{1}{6}$.*” (O’Donnell, 2011, p. 55)

If we assume as before that the author refers to step $s = 1$ here, then according to our Theorem 3.3 this statement is true. However, even if we assume that all statements in the previous lemmas were true (and proven properly), the proof of Lemma 4.7 is again not exhaustive. To be more precise, the reasoning that “ $\frac{1}{2} = r_s$ implies one or more disjoint line segments of three players” (O’Donnell, 2011, p. 55) is not correct. For instance, this becomes clear from examining the following counterexample.

Counterexample 3.13. *Consider the player set N with $n = 6$ and the network $G := \{12, 13, 24, 34, 45, 46\}$ as sketched in Figure 3.8.*

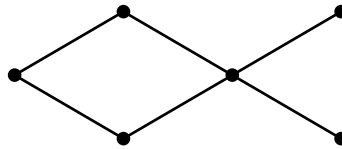


Figure 3.8: A sketch of the network G considered in Counterexample 3.13

For the considered network the algorithm $\mathcal{A}(G)$ yields $r_1 = \frac{1}{2}$ though it does not contain a (disjoint) line of length three.

3.C.6 Lemma 4.9 of O’Donnell (2011)

In his Lemma 4.9, O’Donnell (2011, p. 57) considers components of pairwise stable networks which are neither isolated players nor lines. For such a component G' O’Donnell states the following.

“**Lemma 4.9:** *The network G' is of degree two.*”⁷⁷ (O’Donnell, 2011, p. 57)

Though we suppose that this is indeed true, the following crucial inference in the proof of Lemma 4.9 is not correct. O’Donnell argues that “a player i in the smallest circle segment, [...] having more than two links” (pp. 57–58) and receiving, just as every other player, a payoff of $\frac{1}{2}$ is “strictly worse off than if she were just to have the

⁷⁷A network is said to be of degree two if every player in this network has exactly two links.

two links that keep him in the circle” (p. 58).⁷⁸ This does not need to be the case as the following counterexample reveals.

Counterexample 3.14. Consider the player set N with $n = 9$ and the network $G' := \{12, 17, 19, 23, 34, 45, 47, 56, 67, 78, 89\}$ as sketched in Figure 3.9. Further let $c \in (0, \frac{1}{18}]$.

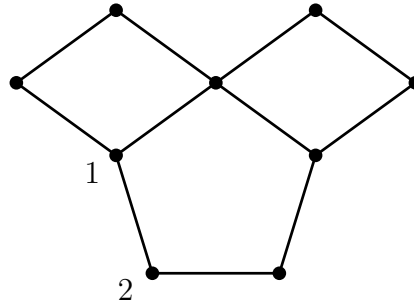


Figure 3.9: A sketch of the network G' considered in Counterexample 3.14

Note that player 1 has more than two links and is contained in the smallest circle segment, that is in one of the two smallest subnetworks of G' which are a circle. Evaluating the algorithm $\mathcal{A}(G')$ gives that, as required, every player receives a payoff of $\frac{1}{2}$ in G' . Further, if player 1 deletes her link to player 2, she would still be contained in the smallest circle (segment). However, if we consider the network $G' - 12$ and apply the algorithm $\mathcal{A}(G' - 12)$, we get

$$v_1(G' - 12) - 2c = \frac{4}{9} - 2c \leq \frac{1}{2} - 3c = v_1(G') - 3c.$$

Thus, player 1 is not worse off in G' compared to $G' - 12$.

In summary, one can state that, concerning the characterization of pairwise stable networks, O'Donnell's work is roughly comparable with the findings we derive in Theorem 3.1 and Corollary 3.1 if one only considers the results he established properly and correctly. In this context, notice again Footnote 74.

⁷⁸A “circle (segment)” here is meant to be a (not necessarily component-induced) subnetwork which is a circle.

Chapter 4

Continuous Homophily and Clustering in Random Networks

4.1 Introduction

Suppose you own a firm and want to fill an open vacancy through the social contacts of one of your current employees. Whom would you ask to recommend someone? Most probably you would address the worker who would himself perform best in the position in question. While this seems to be intuitively reasonable, why do we expect it to be optimal? One important reason is that people tend to connect to similar others. This phenomenon is known as homophily (Lazarsfeld and Merton, 1954).

In this chapter, we introduce a continuous notion of homophily based on incorporating heterogeneity of agents into the Bernoulli Random Graph (BRG) model as examined by Erdős and Rényi (1959). To this end, we propose a two-stage random process which we call *Homophilous Random Network* model. First, agents are assigned characteristics independently drawn from a continuous interval and second a network realizes, linking probabilities being contingent on a homophily parameter and the pairwise distance between agents' characteristics. This enables us to account for homophily in terms of similarity rather than equality of agents, capturing the original sociological definition instead of the stylized version up to now commonly used in the economic literature.

As a first result, we determine the expected linking probabilities between agents (Proposition 4.1) as well as the expected number of links (Corollary 4.2). We then calculate the expected probability that an agent has a certain number of links (Proposition 4.2), showing that the according binomial distribution of the original BRG

model is preserved to some degree. Further, we establish a threshold theorem for any given agent to be connected (Proposition 4.3). For all these (and further) results we demonstrate that the BRG model is comprised as the limit case of no homophily and we thus provide a generalization thereof. As a main result, we show that in our model homophily induces clustering (Theorem 4.1), two stylized facts frequently observed in real-world networks which are not captured by the BRG model.⁷⁹ Furthermore, clustering proves to be strictly increasing in homophily. As a second important feature of our model, two simulations indicate that even at high homophily levels the well-known small-world phenomenon is preserved.⁸⁰ We finally provide an application of the Homophilous Random Network model within a stylized labor market setting to answer the introductory questions.

In the literature the presence of homophily has been established in a wide range of sociological and economic settings. Empirical studies on social networks discovered strong evidence for the similarity of connected individuals with respect to age (see e.g. Verbrugge, 1977; Marsden, 1988; Burt, 1991), education (see e.g. Marsden, 1987; Kalmijn, 2006), income (see e.g. Laumann, 1966, 1973), ethnicity (see e.g. Baerveldt et al., 2004; Ibarra, 1995) or geographical distance (see e.g. Campbell, 1990; Wellman, 1996). For an extensive survey see McPherson et al. (2001). In recent years, economists have developed an understanding of the relevance of network effects in a range of economic contexts. Thus, bearing in mind the presence of homophily in real-world networks can be of great importance for creating meaningful economic models.

There already exists a strand of economic literature examining homophily effects in different settings (see e.g. Currarini et al., 2009). Most of the models assume a finite type space and binary homophily in the sense that an agent prefers to connect to others that are of the same type while not distinguishing between other types.⁸¹ Thus, these models rather capture the idea of equality than of similarity. However, in reality people are in many respects neither “equal” nor “different”. We therefore believe that a notion that provides an ordering of the “degree of similarity” with respect to which an agent orders his preference for connections can capture real-world effects more accurately. This gives rise to a continuous notion of homophily in networks.

⁷⁹A network exhibits clustering if two individuals with a common neighbor have an increased probability of being connected.

⁸⁰The small-world phenomenon describes the observation that even in large networks on average there exist relatively short paths between two individuals.

⁸¹For several homophily measures of this kind see Currarini et al. (2009).

This approach is followed by Gilles and Johnson (2000) and Iijima and Kamada (2014) who examine strategic, deterministic models of network formation. In both models individual utility is shaped directly by homophily such that individuals connect if (and only if) they are sufficiently similar. Iijima and Kamada (2014) consider the extreme case of purely homophilous utility functions, entailing that a high level of homophily is directly identified with efficiency. As opposed to this, in our random graph model, a novel continuous homophily measure is incorporated as a parameter that may be freely chosen to reflect a broad range of possible situations. In their multi-dimensional framework, Iijima and Kamada (2014) examine clustering and the average path length as functions of the number of characteristics agents take into account when evaluating their social distance to others. In contrast, we investigate the direct relation between homophily and these network statistics. The differences in methodology especially lead to opposing results concerning the small-world phenomenon. While in Iijima and Kamada (2014) small worlds only arise if agents disregard a subset of characteristics, we show that this phenomenon is well present in our one-dimensional setting.

Besides the presence of homophily, stylized facts such as the small-world phenomenon and high levels of clustering have indeed been empirically identified in real-world networks (see e.g. Milgram, 1967; Watts and Strogatz, 1998). As in many cases these networks are very large and remain unknown for an analysis, typically random networks are used as an approximation. This constitutes a challenge to design the random network formation process in a way to ensure it complies with the observed stylized facts.

Since the seminal work of Erdős and Rényi (1959), who developed and analyzed a random graph model where a fixed number out of all possible bilateral connections is randomly chosen, a lot of different models have been proposed (see e.g. Wasserman and Pattison, 1996; Watts and Strogatz, 1998; Barabási and Albert, 1999). The most commonly used until today is the BRG model where connections between any two agents are established with the same constant probability. It has been shown that for large networks this model is almost equal to the original model of Erdős and Rényi (1959) (for details see Jackson, 2006; Bollobás, 2001).⁸² It is well understood that this model reproduces the small-world phenomenon but does not exhibit clustering. Also, a notion of homophily is not present as the described random process does not rely on individual characteristics. The latter is also true for the small-world model proposed by Watts and Strogatz (1998). Starting from a network built on a

⁸²In fact, the BRG model rather than their original one is nowadays also known as the Erdős-Rényi model.

low-dimensional regular lattice, they reallocate randomly chosen links and obtain a random network showing a small-world phenomenon. According to their notion this encompasses an increased level of clustering. However, the socio-economic causality of this occurrence remains uncertain. In this regard our model can to some extent serve as a socio-economic foundation of the work of Watts and Strogatz (1998). An approach to generate random graphs more similar to ours is proposed by the recently emerging graph-theoretic literature on random intersection graphs (see e.g. Karonski et al., 1999). Here, each node is randomly assigned a set of features. Connections are then established between any two nodes sharing a given number of features. It has been shown that the resulting graphs also exhibit clustering (Bloznelis, 2013).

In general, not much work has yet been dedicated to the incorporation of homophily into random networks. However, some papers exist that include similar ideas. Jackson (2008a) analyzes the impact of increasing homophily on network statistics such as clustering and the average distance of nodes. A finite number of types as well as linking probabilities between them are exogenously given. Though linking probabilities may vary among types, which allows for cases where similar types are preferred, his notion of homophily remains binary. Golub and Jackson (2012) also assume a finite number of types as well as the linking probabilities between them to be exogenously given. Based on this they analyze the implications of homophily in the framework of dynamic belief formation on networks. Bramoullé et al. (2012) combine random link formation and local search in a sequentially growing society of heterogeneous agents and establish a version of binary homophily along with a degree distribution. Besides the continuous notion of homophily, a major distinction of our approach is the sequential combination of two random processes where agents' characteristics are considered as random variables that influence the random network formation. We thus account for the fact that in many applications, in which the network remains unobserved, it seems unnatural to assume that individual characteristics, which in fact may depict attitudes, beliefs or abilities, are perfectly known.

We conclude this chapter by providing an application of our model for the labor market, proposing an analysis of the introductory question: When is it optimal for a firm to search for a new employee via the contacts of a current employee? We assume the characteristic of each worker to be her individual ability to fill the open vacancy and use our Homophilous Random Network model as an approximation of the workers' network. Given an agent and her characteristic, we determine the expected characteristic of a random contact (Proposition 4.4). This gives rise to a

simple decision rule stating in which constellations firms should hire via the social network. In particular, given sufficiently high levels of homophily and the current employee's ability, it proves to be optimal to always hire via the social network.

Within the job search literature, Horváth (2013) and Zaharieva (2013) incorporate homophily among contacts into job search models. However, these models are again based on a binary concept of homophily and do not include an explicit notion of networks. This research strand traces back to the work of Montgomery (1991) who was the first to address this issue. Finally, our application to some extent captures an idea proposed by Ioannides and Loury (2004) to combine this class of models with a random network setting à la Erdős-Rényi.⁸³

The rest of the chapter is organized as follows. In Section 4.2 we set up the model. Section 4.3 reveals basic properties of homophilous random networks while results on clustering can be found in Section 4.4. In Section 4.5 we simulate the model focusing on the small-world phenomenon. Section 4.6 contains the labor market application and Section 4.7 concludes. Proofs of most results are provided in the appendix.

4.2 The Model

We set up a model of random network formation where first each agent is randomly assigned a continuous characteristic which then influences the respective linking probabilities. We refer to this as the *Homophilous Random Network* model. Consider a set of agents $N = \{1, 2, \dots, n\}$. A *connection* or (*undirected*) *link* between two agents $i, j \in N$ is denoted by $ij = ji := \{i, j\}$. By $g^N := \{ij \mid i, j \in N\}$ we denote the complete network, that is the network where any two agents are connected. Then, we let $\mathbb{G} := \{g \mid g \subseteq g^N\}$ be the set of all possible *non-directed graphs* or *networks*. Further, we define $N_i(g) := \{j \in N \mid ij \in g\}$ to be the set of *neighbors* of agent i in network g , and let $\eta_i(g) := |N_i(g)|$ denote the number of her neighbors. This is sometimes also referred to as the *degree* of agent i . Each agent is assigned a *characteristic* p_i where the vector $p = (p_1, p_2, \dots, p_n)$ denotes a certain realization of the random variable $P = (P_1, P_2, \dots, P_n)$. The underlying distribution of each P_i is assumed to be standard uniform. Hence, all P_i are identically and independently distributed.

⁸³Ioannides and Loury (2004, p. 1068) state “It would be interesting to generalize the model of social structure employed by Montgomery, by assuming groups of different sizes. For example, one may invoke a random graphs setting (Paul Erdős and Alfred Rényi 1960; Ioannides 1997), where a fraction of the entire economy may be in groups whose sizes are denumerable but possibly large.”

Subsequent to the assignment of characteristics a random network forms. Here, based on the Bernoulli Random Graph (BRG) model as introduced by Erdős and Rényi (1959), we assume the following variation. The *linking probability* of two agents $i, j \in N$ is given by

$$q(p_i, p_j) := \lambda a^{|p_i - p_j|}, \tag{4.1}$$

where the *scaling parameter* $\lambda \in [0, 1]$ and the *homophily parameter* $a \in [0, 1]$ are exogenously given and independent of agents i and j . Note that, in situations where the vector of characteristics is unknown, $q(P_i, P_j)$ is a random variable such that the linking probability $q(p_i, p_j)$ is in fact a conditional probability. Figure 4.1 depicts the linking probabilities $q(p_i, p_j)$ for different homophily parameters a , first as a function of the distance of characteristics and second as a function of p_j for given $p_i = 0.25$. As in our model λ simply serves as a scaling parameter corresponding to the linking probability in the BRG model, in Figure 4.1 it is fixed to one for simplicity.

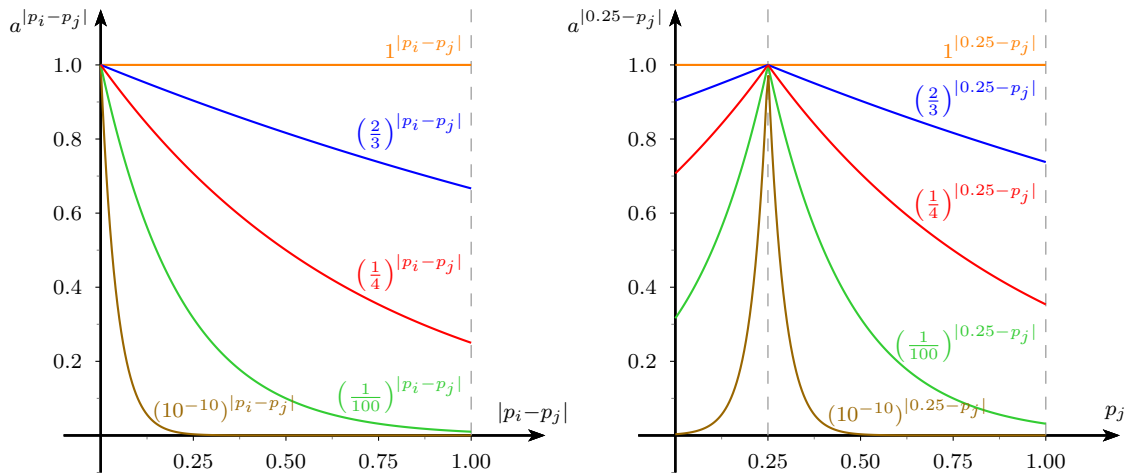


Figure 4.1: (a) Linking probability for all distances of characteristics for several homophily parameters a ; (b) Linking probabilities for an agent with characteristic $p_i = 0.25$ for several homophily parameters a

Let us shortly elaborate on the role of the homophily parameter a . Observe that the linking probability q is decreasing in $|p_i - p_j|$ as a takes values only in $[0, 1]$. In particular, for $a = 1$ the model is equal to the BRG model as all linking probabilities are equal to λ and hence independent of the agents' characteristics. On the contrary, if we have $a = 0$, then solely agents with identical characteristics $p_i = p_j$ get connected with probability λ while all other linking probabilities are zero. Insofar, the parameter a serves as a measure of homophily in the model. Here, lower

parameter values correspond to a higher homophily level in the network. The notion at hand measures homophily in a continuous instead of a binary manner since the distance function $|\cdot|$ is continuous. Note, however, that an increase in homophily which leads to a decreased linking probability then also implies a decreased expected degree of agents. Whenever suitable, one may therefore choose the scaling parameter λ dependent on a such that the expected degree is kept constant for any level of homophily.⁸⁴ We make use of this possibility in Section 4.5 (see Simulation 4.1).

4.3 Basic Properties of Homophilous Random Networks

This section constitutes a foundation for the upcoming main results. To this end, we first need to collect several important properties of the Homophilous Random Network model, such as the expected linking probabilities and the number of links of agents. Moreover, we discuss a threshold theorem for an agent to be isolated. This is of particular importance for the labor market application provided in Section 4.6. Throughout this section we explore, on the one hand, situations in which the realization of one considered agent $i \in N$ is known while all others are not and, on the other hand, situations in which the whole vector of characteristics is unknown. In any case we demonstrate that the BRG model is recuperated as the limit case of no homophily and we thus provide a generalization thereof.

We start by determining the expected linking probabilities for two given agents $i, j \in N$ in the following proposition.

Proposition 4.1. *Given agent i 's realized characteristic $P_i = p_i$ while all other characteristics p_{-i} are unknown, the expected probability that a certain link ij forms is*

$$\mathbb{E}^P \left[\mathbb{P}^G (ij \in G \mid P) \mid P_i = p_i \right] = \frac{\lambda}{\ln(a)} \left(a^{p_i} + a^{1-p_i} - 2 \right) =: \varphi(\lambda, a, p_i). \quad (4.2)$$

If the vector p is unknown, the expected probability that the link ij forms is

$$\mathbb{E}^P \left[\mathbb{P}^G (ij \in G \mid P) \right] = \frac{2\lambda}{\ln(a)^2} \left(a - 1 - \ln(a) \right) =: \Phi(\lambda, a). \quad (4.3)$$

⁸⁴According to Corollary 4.2, choosing $\lambda = \frac{\eta^{exp} \ln(a)^2}{2(n-1)(a-1-\ln(a))}$ yields a fixed expected degree of η^{exp} (if compatible).

The proof of Proposition 4.1 as well as all subsequent proofs can be found in the appendix. It is straightforward to understand that the function φ indeed has to depend on characteristic p_i as it makes a difference whether p_i tends to the center or to the boundaries of the interval $[0, 1]$. The closer p_i is to 0.5 the smaller is the expected distance to other agents' characteristics, hence, the higher is the expected linking probability φ . In particular, it is $\arg \max_{p_i} \varphi = 0.5$ and $\arg \min_{p_i} \varphi = \{0, 1\}$ for all $a \in (0, 1)$. To this respect, it is obvious that $\varphi(\lambda, a, 0) \leq \Phi(\lambda, a) \leq \varphi(\lambda, a, 0.5)$ for all $\lambda, a \in [0, 1]$. Also, it is important to note that the expected linking probability is decreasing in homophily, that is for all $a \in (0, 1]$ we have

$$\frac{\partial}{\partial a} \Phi(\lambda, a) = \frac{\partial}{\partial a} \left[2\lambda \frac{a - 1 - \ln(a)}{\ln(a)^2} \right] = 2\lambda \frac{2(1 - a) + \ln(a)(1 + a)}{a \ln(a)^3} > 0.^{85}$$

To verify intuition that our model reproduces the BRG model as a limit case and to gain insights on the behavior in boundary cases, the following corollary is concerned with the limits of the expected linking probabilities with respect to the homophily parameter a .

Corollary 4.1. *For maximal homophily, i.e. for $a \rightarrow 0$, the expected linking probability is*

$$\lim_{a \rightarrow 0} \varphi(\lambda, a, p_i) = \lim_{a \rightarrow 0} \Phi(\lambda, a) = 0. \quad (4.4)$$

In case of no homophily, i.e. for $a \rightarrow 1$, the expected linking probability is

$$\lim_{a \rightarrow 1} \varphi(\lambda, a, p_i) = \lim_{a \rightarrow 1} \Phi(\lambda, a) = \lambda. \quad (4.5)$$

As usual, a proof is provided in the appendix. Maximal homophily in this model means that only agents with identical characteristics would have a strictly positive linking probability. However, since the standard uniform distribution has no mass point, such two agents do not exist with positive probability. Therefore, both according expected linking probabilities φ and Φ tend to zero. In case of no homophily, as mentioned before, the model indeed reproduces the BRG model such that all linking probabilities are alike, independent of individual characteristics p .

Based on Proposition 4.1, we also immediately get the expected number of links of an agent.

⁸⁵We indeed can include the value $a = 1$ here as it happens to be a removable discontinuity of the derivative. On the contrary, at $a = 0$ the right-handed derivative is infinity as the expected number of links is zero with probability one.

Corollary 4.2. *The expected number of links of an agent i with given characteristic $P_i = p_i$ is*

$$\mathbb{E}^P \left[\mathbb{E}^G [\eta_i(G) \mid P] \mid P_i = p_i \right] = (n - 1)\varphi(\lambda, a, p_i). \quad (4.6)$$

Similarly, if p is unknown, we have

$$\mathbb{E}^P \left[\mathbb{E}^G [\eta_i(G) \mid P] \right] = (n - 1)\Phi(\lambda, a). \quad (4.7)$$

A proof of this corollary is omitted as it is clear that all expected linking probabilities are independent and, hence, the result follows directly from the proof of Proposition 4.1. Observe that from this result, we can also calculate the expected number of links in a network to be

$$\frac{n(n - 1)}{2}\Phi(\lambda, a).$$

Together with Corollary 4.1 this gives that the expected number of links is zero for maximal homophily while in case of no homophily, again as in the BRG model, one gets $\lambda n(n - 1)/2$ links in expectation.

In what follows, we calculate the expected probability for an agent with given characteristic to have a certain number of links. This entails that the model inherits a version of the binomial distribution known from the BRG model.

Proposition 4.2. *The expected probability that an agent i with given characteristic $P_i = p_i$ has $k \in \{0, 1, \dots, n - 1\}$ links is given by*

$$\mathbb{E}^P \left[\mathbb{P}^G (\eta_i(G) = k \mid P) \mid P_i = p_i \right] = \binom{n - 1}{k} \cdot \varphi(\lambda, a, p_i)^k \cdot (1 - \varphi(\lambda, a, p_i))^{n-k-1}. \quad (4.8)$$

Observe that this form can be interpreted as a binomial distribution with parameters $\varphi(\lambda, a, p_i)$ and $n - 1$. Further, it is worth noting that the extreme cases meet the expected outcome as we have

$$\lim_{a \rightarrow 0} \mathbb{E}^P \left[\mathbb{P}^G (\eta_i(G) = k \mid P) \mid P_i = p_i \right] \stackrel{(4.4)}{=} \binom{n - 1}{k} \cdot 0^k \cdot 1^{n-k-1} = \begin{cases} 1, & \text{if } k = 0 \\ 0, & \text{else} \end{cases},$$

$$\lim_{a \rightarrow 1} \mathbb{E}^P \left[\mathbb{P}^G (\eta_i(G) = k \mid P) \mid P_i = p_i \right] \stackrel{(4.5)}{=} \binom{n - 1}{k} \cdot \lambda^k \cdot (1 - \lambda)^{n-k-1},$$

where the latter term, unsurprisingly, is equal to the probability for any agent to have k links in the BRG model with independent linking probability λ . Unfortunately, the calculation in case that the whole vector of characteristics p is unknown is analytically not tractable.

One major reason why random network models are used frequently is to match qualitative characteristics of real world networks. The Law of Large Numbers in this case yields that large networks indeed meet these characteristics with a high probability (see e.g. Jackson, 2008b, Chapter 4). A seminal contribution of Erdős and Rényi (1959) was to provide so called threshold theorems for the case of the BRG model. These results state that, if the network size n goes to infinity while the linking probability $\lambda(n)$ goes to zero slower than some threshold $t(n)$, then the limit network has a certain property with probability one. On the contrary, if $\lambda(n)$ goes to zero faster than $t(n)$, then the limit network has the same property only with probability zero.⁸⁶ It is clear that this kind of results can only be found for monotone properties, that is for those which yield that, if any network g has the property, then also any network $g' \supseteq g$ has it. One example is the property that a given agent has at least one link which we establish in the next proposition. For instance, regarding our application of the labor market (Section 4.6) this feature is of great importance. In that context, we assume this as a prerequisite as determining the expected characteristic of a given agent's contact is meaningful only if this agent is not isolated.

Proposition 4.3. *Assume a minimal level of homophily to be guaranteed as the network size becomes large. Then the function $t(n) = 1/(n - 1)$ is a threshold for a given agent to be non-isolated in the following sense:*

$$\begin{aligned} \mathbb{E}^P \left[\mathbb{P}^G (\eta_i(G) \geq 1 \mid P) \mid P_i = p_i \right] &\rightarrow 1 \quad \forall p_i \in [0, 1] \quad \text{if} \quad \frac{-\lambda(n)/\ln(a(n))}{t(n)} \rightarrow \infty, \\ \mathbb{E}^P \left[\mathbb{P}^G (\eta_i(G) \geq 1 \mid P) \mid P_i = p_i \right] &\rightarrow 0 \quad \forall p_i \in [0, 1] \quad \text{if} \quad \frac{-\lambda(n)/\ln(a(n))}{t(n)} \rightarrow 0. \end{aligned}$$

First, note that in Proposition 4.3 the right-hand side conditions are equivalent to $\varphi(\lambda(n), a(n), \hat{p})/t(n)$ converging to infinity or zero, respectively, for any arbitrary $\hat{p} \in [0, 1]$. For details refer to the proof in the appendix. What is surprising about this (as well as about other threshold theorems), is the sharp distinction made by the threshold $t(n)$, in the sense that if the growth of probability φ passes the threshold $t(n)$, then the probability of any agent to be isolated changes “directly” from zero to

⁸⁶For a more elaborate characterization of thresholds as well as several results see Bollobás (1998).

one. What is more, notice that the threshold $t(n) = 1/(n-1)$ is actually the same as in the BRG model. However, it has to hold for φ rather than just for λ since in this model both λ and a may vary with respect to the size of the network. Indeed, it does not seem farfetched to assume that homophily increases with the network size as the assortment of similar agents gets larger. Having understood this, one can directly deduce the cases where only one of the two parameters varies with n .

Corollary 4.3. *If $a \equiv a(n)$ depends on n but λ does not, one gets that if $a(n)$ goes toward zero faster than $\exp(-n)$, then any given agent is isolated with probability one in the limit while if $a(n)$ does not go toward zero or at least slower than $\exp(-n)$, then any given agent has at least one link with probability one in the limit.*

If $\lambda \equiv \lambda(n)$ depends on n but a does not, the condition collapses to the threshold of $t(n)$ for $\lambda(n)$ as in the BRG model where any given agent has at least one link if $\lambda(n)$ grows faster than $t(n)$ while if $\lambda(n)$ grows slower than $t(n)$, any given agent is isolated with probability one.

Both parts of the corollary follow directly from Proposition 4.3 such that a proof can be omitted.

4.4 Clustering

As mentioned in the introduction, a main criticism of the Bernoulli Random Graph (BRG) model is that the resulting networks do not exhibit clustering while most examples of real-world networks do so (see e.g. Watts and Strogatz, 1998; Newman, 2003, 2006). In this section, we show that our Homophilous Random Network model indeed exhibits clustering and one can use the homophily parameter a to calibrate it to a broad range of degrees of clustering.

The notion of clustering in general captures the extent to which connections in networks are transitive, that is the frequency with which two agents are linked to each other given that they have a common neighbor. Watts and Strogatz (1998), who introduced this concept, measure the transitivity of a network by a global clustering coefficient which denotes the average probability that two neighbors of a given agent are directly linked as well. A random graph model is said to exhibit clustering if the coefficient is larger than the general, unconditional linking probability of two agents (see Newman, 2006). Considering the set of networks that contain some link $ij \in g^N$, that is $\mathbb{G}_{ij} := \{g \subseteq g^N \mid ij \in g\} \subset \mathbb{G}$, this can be transferred to our model in the following way:

Definition 4.1 (Clustering). *For the Homophilous Random Network model with $\lambda \in [0, 1]$ and $a \in (0, 1)$ the clustering coefficient is defined as*

$$C(\lambda, a) := \mathbb{E}^P \left[\mathbb{P}^G (G \in \mathbb{G}_{jk} \mid P) \mid G \in \mathbb{G}_{ij} \cap \mathbb{G}_{ik} \right]$$

where $i, j, k \in N$. The model is said to exhibit clustering if we have $C(\lambda, a) > \Phi(\lambda, a)$.

The choice of the agents i , j and k obviously cannot have an influence in this context since ex ante, i.e. before characteristics realize, all agents are assumed to be equal. Further, recall that Φ gives the probability of two agents to be connected, characteristics being unknown. The function C captures this probability as well, however, conditional on the existence of a common neighbor. It should be clear that the original BRG model does not exhibit clustering since every link is formed with the same independent probability. As a main result of this chapter, we discover next that, apart from the limit case of no homophily, our Homophilous Random Network model has this feature and is insofar more realistic.

Theorem 4.1 (Clustering in Homophilous Random Networks). *In the Homophilous Random Network model the clustering coefficient is given by*

$$C(\lambda, a) = \lambda \frac{3(\ln(a)a^2 + \ln(a) - a^2 + 1)}{2(2\ln(a)a + 4\ln(a) + a^2 - 8a + 7)}.$$

Given a non-extreme homophily parameter, the model exhibits clustering, that is we have

$$C(\lambda, a) > \Phi(\lambda, a)$$

for all $\lambda \in (0, 1]$, $a \in (0, 1)$.

The intuition for the proof of this theorem (which is again presented in the appendix) is the following: If there is homophily to some degree and two agents have a common neighbor, then this fact contains additional information. The expected distance between these two agents is smaller than if there is no assumption about a common neighbor. Again due to homophily, it is therefore more likely that a link between these two agents forms. Also, Figure 4.2 might contribute to a better understanding of the situation. Note here that $C(\lambda, a)/\lambda \equiv C(1, a)$ and $\Phi(\lambda, a)/\lambda \equiv \Phi(1, a)$. One can additionally perceive that the difference $C(\lambda, a) - \Phi(\lambda, a)$ is strictly decreasing in $a \in (0, 1)$ for all $\lambda \in (0, 1]$, that is clustering is strictly increasing in the degree of homophily.

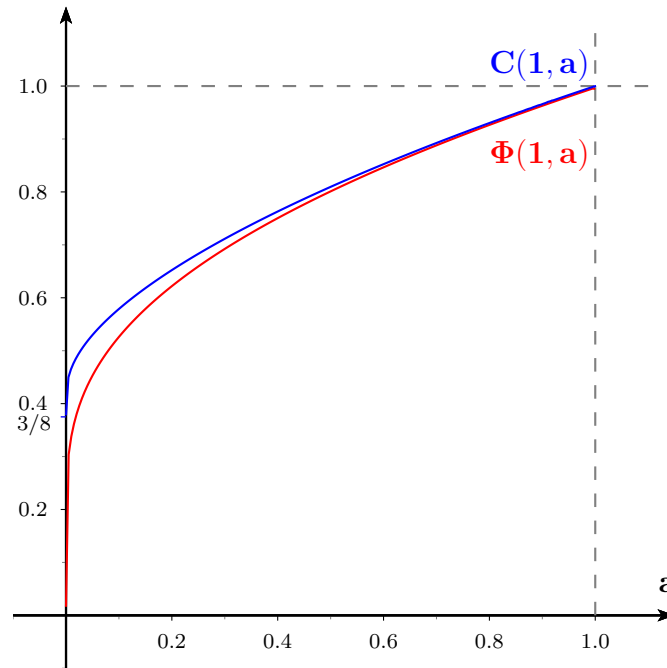


Figure 4.2: Clustering coefficient $C(1, a)$ and unconditional linking probability $\Phi(1, a)$ for all homophily parameters $a \in (0, 1)$

Again, it is of interest to consider the limit cases of maximal and no homophily which we do in the following corollary.

Corollary 4.4. *For maximal homophily, i.e. for $a \rightarrow 0$, we have*

$$\lim_{a \rightarrow 0} C(\lambda, a) = \lim_{a \rightarrow 0} [C(\lambda, a) - \Phi(\lambda, a)] = \frac{3}{8}\lambda.$$

In case of no homophily, i.e. for $a \rightarrow 1$, we get

$$\lim_{a \rightarrow 1} C(\lambda, a) = \lim_{a \rightarrow 1} \Phi(\lambda, a) = \lambda.$$

If there is no homophily, we are again back in the BRG model which we already know not to exhibit clustering. Insofar, the second part of the corollary is consistent. However, the more interesting case is the one of maximal homophily. Though in the limit no link forms with positive probability, one can deduce properties regarding the case of homophily being high, yet not maximal, due to continuity of the functional forms. Let us clarify the intuition why the clustering coefficient takes a value strictly between zero and λ if homophily is maximal. Recall first that we have $\lim_{a \rightarrow 0} \Phi(\lambda, a) = 0$ since for maximal homophily only agents with identical characteristics are linked with positive probability and such two agents exist with probability

zero. However, the clustering coefficient is a probability conditioned on the existence of links to a common neighbor. This additional information implies that either characteristics are equal or links have formed despite differing characteristics. Though both events occur only with probability zero, this does not preclude them *per se*. Having understood this, it should be clear that in the former case the probability of the third link would indeed be λ while in the latter case it would still be zero. Taken together, this yields $\lim_{a \rightarrow 0} C(\lambda, a) \in (0, \lambda)$. It remains surprising, however, that the clustering coefficient takes the specific value $\frac{3}{8}\lambda$.

4.5 The Small-World Phenomenon

Besides the presence of homophily and clustering, another stylized fact is frequently observed in real-world networks which is widely known as the small-world phenomenon. It captures the finding that, even in large networks, there typically exist remarkably short paths between two individuals. The original BRG model is known to reproduce this characteristic (see e.g. Bollobás, 2001; Chung and Lu, 2002).

Thus, in this section, we aim to establish the small-world phenomenon to be preserved in our Homophilous Random Network (HRN) model even in case of homophily being high. For this purpose, we present and analyze simulations of homophilous random networks as this issue seems to be no longer analytically tractable. Our simulations provide a strong indication that also in cases of high homophily the small-world phenomenon remains present. Additionally, we apply two alternative statistical notions of clustering. It turns out that their values are not significantly different from the analytical measure given in Definition 4.1. In the following, Figure 4.3 may already provide a first intuition regarding the differences between cases of high and low homophily. In particular, while the total number of links is almost the same in both simulated 100-agent networks, one observes clustering merely in the first case.

The notion of the small-world phenomenon usually grounds on the average shortest path length between all pairs of agents belonging to a network and having a connecting path. With regard to real-world networks the small-world phenomenon is a rather vague concept since it is typically based on subjective assessments of path lengths rather than on verifiable, definite criteria. However, most people will agree that the values for several real-world networks as for instance compiled by Watts and Strogatz (1998) and Newman (2003) are surprisingly low. Insofar, it could be said that most of these networks exhibit the small-world phenomenon. A formal definition of the small-world phenomenon applicable to most random network models is

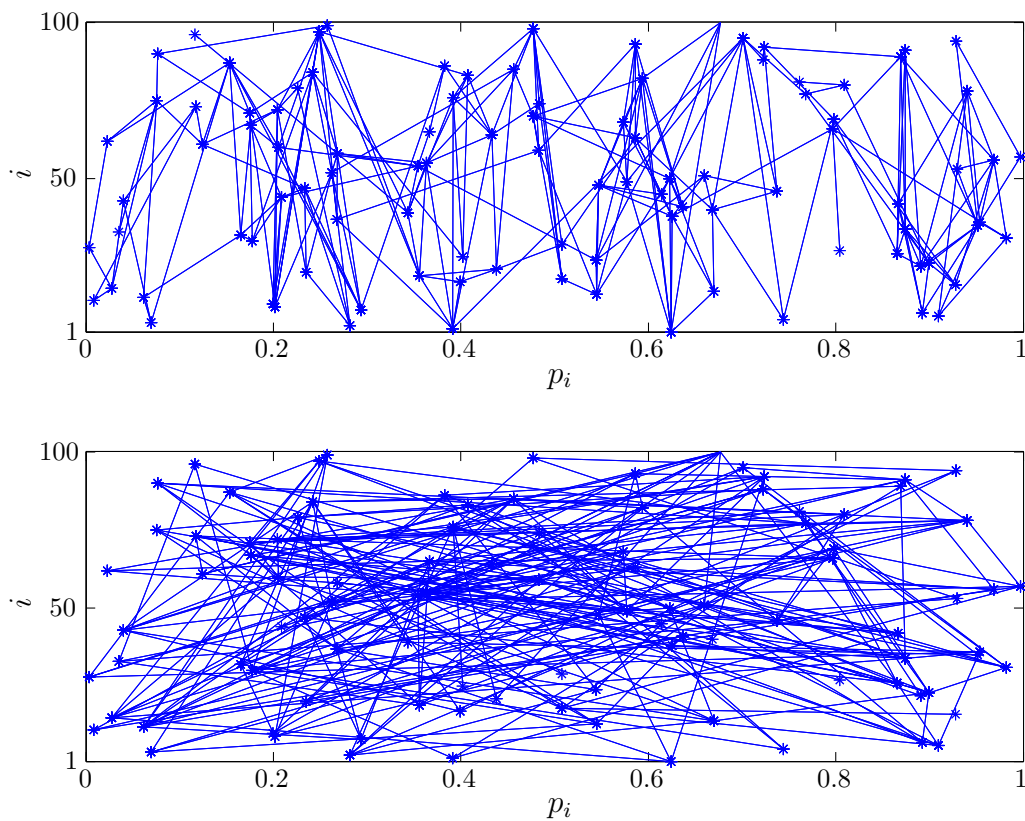


Figure 4.3: (a) HRN with $\lambda = 0.5$, $a = 10^{-8}$; $\#links = 484$
 (b) BRG with linking probability $\Phi(0.5, 10^{-8}) = 0.0513$; $\#links = 496$
 (created with MATLAB, 2014)

formulated by Newman (2003) and reads as follows:

Definition 4.2 (Small-World Phenomenon). *A random network is said to exhibit the small-world phenomenon if the average shortest path length \bar{d} between pairs of agents having a connecting path scales logarithmically or slower with network size n while keeping agents' expected degree constant, that is if $\bar{d}/\ln(n)$ is non-increasing in n .*

As already mentioned, it has been established that the original BRG model exhibits the small-world phenomenon according to Definition 4.2 (see e.g. Bollobás, 2001; Chung and Lu, 2002). It is not clear, however, whether this still holds for our generalization, given a considerably high level of homophily, but the results of the following simulations provide some indication.

Prior to this, let us additionally introduce two statistical notions of clustering which are frequently used in the literature and closely related to the one given in Definition 4.1. The simulations allow to compare these for our model. Here, clustering is associated with an increased number of triangles in the network. More precisely,

both alternative clustering measures are defined based on the ratio of the number of triangles and the number of connected triples. A triangle is a subnetwork of three agents all of whom being connected to each other while a connected triple is a subnetwork of three agents such that at least one of them is linked to the other two. Formally, this amounts to the following definition.

Definition 4.3 (Statistical Clustering). *For a given network with set of agents $N = \{1, \dots, n\}$, the (statistical) clustering coefficients $C^{(1)}$ and $C^{(2)}$ are determined by*

$$C^{(1)} := \frac{3 \times \text{number of triangles in the network}}{\text{number of connected triples in the network}} \quad \text{and}$$

$$C^{(2)} := \frac{1}{n} \sum_{i \in N} \frac{\text{number of triangles containing agent } i}{\text{number of connected triples centered on agent } i}.$$

The coefficient $C^{(1)}$ counts the overall number of triangles and relates it to the overall number of connected triples in the network. The factor of three accounts for the fact that each triangle contributes to three connected triples. The second one, $C^{(2)}$, which goes back to Watts and Strogatz (1998), first calculates an individual clustering coefficient for each agent and then averages these. Compared to the first one, $C^{(2)}$ gives more weight to low-degree agents.⁸⁷ Additionally, note that $C^{(2)}$ is only well-defined if there are no isolated or loose-end agents in the network.

To capture both the heuristic and the formal approach to the small-world phenomenon, we present the outcomes of two different simulations. In the first one, we fix the number of agents $n = 500$ and the ex-ante expected degree of any agent i , here denoted by η^{exp} , to $\eta^{exp} := \mathbb{E}[\eta_i] = 15$. Furthermore, we select several homophily levels ranging from no homophily, i.e. the limit case of the BRG model, to very high homophily, represented by $a = 10^{-8}$. For each parameter value of a , we then simulate a homophilous random network $R = 1000$ times and assess the averaged network statistics. The parameters and network statistics of the simulation are stated in Table 4.1. We fix the expected degree by choosing $\lambda = \frac{15 \ln(a)^2}{2(n-1)(a-1-\ln(a))}$ (recall Corollary 4.2) which enables us to compare the results for different homophily levels as this leads to identical values for $\Phi(\lambda, a)$ in all cases. Recall that Φ captures the expected probability of two agents to be connected, characteristics being unknown (recall Proposition 4.1).

Regarding the results of the simulation, we find that the average path length increases in homophily. This is in line with intuition since agents with distant char-

⁸⁷Referring to $C^{(2)}$, Newman (2003, p. 184) states “This definition effectively reverses the order of the operations of taking the ratio of triangles to triples and of averaging over vertices – one here calculates the mean of the ratio, rather than the ratio of the means.”

Parameter / Statistics	$a = 1$	$a = 10^{-2}$	$a = 10^{-4}$	$a = 10^{-6}$	$a = 10^{-8}$
n			500		
R			1000		
Exp. Degree η^{exp}			15		
Exp. Linking Prob. Φ			0.0301		
λ	0.0301	0.0882	0.1553	0.2239	0.2928
Avg. Degree $\bar{\eta}$	14.9990 (0.2475)	15.0074 (0.3064)	15.0098 (0.2986)	14.9899 (0.2925)	15.0037 (0.2839)
Avg. Shortest Path \bar{d}	2.5944 (0.0113)	2.6288 (0.0164)	2.8086 (0.0277)	3.0806 (0.0429)	3.3939 (0.0611)
$\bar{d}/\ln(n)$	0.4175 (0.0018)	0.4230 (0.0026)	0.4519 (0.0045)	0.4957 (0.0069)	0.5461 (0.0098)
Clustering Coeff. C	0.0301	0.0411	0.0641	0.0892	0.1147
Clustering Coeff. $C^{(1)}$	0.0301 (0.0013)	0.0411 (0.0016)	0.0642 (0.0023)	0.0891 (0.0029)	0.1147 (0.0035)
Clustering Coeff. $C^{(2)}$	0.0301 (0.0015)	0.0411 (0.0019)	0.0642 (0.0026)	0.0892 (0.0032)	0.1148 (0.0039)

Table 4.1: Results of Simulation 4.1 comparing network statistics for different homophily levels ranging from no homophily (BRG) to extreme homophily; Standard errors stated in parentheses (carried out with MATLAB, 2014)

acteristics are increasingly likely to be distant in the network. However, it increases by less than one link from no to highest homophily. Also, an average distance of less than 3.4 between two agents can still be considered relatively small in a network of 500 agents with about 15 links on average. Thus, regarding the heuristic approach, it seems reasonable to accept the small-world phenomenon to be exhibited for all homophily levels.⁸⁸

Furthermore, we observe an increasing level of clustering for the simulated homophilous random networks. This is in line with the findings in Section 4.4. If homophily is highest, the probability that two agents are linked, given they have a common neighbor, is about four times as high as in the case of the Bernoulli Random Graphs where this probability coincides with the unconditional linking probability $\Phi(\lambda, a)$. Another expectable, yet important observation is that there are no significant differences between the expected clustering coefficient C (recall Definition 4.1) and the values we determined for the statistical coefficients $C^{(1)}$ and $C^{(2)}$ (recall Definition 4.3).⁸⁹ To sum up, Simulation 4.1 indicates that the Homophilous Random

⁸⁸To calculate average shortest paths, one commonly restricts to agents having a connecting path if the network has more than one component. However, such a network realized extremely rarely in this simulation, namely only in 0.06% of all cases.

⁸⁹Note that isolated and loose-end agents never appeared in the simulation guaranteeing that $C^{(2)}$

Network model exhibits the small-world phenomenon and clustering at the same time for all $a \in (0, 1)$. In what follows, we consider the most interesting case of highest homophily captured by $a = 10^{-8}$ in more detail.

The second simulation focuses on the formal Definition 4.2 of the small-world phenomenon. For this purpose, we simulate a collection of $R = 100$ networks for each size $n = 150, 200, 250, \dots, 1000$ and compute the respective averages of the relevant network statistics. To this end, we consider the parameter of highest homophily that is regarded in Simulation 4.1. The precise data is stated in Table 4.2. Note that the simulation for each network size is structurally the same as in the first simulation, merely a smaller number of iterations is chosen due to computational restrictions. However, as can be seen in Table 4.1, all standard errors and especially the one of the ratio $\bar{d}/\ln(n)$ are very low. Thus, 100 iterations should be sufficient to generate a precise estimate.

In Figure 4.4, we plot the ratio of the average shortest path length and the logarithm of the network size $\bar{d}/\ln(n)$ for the different network sizes n . This ratio is decreasing in n as the illustration reveals. From this, we deduce that the average path length \bar{d} increases slower in n than $\ln(n)$ does. Thus, the homophilous random networks exhibit the small-world phenomenon according to Definition 4.2.

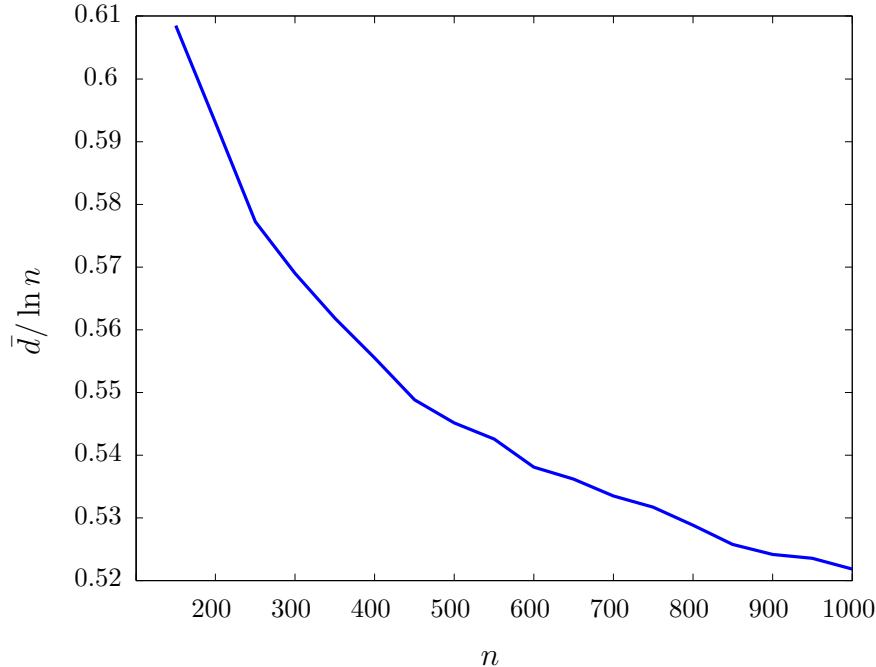


Figure 4.4: Small World of HRN with n from 150 to 1000 and constant expected degree 15 (created with MATLAB, 2014)

was steadily well-defined.

Parameter / Statistics	$n = 150$	200	250	300	350	400
R			100			
a			10^{-8}			
Expected Degree η^{exp}			15			
Average Degree $\bar{\eta}$	14,99	15,02	14,98	15,02	14,97	15,00
Average Shortest Path \bar{d}	3,05	3,14	3,19	3,25	3,29	3,33
$\bar{d}/\ln(n)$	0,609	0,593	0,577	0,569	0,562	0,556

Parameter / Statistics	$n = 450$	500	550	600	650	700
R			100			
a			10^{-8}			
Expected Degree η^{exp}			15			
Average Degree $\bar{\eta}$	15,01	15,03	15,02	15,01	15,00	15,01
Average Shortest Path \bar{d}	3,35	3,39	3,42	3,44	3,47	3,50
$\bar{d}/\ln(n)$	0,549	0,545	0,543	0,538	0,536	0,534

Parameter / Statistics	$n = 750$	800	850	900	950	1000
R			100			
a			10^{-8}			
Expected Degree η^{exp}			15			
Average Degree $\bar{\eta}$	14,99	14,98	15,03	15,04	14,97	15,01
Average Shortest Path \bar{d}	3,52	3,54	3,55	3,57	3,59	3,61
$\bar{d}/\ln(n)$	0,532	0,529	0,526	0,524	0,524	0,522

Table 4.2: Results of Simulation 4.2 computing average degrees, shortest paths and small world ratios of the HRN model for a growing network size (carried out with MATLAB, 2014)

4.6 An Example of the Labor Market

While in the previous sections, a theoretical analysis of the suggested Homophilous Random Network model is presented, we now provide one possible economic application. In recent years, more and more research in the field of labor economics has been dedicated to understanding the mechanisms of different hiring channels. One of these channels which is commonly used in reality relies on the contacts of current employees. Starting with the seminal contribution of Montgomery (1991), a lot of researchers decided to model connections between workers as a social network (see e.g. Calvó-Armengol, 2004; Calvó-Armengol and Jackson, 2007; Dawid and Gemkow,

2014).⁹⁰ As known from the extensive sociological literature (see Section 4.1), in these social networks, one should expect to observe homophily with respect to skills or competence, performance, education, level of income, and geographical distance. While there are lots of empirical studies confirming the existence of homophily in workers' social contacts and analyzing the implications thereof (see e.g. Mayer and Puller, 2008; Rees, 1966), only few work has yet been dedicated to developing theoretical models capturing this effect.⁹¹

In our application, we consider a risk-neutral firm that plans to fill an open vacancy. Two possible hiring channels are available. On the one hand, there is the formal job market and, on the other hand, the possibility to hire a contact of its current employee. Based on the model introduced in Section 4.2, we consider n workers and a vector of characteristics p capturing the ability of each worker to do the vacant job. W.l.o.g. we assume that agent 1 is the current employee of the firm while all other agents $2, \dots, n$ are supposed to be available on the job market. While we fix p_1 as a parameter of the model, meaning that the firm knows the ability of its current employee, $p_{-1} = (p_2, \dots, p_n)$ is again considered as a realization of the $(n-1)$ -dimensional random variable P_{-1} . Given this situation and based on individual linking probabilities (4.1) for parameters $\lambda, a \in (0, 1)$, we assume that a homophilous random network forms.

Knowing the distribution function of the random variable P_{-1} and the conditional linking probabilities but not the realization, the firm has to decide on one hiring channel. For this purpose, the expected characteristic of a contact of agent 1 is the crucial statistic. It can be calculated as follows.⁹²

Proposition 4.4. *Given some homophily parameter $a \in (0, 1)$, the expected characteristic of a neighbor $j \in \{2, \dots, n\}$ of agent 1 with given characteristic $p_1 \in [0, 1]$ is*

$$\mathbb{E}^P [P_j \mid G \in \mathbb{G}_{1j}] = \frac{1}{2} + \frac{(a^{p_1} - a^{1-p_1})\left(\frac{1}{2} - \frac{1}{\ln(a)}\right) + 2p_1 - 1}{2 - a^{p_1} - a^{1-p_1}}. \quad (4.9)$$

A plot of function (4.9) is given in Figure 4.5. However, investigating the expected characteristic analytically, reveals some intuitive properties, at least for some special

⁹⁰For an extensive survey including both empirical and theoretic literature from sociology and economics see Ioannides and Loury (2004).

⁹¹Exceptions are Horváth (2013), van der Leij and Buhai (2008) and Zaharieva (2013), however, all using binary notions of homophily.

⁹²Note that this probability is meaningful only if agent 1 has at least one link. For large networks, however, this is guaranteed whenever the corresponding condition of the threshold theorem (recall Proposition 4.3) is fulfilled.

cases. These might contribute to a better understanding of the rather complicated functional form and its appearance. We collect these properties in the following corollary. Note that all of them can be detected in Figure 4.5.

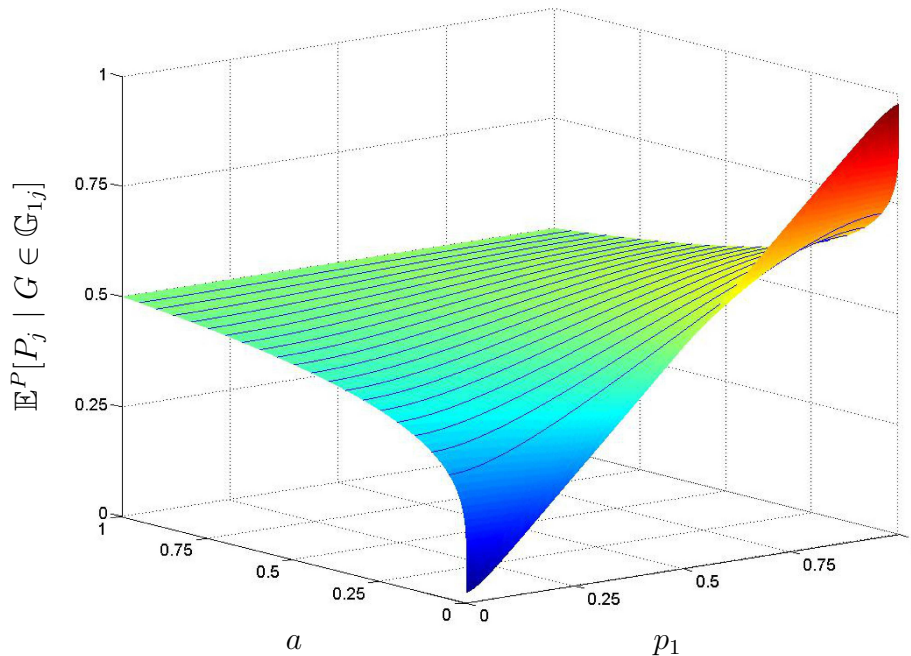


Figure 4.5: Expected characteristic of a contact of agent 1 (carried out and created with MATLAB, 2014)

Corollary 4.5. *Function (4.9) in Proposition 4.4 yields:*

- (i) $\mathbb{E}^P[P_j | G \in \mathbb{G}_{1j}] \Big|_{p_1=\frac{1}{2}} = \frac{1}{2} \quad \forall a \in (0, 1),$
- (ii) $\lim_{a \rightarrow 0} \mathbb{E}^P[P_j | G \in \mathbb{G}_{1j}] = p_1 \quad \forall p_1 \in [0, 1],$ and
- (iii) $\lim_{a \rightarrow 1} \mathbb{E}^P[P_j | G \in \mathbb{G}_{1j}] = \frac{1}{2} \quad \forall p_1 \in [0, 1].$

Finally assume, for simplicity, that the expected characteristic or rather ability of a worker hired via the formal job market is some value $\bar{p} \in (0, 1)$ which is independent of the homophily parameter a and the ability of the current employee p_1 . Given this situation, the firm faces a simple decision rule when to hire via the social network. We have that, for sufficiently high p_1 and low a , respectively, the expected ability of the current employee's contact exceeds any ability level \bar{p} . More precisely, for any parameter value $a \in (0, 1)$, solving the equation $\mathbb{E}^P[P_j | G \in \mathbb{G}_{1j}] = \bar{p}$ yields a minimum ability level p_1 (if existing at this homophily level) that has to be reached for the expected ability of the current employee's contact to exceed \bar{p} . Similarly, given

$p_1 \in [0, 1]$, we obtain a maximum level of a , that is a minimum level of homophily. Thus, the decision rule is that the firm should hire a randomly chosen contact instead of recruiting via the formal job market if and only if the respective calculated minimum level is exceeded (or at least reached).

4.7 Conclusion

In this chapter, we set up a novel Homophilous Random Network model incorporating heterogeneity of agents. In a two-stage random process, first each agent (or vertex) is assigned a one-dimensional characteristic. Second, based on these realized characteristics, the links of a random network form whilst taking into account a continuous notion of homophily. This captures the frequently observed propensity of individuals to connect with similar others. Exploiting this continuous formalization of homophily, our approach allows for a broad range of homophily levels ranging from the extreme case of maximal homophily where only equal agents get linked with positive probability up to the case where there is no homophily at all. The latter case corresponds to the Bernoulli Random Graph (BRG) model, often referred to as the Erdős-Rényi model. Insofar, our model can also be regarded as a generalization thereof. Most importantly, unlike the vast majority of related economic models, we indeed capture homophily as it is defined and used in the sociological literature, namely in terms of similarity rather than equality.

In our work, we first reveal some basic properties and network statistics of the Homophilous Random Network model and establish a threshold theorem. The comparison with the BRG model provides additional insight. To derive one of our main results, we focus on another stylized fact of real-world networks, namely the occurrence of clustering. Although homophily and clustering are frequently observed in reality, both phenomena are not captured by the original BRG model. While revealing by simulations that the small-world phenomenon is apparently preserved, we are able to show analytically that homophily induces clustering in our model. This gives rise to the conjecture that also in reality there might be a considerable causality between the two. It might be worthwhile for future research to pursue this question. Finally, we provide an easily accessible application of our model for labor economics. Assuming homophily with respect to abilities to do a certain job, we consider workers being connected through a homophilous random network. We determine the expected ability of a given worker's random contact depending on the level of homophily and the given worker's own ability. This yields a simple decision rule for a firm which

intends to fill an open vacancy and needs to decide whether to hire through a current employee's contacts or the formal job market.

Furthermore, our Homophilous Random Network model is now available as a tool which can be used to understand and predict diffusion processes in social networks. As it complies with those important stylized facts which we frequently observe in social networks, it might yield meaningful results, for instance, regarding the spread of information or a disease. Beyond that, there are certainly several further questions which remain open for future research. Although our simulation results yield a strong indication in this direction, one task would be to show analytically that the small-world phenomenon is generally preserved in our model. As a second point, it could be of interest to expand our considerations about threshold theorems and to establish those for different properties such as connectedness in our model. Further, it would be a natural, yet analytically challenging extension to check the qualitative robustness of the findings for different distributions of characteristics. For many applications, a distribution that puts more weight on intermediate characteristics might capture reality more accurately. Also, an extension of the model to multi-dimensional characteristics would be valuable, in particular if one would succeed to combine characteristics of both continuous and binary nature. Finally, a calibration of the model to real-world data is yet to be done. Performing this in a meaningful way is most certainly a challenge, especially as the level of homophily in a given network is not clearly observable. However, one way to deal with this could be to calibrate the model to the observable degree of clustering which we showed to be directly connected to homophily in our model.

Appendix 4.A Proofs

4.A.1 Proof of Proposition 4.1

We calculate the expected probability:

$$\begin{aligned}
 \mathbb{E}^P \left[\mathbb{P}^G (ij \in G \mid P) \mid P_i = p_i \right] &= \mathbb{E}^P \left[\lambda a^{|P_i - P_j|} \mid P_i = p_i \right] \\
 &= \lambda \left(\int_0^1 \underbrace{f_{P_j}(p_j)}_1 a^{|p_i - p_j|} dp_j \right) \\
 &= \lambda \left(\int_0^{p_i} a^{p_i - p_j} dp_j + \int_{p_i}^1 a^{p_j - p_i} dp_j \right) \\
 &= \lambda \left(a^{p_i} \int_0^{p_i} a^{-p_j} dp_j + a^{-p_i} \int_{p_i}^1 a^{p_j} dp_j \right)
 \end{aligned}$$

$$\begin{aligned}
&= \lambda \left(a^{p_i} \frac{1 - a^{-p_i}}{\ln(a)} + a^{-p_i} \frac{a - a^{p_i}}{\ln(a)} \right) \\
&= \frac{\lambda}{\ln(a)} (a^{p_i} + a^{1-p_i} - 2). \tag{4.10}
\end{aligned}$$

Moreover, by integrating equation (4.10) with respect to p_i , we get the expected probability if p is unknown:

$$\begin{aligned}
\mathbb{E}^P [\mathbb{P}^G [ij \in G \mid P]] &= \mathbb{E}^P [\lambda a^{|P_i - P_j|}] \\
&= \lambda \left(\int_{[0,1]^2} \underbrace{f_{P_i, P_j}(p_i, p_j)}_{= f_{P_i}(p_i) f_{P_j}(p_j) = 1} a^{|p_i - p_j|} d(p_i, p_j) \right) \\
&\stackrel{(4.10)}{=} \lambda \left(\int_0^1 \frac{(a^{p_i} + a^{1-p_i} - 2)}{\ln(a)} dp_i \right) \\
&= \frac{\lambda}{\ln(a)} \left[\frac{a^{p_i} - a^{1-p_i} - 2p_i \ln(a)}{\ln(a)} \right] \Bigg|_{p_i=0}^{p_i=1} \\
&= \frac{\lambda}{\ln(a)^2} [a - 1 - 2 \ln(a) - 1 + a] \\
&= \frac{2\lambda}{\ln(a)^2} [a - 1 - \ln(a)].
\end{aligned}$$

□

4.A.2 Proof of Corollary 4.1

Using l'Hôpital's rule, we calculate the limit of φ as

$$\begin{aligned}
\lim_{a \rightarrow 0} \varphi(\lambda, a, p_i) &= \lim_{a \rightarrow 0} \frac{\lambda(a^{p_i} + a^{1-p_i} - 2)}{\ln(a)} = \lim_{a \rightarrow 0} \frac{\lambda(p_i a^{p_i-1} + (1-p_i)a^{-p_i})}{1/a} \\
&= \lim_{a \rightarrow 0} \lambda(p_i a^{p_i} + (1-p_i)a^{1-p_i}) = 0.
\end{aligned}$$

Similarly, we get

$$\begin{aligned}
\lim_{a \rightarrow 1} \varphi(\lambda, a, p_i) &= \lim_{a \rightarrow 1} \frac{\lambda(a^{p_i} + a^{1-p_i} - 2)}{\ln(a)} = \lim_{a \rightarrow 1} \frac{\lambda(p_i a^{p_i-1} + (1-p_i)a^{-p_i})}{1/a} \\
&= \lim_{a \rightarrow 1} \lambda(p_i a^{p_i} + (1-p_i)a^{1-p_i}) = \lambda.
\end{aligned}$$

For the case of Φ , by now using l'Hôpital's rule twice, we get

$$\lim_{a \rightarrow 0} \Phi(\lambda, a) = \lim_{a \rightarrow 0} 2\lambda \frac{a - 1 - \ln(a)}{\ln(a)^2} = \lim_{a \rightarrow 0} 2\lambda \frac{1 - 1/a}{2 \ln(a)/a} = \lim_{a \rightarrow 0} \lambda \frac{a - 1}{\ln(a)} = 0,$$

as well as

$$\lim_{a \rightarrow 1} \Phi(\lambda, a) = \lim_{a \rightarrow 1} 2\lambda \frac{a - 1 - \ln(a)}{\ln(a)^2} = \lim_{a \rightarrow 1} 2\lambda \frac{a - 1}{2 \ln(a)} = \lim_{a \rightarrow 1} \lambda \frac{1}{1/a} = \lambda.$$

□

4.A.3 Proof of Proposition 4.2

Taking into account equation (4.2), we calculate

$$\begin{aligned} & \mathbb{E}^P \left[\mathbb{P}^G (\eta_i(G) = k \mid P) \mid P_i = p_i \right] \\ &= \mathbb{E}^P \left[\sum_{K \subseteq N \setminus \{i\}; |K|=k} \left(\prod_{j \in K} (q(P_i, P_j)) \cdot \prod_{l \in N \setminus K \setminus \{i\}} (1 - q(P_i, P_l)) \right) \mid P_i = p_i \right] \\ &= \sum_{K \subseteq N \setminus \{i\}; |K|=k} \left(\mathbb{E}^P \left[\prod_{j \in K} (q(P_i, P_j)) \cdot \prod_{l \in N \setminus K \setminus \{i\}} (1 - q(P_i, P_l)) \mid P_i = p_i \right] \right) \\ &= \sum_{K \subseteq N \setminus \{i\}; |K|=k} \left(\int_{[0,1]^{n-1}} \left(\underbrace{f_{P_i}(p_{-i})}_{=1} \cdot \prod_{j \in K} (q(p_i, p_j)) \cdot \prod_{l \in N \setminus K \setminus \{i\}} (1 - q(p_i, p_l)) \right) dp_{-i} \right) \\ &= \sum_{K \subseteq N \setminus \{i\}; |K|=k} \left(\prod_{j \in K} \left(\int_0^1 (q(p_i, p_j)) dp_j \right) \cdot \prod_{l \in N \setminus K \setminus \{i\}} \left(\int_0^1 (1 - q(p_i, p_l)) dp_l \right) \right) \\ &\stackrel{(4.2)}{=} \sum_{K \subseteq N \setminus \{i\}; |K|=k} \left(\left(\frac{\lambda}{\ln(a)} (a^{p_i} + a^{1-p_i} - 2) \right)^k \cdot \left(1 - \frac{\lambda}{\ln(a)} (a^{p_i} + a^{1-p_i} - 2) \right)^{n-k-1} \right) \\ &\stackrel{(4.2)}{=} \binom{n-1}{k} \cdot (\varphi(\lambda, a, p_i))^k \cdot (1 - \varphi(\lambda, a, p_i))^{n-k-1}. \end{aligned}$$

□

4.A.4 Proof of Proposition 4.3

The probability that an agent i with given characteristic p_i is isolated is

$$\mathbb{E}^P \left[\mathbb{P}^G (\eta_i(G) = 0 \mid P) \mid P_i = p_i \right] \stackrel{(4.8)}{=} (1 - \varphi(\lambda(n), a(n), p_i))^{n-1}.$$

If we assume that there is at least some homophily as the size of the network becomes large, that is formally

$$\exists \tilde{\epsilon} > 0, \bar{n} \in \mathbb{N} : a(n) \leq 1 - \tilde{\epsilon} \quad \forall n \geq \bar{n},$$

then we have that

$$\exists \epsilon > 0 : 2 - a(n)^{\hat{p}} - a(n)^{1-\hat{p}} \in [\epsilon, 2] \quad \forall n \geq \bar{n}.$$

Now it holds that if $\lim_{n \rightarrow \infty} [-\lambda(n)/(\ln(a(n))t(n))] = \infty$, then we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} (1 - \varphi(\lambda(n), a(n), p_i))^{n-1} \\ &= \lim_{n \rightarrow \infty} \left(1 - \frac{\varphi(\lambda(n), a(n), p_i)/t(n)}{n-1} \right)^{n-1} \\ &\stackrel{(4.2)}{=} \lim_{n \rightarrow \infty} \left(1 - \frac{\frac{\lambda(n)(n-1)}{\ln(a(n))} (a(n)^{p_i} + a(n)^{1-p_i} - 2)}{n-1} \right)^{n-1} \\ &= \lim_{n \rightarrow \infty} \exp \left(\underbrace{-\frac{\lambda(n)(n-1)}{\ln(a(n))}}_{\rightarrow \infty} \underbrace{(a(n)^{p_i} + a(n)^{1-p_i} - 2)}_{\in [-2, -\epsilon]} \right) = 0, \end{aligned}$$

On the contrary, if $\lim_{n \rightarrow \infty} [-\lambda(n)/(\ln(a(n))t(n))] = 0$, then we get

$$\begin{aligned} & \lim_{n \rightarrow \infty} (1 - \varphi(\lambda(n), a(n), p_i))^{n-1} \\ &= \lim_{n \rightarrow \infty} \exp \left(\underbrace{-\frac{\lambda(n)(n-1)}{\ln(a(n))}}_{\rightarrow 0} \underbrace{(a(n)^{p_i} + a(n)^{1-p_i} - 2)}_{\in [-2, -\epsilon]} \right) = 1. \end{aligned}$$

□

4.A.5 Proof of Theorem 4.1

We calculate the clustering coefficient

$$\begin{aligned} & C(\lambda, a) \\ &= \mathbb{E}^P \left[\lambda a^{|P_j - P_k|} \mid G \in \mathbb{G}_{ij} \cap \mathbb{G}_{ik} \right] \\ &= \lambda \int_{[0,1]^n} a^{|p_j - p_k|} f_P(p \mid G \in \mathbb{G}_{ij} \cap \mathbb{G}_{ik}) dp \end{aligned}$$

$$\begin{aligned}
&= \lambda \int_{[0,1]^n} a^{|p_j-p_k|} \frac{f_{P,G}(p, \mathbb{G}_{ij} \cap \mathbb{G}_{ik})}{f_G(\mathbb{G}_{ij} \cap \mathbb{G}_{ik})} dp \\
&= \frac{\lambda}{f_G(\mathbb{G}_{ij} \cap \mathbb{G}_{ik})} \int_{[0,1]^n} a^{|p_j-p_k|} f_{P,G}(p, \mathbb{G}_{ij} \cap \mathbb{G}_{ik}) dp \\
&= \frac{\lambda}{f_G(\mathbb{G}_{ij} \cap \mathbb{G}_{ik})} \int_{[0,1]^n} a^{|p_j-p_k|} f_G(\mathbb{G}_{ij} \cap \mathbb{G}_{ik} \mid P=p) \overbrace{f_P(p)}^{=1} dp \\
&= \frac{\lambda}{\int_{[0,1]^n} \underbrace{f_P(x)}_{=1} f_G(\mathbb{G}_{ij} \cap \mathbb{G}_{ik} \mid P=x) dx} \int_{[0,1]^n} a^{|p_j-p_k|} f_G(\mathbb{G}_{ij} \cap \mathbb{G}_{ik} \mid P=p) dp \\
&= \frac{\lambda}{\int_{[0,1]^n} \mathbb{P}^G(G \in \mathbb{G}_{ij} \cap \mathbb{G}_{ik} \mid P=x) dx} \int_{[0,1]^n} a^{|p_j-p_k|} \mathbb{P}^G(G \in \mathbb{G}_{ij} \cap \mathbb{G}_{ik} \mid P=p) dp \\
&= \frac{\lambda}{\int_{[0,1]^n} \lambda a^{|x_i-x_j|} \lambda a^{|x_i-x_k|} dx} \int_{[0,1]^n} a^{|p_j-p_k|} \lambda a^{|p_i-p_j|} \lambda a^{|p_i-p_k|} dp \\
&= \lambda \frac{\int_{[0,1]^n} a^{|p_j-p_k|+|p_i-p_j|+|p_i-p_k|} dp}{\int_{[0,1]^n} a^{|x_i-x_j|+|x_i-x_k|} dx} = \lambda \frac{\int_{[0,1]^3} a^{|p_j-p_k|+|p_i-p_j|+|p_i-p_k|} d(p_i, p_j, p_k)}{\int_{[0,1]^3} a^{|x_i-x_j|+|x_i-x_k|} d(x_i, x_j, x_k)}.
\end{aligned} \tag{4.11}$$

Let us solve the integral in the denominator first. For the sake of readability denote $x = (x_i, x_j, x_k)$. We have

$$\begin{aligned}
\int_{[0,1]^3} a^{|x_i-x_j|+|x_i-x_k|} dx &= \int_{\substack{x \in [0,1]^3: \\ x_j, x_k \leq x_i}} a^{2x_i-x_j-x_k} dx + \int_{\substack{x \in [0,1]^3: \\ x_i \leq x_j, x_k}} a^{x_j+x_k-2x_i} dx \\
&\quad + \int_{\substack{x \in [0,1]^3: \\ x_j \leq x_i \leq x_k}} a^{x_k-x_j} dx + \int_{\substack{x \in [0,1]^3: \\ x_k \leq x_i \leq x_j}} a^{x_j-x_k} dx \\
&= \frac{2 \ln(a) - 4a + a^2 + 3}{2(\ln(a))^3} + \frac{2 \ln(a) - 4a + a^2 + 3}{2(\ln(a))^3} \\
&\quad + \frac{2 \ln(a) - 4a + 2a \ln(a) + 4}{2(\ln(a))^3} + \frac{2 \ln(a) - 4a + 2a \ln(a) + 4}{2(\ln(a))^3} \\
&= \frac{1}{2(\ln(a))^3} [8 \ln(a) - 16a + 2a^2 + 4 \ln(a)a + 14].
\end{aligned}$$

Next, we solve the integral in the numerator of (4.11), substituting x for p in order to use the same notation as above. This yields

$$\begin{aligned}
&\int_{[0,1]^3} a^{|x_j-x_k|+|x_i-x_j|+|x_i-x_k|} dx \\
&= \int_{\substack{x \in [0,1]^3: \\ x_i \leq x_j \leq x_k}} a^{2x_k-2x_i} dx + \int_{\substack{x \in [0,1]^3: \\ x_i \leq x_k \leq x_j}} a^{2x_j-2x_i} dx + \int_{\substack{x \in [0,1]^3: \\ x_j \leq x_i \leq x_k}} a^{2x_k-2x_j} dx \\
&\quad + \int_{\substack{x \in [0,1]^3: \\ x_j \leq x_k \leq x_i}} a^{2x_i-2x_j} dx + \int_{\substack{x \in [0,1]^3: \\ x_k \leq x_i \leq x_j}} a^{2x_j-2x_k} dx + \int_{\substack{x \in [0,1]^3: \\ x_k \leq x_j \leq x_i}} a^{2x_i-2x_k} dx
\end{aligned}$$

$$= 6 \frac{\ln(a) - a^2 + a^2 \ln(a) + 1}{4(\ln(a))^3} = \frac{1}{2(\ln(a))^3} [3 \ln(a) - 3a^2 + 3a^2 \ln(a) + 3].$$

Taken together, this gives

$$C(\lambda, a) = \lambda \frac{3 \ln(a) - 3a^2 + 3a^2 \ln(a) + 3}{8 \ln(a) - 16a + 2a^2 + 4 \ln(a)a + 14}.$$

By using this, we can now start with the actual proof. We have

$$\begin{aligned} & C(\lambda, a) - \Phi(\lambda, a) \\ &= \lambda \left(\frac{3(\ln(a)a^2 + \ln(a) - a^2 + 1)}{2(2 \ln(a)a + 4 \ln(a) + a^2 - 8a + 7)} + \frac{2(\ln(a) - a + 1)}{\ln(a)^2} \right) \\ &= \lambda \frac{3 \ln(a)^3(a^2+1) + \ln(a)^2(-3a^2+8a+19) + \ln(a)(-4a^2-40a+44) + (-4a^3+36a^2-60a+28)}{2 \ln(a)^2(2 \ln(a)a + 4 \ln(a) + a^2 - 8a + 7)} \end{aligned} \quad (4.12)$$

In what follows, we use that for $a \in (0, 1)$ we have

$$\ln(a) = - \sum_{m=0}^{\infty} \frac{(1-a)^{m+1}}{m+1}$$

which implies that $\ln(a) < - \sum_{m=0}^M \frac{(1-a)^{m+1}}{m+1} < 0$ for all $M \in \mathbb{N}$. The first and easier part is to show that the denominator of the term on the right-hand side of equation (4.12) is negative for all $a \in (0, 1)$. We calculate

$$\begin{aligned} & 2 \ln(a)a + 4 \ln(a) + a^2 - 8a + 7 \\ &= 2(a+2) \ln(a) + a^2 - 8a + 7 \\ &< -2(a+2) \left(1-a + \frac{1}{2}(1-a)^2 + \frac{1}{3}(1-a)^3 \right) + a^2 - 8a + 7 \\ &= \frac{1}{3}(a+2)(2a^3 - 9a^2 + 18a - 11) + a^2 - 8a + 7 \\ &= \frac{1}{3}(2a^4 - 5a^3 + 3a^2 + a - 1) = -\frac{1}{3}(1-a)^3(2a+1) < 0 \end{aligned}$$

Further, we define

$$\begin{aligned} g(a) &:= 3 \ln(a)^3(a^2 + 1) + \ln(a)^2(-3a^2 + 8a + 19) + \ln(a)(-4a^2 - 40a + 44) \\ &\quad + (-4a^3 + 36a^2 - 60a + 28). \end{aligned}$$

Then $\lambda g(a)$ is the numerator of the term on the right-hand side of equation (4.12). We calculate the derivatives

$$\begin{aligned}\frac{dg}{da}(a) &= \frac{1}{a} [6 \ln(a)^3 a^2 + \ln(a)^2 (3a^2 + 8a + 9) + 2 \ln(a) (-7a^2 - 12a + 19) \\ &\quad + 4(-3a^3 + 17a^2 - 25a + 11)], \\ \frac{d^2g}{da^2}(a) &= \frac{1}{a^2} [6 \ln(a)^3 a^2 + 3 \ln(a)^2 (7a^2 - 3) + 4 \ln(a) (-2a^2 + 4a - 5) \\ &\quad + 6(-4a^3 + 9a^2 - 4a - 1)], \\ \frac{d^3g}{da^3}(a) &= \frac{1}{a^3} [18 \ln(a)^2 (a^2 + 1) + 2 \ln(a) (21a^2 - 8a + 11) + 8(-3a^3 - a^2 + 5a - 1)], \\ \frac{d^4g}{da^4}(a) &= \frac{1}{a^4} [18 \ln(a)^2 (-a^2 - 3) + 2 \ln(a) (-3a^2 + 16a - 15) + 2(25a^2 - 48a + 23)], \\ \frac{d^5g}{da^5}(a) &= \frac{1}{a^5} [36 \ln(a)^2 (a^2 + 6) + 12 \ln(a) (-2a^2 - 8a + 1) + 2(-53a^2 + 160a - 107)], \\ \frac{d^6g}{da^6}(a) &= \frac{1}{a^6} [108 \ln(a)^2 (-a^2 - 10) + 12 \ln(a) (12a^2 + 32a + 31) + 2(147a^2 - 688a + 541)].\end{aligned}$$

Notice here that

$$g(1) = \frac{dg}{da}(1) = \frac{d^2g}{da^2}(1) = \frac{d^3g}{da^3}(1) = \frac{d^4g}{da^4}(1) = \frac{d^5g}{da^5}(1) = 0.$$

Moreover, we have

$$\begin{aligned}\frac{d^6g}{da^6}(a) &= \frac{1}{a^6} \left[108 \ln(a)^2 \underbrace{(-a^2 - 10)}_{<0} + 12 \ln(a) \underbrace{(12a^2 + 32a + 31)}_{>0} \right. \\ &\quad \left. + 2(147a^2 - 688a + 541) \right] \\ &< \frac{1}{a^6} \left[108(1-a)^2(-a^2 - 10) - 12(1-a)(12a^2 + 32a + 31) \right. \\ &\quad \left. + 2(147a^2 - 688a + 541) \right] \\ &= \frac{2}{a^6} \left[-54a^4 + 180a^3 - 327a^2 + 386a - 185 \right] \\ &= \frac{2}{a^6} (1-a) \left[54(a - \frac{7}{9})^3 + 103(a - \frac{7}{9}) - \frac{2146}{27} \right] \\ &< \frac{2}{a^6} (1-a) \left[54 \cdot (\frac{2}{9})^3 + 103 \cdot \frac{2}{9} - \frac{2146}{27} \right] = -\frac{112}{a^6} (1-a) < 0.\end{aligned}$$

Combining this, it follows for all $a \in (0, 1)$ that

$$\begin{aligned}\frac{d^5g}{da^5}(a) > 0 &\Rightarrow \frac{d^4g}{da^4}(a) < 0 \Rightarrow \frac{d^3g}{da^3}(a) > 0 \Rightarrow \frac{d^2g}{da^2}(a) < 0 \Rightarrow \frac{dg}{da}(a) > 0 \\ &\Rightarrow g(a) < 0.\end{aligned}$$

Taken together, we have indeed that

$$C(\lambda, a) - \Phi(\lambda, a) = \lambda \frac{g(a)}{2 \ln(a)^2 (2 \ln(a)a + 4 \ln(a) + a^2 - 8a + 7)} > 0$$

which concludes the proof of the theorem. \square

4.A.6 Proof of Corollary 4.4

By applying l'Hôpital's rule three times, we calculate

$$\begin{aligned}
\lim_{a \rightarrow 0} C(\lambda, a) &= \lambda \lim_{a \rightarrow 0} \frac{3 \ln(a) - 3a^2 + 3a^2 \ln(a) + 3}{8 \ln(a) - 16a + 2a^2 + 4 \ln(a)a + 14} \\
&= \lambda \lim_{a \rightarrow 0} \frac{3/a - 6a + 6a \ln(a) + 3a}{8/a - 16 + 4a + 4 \ln(a) + 4} \\
&= \frac{3\lambda}{4} \lim_{a \rightarrow 0} \frac{1 - a^2 + 2a^2 \ln(a)}{2 - 3a + a^2 + a \ln(a)} \\
&= \frac{3\lambda}{4} \frac{\lim_{a \rightarrow 0}[1 - a^2 + 2a^2 \ln(a)]}{\lim_{a \rightarrow 0}[2 - 3a + a^2 + a \ln(a)]} \\
&= \frac{3\lambda}{4} \frac{\lim_{a \rightarrow 0}[1] - \lim_{a \rightarrow 0}[a^2] + \lim_{a \rightarrow 0}[2a^2 \ln(a)]}{\lim_{a \rightarrow 0}[2] - \lim_{a \rightarrow 0}[3a] + \lim_{a \rightarrow 0}[a^2] + \lim_{a \rightarrow 0}[a \ln(a)]} \\
&= \frac{3\lambda}{4} \frac{1 - 0 + \lim_{x \rightarrow \infty}[2 \ln(1/x)/x^2]}{2 - 0 + 0 + \lim_{x \rightarrow \infty}[\ln(1/x)/x]} \\
&= \frac{3\lambda}{4} \frac{1 + \lim_{x \rightarrow \infty}[-2x(1/x^2)/2x]}{2 + \lim_{x \rightarrow \infty}[-x(1/x^2)/1]} = \frac{3\lambda}{4} \frac{1 + \lim_{x \rightarrow \infty}[-1/x^2]}{2 + \lim_{x \rightarrow \infty}[-1/x]} = \frac{3\lambda}{8}.
\end{aligned}$$

The stated result follows immediately since we established in Corollary 4.1 that $\lim_{a \rightarrow 0} \Phi(\lambda, a) = 0$. On the contrary, by again using l'Hôpital's rule three times, we get

$$\begin{aligned}
\lim_{a \rightarrow 1} C(\lambda, a) &= \lambda \lim_{a \rightarrow 1} \frac{3 \ln(a) - 3a^2 + 3a^2 \ln(a) + 3}{8 \ln(a) - 16a + 2a^2 + 4 \ln(a)a + 14} \\
&= \lambda \lim_{a \rightarrow 1} \frac{3/a - 6a + 6a \ln(a) + 3a}{8/a - 16 + 4a + 4 \ln(a) + 4} \\
&= \lambda \lim_{a \rightarrow 1} \frac{3 - 3a^2 + 6a^2 \ln(a)}{8 - 12a + 4a^2 + 4a \ln(a)} \\
&= \lambda \lim_{a \rightarrow 1} \frac{-6a + 12a \ln(a) + 6a}{-12 + 8a + 4 \ln(a) + 4} = \lambda \lim_{a \rightarrow 1} \frac{12 \ln(a) + 12}{8 + 4/a} = \lambda.
\end{aligned}$$

According to Corollary 4.1, we have $\lim_{a \rightarrow 1} \Phi(\lambda, a) = \lambda$ which concludes the proof. \square

4.A.7 Proof of Proposition 4.4

We calculate

$$\begin{aligned}
\mathbb{E}^P [P_j \mid G \in \mathbb{G}_{1j}] &= \int_0^1 p_j f_{P_j|G}(p_j, \mathbb{G}_{1j}) dp_j = \int_0^1 p_j f_{P_j}(p_j \mid G \in \mathbb{G}_{1j}) dp_j \\
&= \int_0^1 p_j \frac{f_{P_j, G}(p_j, \mathbb{G}_{1j})}{f_G(\mathbb{G}_{1j})} dp_j \\
&= \int_0^1 p_j \frac{f_G(\mathbb{G}_{1j} \mid P_j = p_j) \overbrace{f_{P_j}(p_j)}^1}{f_G(\mathbb{G}_{1j})} dp_j \\
&= \int_0^1 p_j \frac{f_G(\mathbb{G}_{1j} \mid P_j = p_j)}{\underbrace{\int_0^1 f_{P_j}(x) dx}_1 \underbrace{f_G(\mathbb{G}_{1j} \mid P_j = x)}_{\mathbb{P}(G \in \mathbb{G}_{1j} \mid P_j = x)}} dp_j \\
&= \int_0^1 p_j \frac{\overbrace{f_G(\mathbb{G}_{1j} \mid P_j = p_j)}^{\lambda a^{|p_1 - p_j|}}}{\underbrace{\int_0^1 \lambda a^{|p_1 - x|} dx}_{\frac{\lambda}{\ln(a)}(a^{p_1} + a^{1-p_1} - 2)}} dp_j \\
&= \frac{\ln(a)}{a^{p_1} + a^{1-p_1} - 2} \int_0^1 p_j a^{|p_1 - p_j|} dp_j.
\end{aligned}$$

Focusing on the integral first gives

$$\begin{aligned}
\int_0^1 p_j a^{|p_1 - p_j|} dp_j &= \int_0^{p_1} p_j a^{(p_1 - p_j)} dp_j + \int_{p_1}^1 p_j a^{(p_j - p_1)} dp_j \\
&= \frac{a^{p_1} - p_1 \ln(a) - 1}{\ln(a)^2} + \frac{a^{1-p_1} (\ln(a) - 1) - p_1 \ln(a) + 1}{\ln(a)^2}.
\end{aligned}$$

It follows that

$$\mathbb{E}^P(P_j \mid G \in \mathbb{G}_{1j}) = \frac{a^{p_1} + a^{1-p_1} (\ln(a) - 1) - 2p_1 \ln(a)}{\ln(a)(a^{p_1} + a^{1-p_1} - 2)} \quad (4.13a)$$

$$= \frac{1}{2} + \frac{(a^{p_1} - a^{1-p_1}) \left(\frac{1}{2} - \frac{1}{\ln(a)} \right) + 2p_1 - 1}{2 - a^{p_1} - a^{1-p_1}} \dots \quad (4.13b)$$

□

4.A.8 Proof of Corollary 4.5

Considering the functional form (4.9), we prove the properties in question one after the other. Regarding Part (i), by using equation (4.13b) we calculate for $a \in (0, 1)$

that

$$\mathbb{E}^P[P_j \mid G \in \mathbb{G}_{1j}] \Big|_{p_1=\frac{1}{2}} = \frac{1}{2} + \frac{(\sqrt{a} - \sqrt{a})(\frac{1}{2} - \frac{1}{\ln(a)}) + 1 - 1}{2 - \sqrt{a} - \sqrt{a}} = \frac{1}{2}.$$

Next, we consider Part (ii). Again applying equation (4.13b), we get for $p_1 \in (0, 1)$ that

$$\lim_{a \rightarrow 0} \mathbb{E}^P[P_j \mid G \in \mathbb{G}_{1j}] = \frac{1}{2} + \frac{(0 - 0)(\frac{1}{2} + 0) + 2p_1 - 1}{2 - 0 - 0} = p_1$$

and for the marginals we have

$$\begin{aligned} \lim_{a \rightarrow 0} \mathbb{E}^P[P_j \mid G \in \mathbb{G}_{1j}] \Big|_{p_1=0} &= \frac{1}{2} + \frac{(1 - 0)(\frac{1}{2} + 0) + 0 - 1}{2 - 1 - 0} = 0, \\ \lim_{a \rightarrow 0} \mathbb{E}^P[P_j \mid G \in \mathbb{G}_{1j}] \Big|_{p_1=1} &= \frac{1}{2} + \frac{(0 - 1)(\frac{1}{2} + 0) + 2 - 1}{2 - 0 - 1} = 1. \end{aligned}$$

To establish Part (iii), we have to apply l'Hôpital's rule. For $p_1 \in [0, 1]$ we get

$$\begin{aligned} \lim_{a \rightarrow 1} \mathbb{E}^P[P_j \mid G \in \mathbb{G}_{1j}] &\stackrel{(4.13a)}{=} \lim_{a \rightarrow 1} \frac{a^{p_1} + a^{1-p_1}(\ln(a) - 1) - 2p_1 \ln(a)}{\ln(a)(a^{p_1} + a^{1-p_1} - 2)} \\ &= \lim_{a \rightarrow 1} \frac{p_1 a^{p_1-1} + (1-p_1)a^{-p_1}(\ln(a) - 1) + a^{-p_1} - \frac{2p_1}{a}}{\frac{1}{a}(a^{p_1} + a^{1-p_1} - 2) + \ln(a)(p_1 a^{p_1-1} + (1-p_1)a^{-p_1})} \end{aligned} \quad (4.14)$$

while using l'Hôpital's rule once. However, we obviously need to apply it second time. For this purpose, we calculate the derivatives of the numerator and denominator of the term on the right-hand side in equation (4.14). We get

$$\begin{aligned} &\frac{\partial}{\partial a} \left[p_1 a^{p_1-1} + (1-p_1)a^{-p_1}(\ln(a) - 1) + a^{-p_1} - \frac{2p_1}{a} \right] \\ &= p_1(p_1 - 1)a^{p_1-2} + p_1(p_1 - 1)a^{-p_1-1}(\ln(a) - 1) + (1-p_1)a^{-p_1-1} - p_1 a^{-p_1-1} + \frac{2p_1}{a^2} \end{aligned}$$

and

$$\begin{aligned} &\frac{\partial}{\partial a} \left[\frac{1}{a}(a^{p_1} + a^{1-p_1} - 2) + \ln(a)(p_1 a^{p_1-1} + (1-p_1)a^{-p_1}) \right] \\ &= -\frac{1}{a^2}(a^{p_1} + a^{1-p_1} - 2) + \frac{2}{a}(p_1 a^{p_1-1} + (1-p_1)a^{-p_1}) \\ &\quad + \ln(a)(p_1(p_1 - 1)a^{p_1-2} + p_1(p_1 - 1)a^{-p_1-1}). \end{aligned}$$

By recalling equation (4.14) and using l'Hôpital's rule the second time, this gives

$$\lim_{a \rightarrow 1} \mathbb{E}^P[P_j \mid G \in \mathbb{G}_{1j}] = \frac{p_1(p_1 - 1) + p_1(p_1 - 1)(0 - 1) + (1 - p_1) - p_1 + 2p_1}{-(1 + 1 - 2) + 2(p_1 + (1 - p_1)) + 0} = \frac{1}{2}$$

which concludes the proof. □

References

- Abreu, D. and Manea, M. (2012). Bargaining and efficiency in networks. *Journal of Economic Theory*, 147(1):43–70.
- Amir, R. and Lazzati, N. (2014). Endogenous information acquisition in Bayesian games with strategic complementarities. Working Paper.
- Aumann, R. J. and Maschler, M. (1972). Some thoughts on the minimax principle. *Management Science*, 18(5):54–63.
- Aumann, R. J. and Myerson, R. B. (1988). Endogenous formation of links between players and of coalitions: An application of the Shapley value. In Roth, A. E., editor, *The Shapley Value: Essays in Honor of Lloyd S. Shapley*, pages 175–194. Cambridge University Press, Cambridge.
- Baerveldt, C., Van Duijn, M. A., Vermeij, L., and Van Hemert, D. A. (2004). Ethnic boundaries and personal choice. Assessing the influence of individual inclinations to choose intra-ethnic relationships on pupils’ networks. *Social Networks*, 26(1):55–74.
- Baetz, O. (2015). Social activity and network formation. *Theoretical Economics*, 10(2):315–340.
- Bala, V. and Goyal, S. (2000). A noncooperative model of network formation. *Econometrica*, 68(5):1181–1229.
- Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.
- Bellhouse, D. (2007). The problem of Waldegrave. *Electronic Journal for History of Probability and Statistics*, 3(2):1–12.
- Berge, C. (1981). Some common properties for regularizable graphs, edge-critical graphs and b-graphs. In Saito, N. and Nishizeki, T., editors, *Graph Theory and Algorithms*, pages 108–123. Springer, Berlin Heidelberg.

- Bergemann, D., Shi, X., and Välimäki, J. (2009). Information acquisition in interdependent value auctions. *Journal of the European Economic Association*, 7(1):61–89.
- Bloch, F. and Jackson, M. O. (2006). Definitions of equilibrium in network formation games. *International Journal of Game Theory*, 34(3):305–318.
- Bloznelis, M. (2013). Degree and clustering coefficient in sparse random intersection graphs. *The Annals of Applied Probability*, 23(3):1254–1289.
- Bollobás, B. (1998). *Modern Graph Theory*. Springer, New York, NY.
- Bollobás, B. (2001). *Random Graphs*. Cambridge University Press, Cambridge, second edition.
- Boorman, S. A. (1975). A combinatorial optimization model for transmission of job information through contact networks. *Bell Journal of Economics*, 6(1):216–249.
- Bramoullé, Y., Currarini, S., Jackson, M. O. ., Pin, P., and Rogers, B. W. (2012). Homophily and long-run integration in social networks. *Journal of Economic Theory*, 147(5):1754–1786.
- Burt, R. S. (1991). Measuring age as a structural concept. *Social Networks*, 13(1):1–34.
- Calvó-Armengol, A. (2003). Stable and efficient bargaining networks. *Review of Economic Design*, 7(4):411–428.
- Calvó-Armengol, A. (2004). Job contact networks. *Journal of Economic Theory*, 115(1):191–206.
- Calvó-Armengol, A. and Jackson, M. O. (2007). Networks in labor markets: Wage and employment dynamics and inequality. *Journal of Economic Theory*, 132(1):27–46.
- Campbell, K. E. (1990). Networks past: A 1939 Bloomington neighborhood. *Social Forces*, 69(1):139–155.
- Carlsson, H. and Van Damme, E. (1993). Global games and equilibrium selection. *Econometrica*, 61(5):989–1018.
- Chung, F. and Lu, L. (2002). The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 99(25):15879–15882.

-
- Corominas-Bosch, M. (2004). Bargaining in a network of buyers and sellers. *Journal of Economic Theory*, 115(1):35–77.
- Costa-Gomes, M., Crawford, V. P., and Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, 69(5):1193–1235.
- Costa-Gomes, M. A. and Crawford, V. P. (2006). Cognition and behavior in two-person guessing games: An experimental study. *American Economic Review*, 96(5):1737–1768.
- Crawford, V. P. (2003). Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Economic Review*, 93(1):133–149.
- Crawford, V. P. and Iriberri, N. (2007). Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner’s curse and overbidding in private-value auctions? *Econometrica*, 75(6):1721–1770.
- Currarini, S., Jackson, M. O., and Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1):113.
- Dawid, H. and Gemkow, S. (2014). How do social networks contribute to wage inequality? Insights from an agent-based analysis. *Industrial and Corporate Change*, 23(5):1171–1200.
- de Montmort, P. R. (1713). *Essai d’analyse sur les jeux de hazard*. Quillau, Paris, seconde edition.
- Dekel, E., Ely, J. C., and Yilankaya, O. (2007). Evolution of preferences. *Review of Economic Studies*, 74(3):685–704.
- Dimitrova, M. and Schlee, E. E. (2003). Monopoly, competition and information acquisition. *International Journal of Industrial Organization*, 21(10):1623–1642.
- Ely, J. C. and Yilankaya, O. (2001). Nash equilibrium and evolution of preferences. *Journal of Economic Theory*, 97(2):255–272.
- Erdős, P. and Rényi, A. (1959). On random graphs I. *Publicationes Mathematicae Debrecen*, 6(1):290–297.

- Galeotti, A., Goyal, S., and Kamphorst, J. (2006). Network formation with heterogeneous players. *Games and Economic Behavior*, 54(2):353–372.
- Gilles, R. P. and Johnson, C. (2000). Spatial social networks. *Review of Economic Design*, 5(3):273–299.
- Golub, B. and Jackson, M. O. (2012). How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, 127(3):1287–1338.
- Goyal, S. (2012). *Connections: An Introduction to the Economics of Networks*. Princeton University Press, Princeton, NJ.
- Goyal, S. and Joshi, S. (2003). Networks of collaboration in oligopoly. *Games and Economic Behavior*, 43(1):57–85.
- Goyal, S. and Joshi, S. (2006). Bilateralism and free trade. *International Economic Review*, 47(3):749–778.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380.
- Granovetter, M. S. (1995). *Getting a Job: A Study of Contacts and Careers*. University of Chicago Press, Chicago, IL, second edition.
- Güth, W. (1995). An evolutionary approach to explaining cooperative behavior by reciprocal incentives. *International Journal of Game Theory*, 24(4):323–344.
- Güth, W. and Yaari, M. E. (1992). Explaining reciprocal behavior in simple strategic games: An evolutionary approach. In Witt, U., editor, *Explaining Process and Change: Approaches to Evolutionary Economics*, pages 23–34. University of Michigan Press, Ann Arbor, MI.
- Harsanyi, J. C. (1967). Games with incomplete information played by “Bayesian” players, part I. The basic model. *Management Science*, 14(3):159–182.
- Harsanyi, J. C. (1968a). Games with incomplete information played by “Bayesian” players, part II. Bayesian equilibrium points. *Management Science*, 14(5):320–334.
- Harsanyi, J. C. (1968b). Games with incomplete information played by “Bayesian” players, part III. The basic probability distribution of the game. *Management Science*, 14(7):486–502.

-
- Harsanyi, J. C. (1973). Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory*, 2(1):1–23.
- Hauk, E. and Hurkens, S. (2001). Secret information acquisition in cournot markets. *Economic Theory*, 18(3):661–681.
- Heifetz, A., Shannon, C., and Spiegel, Y. (2007a). The dynamic evolution of preferences. *Economic Theory*, 32(2):251–286.
- Heifetz, A., Shannon, C., and Spiegel, Y. (2007b). What to maximize if you must. *Journal of Economic Theory*, 133(1):31–57.
- Heller, Y. and Mohlin, E. (2015a). Coevolution of deception and preferences: Darwin and Nash meet Machiavelli. Working Paper, Available at SSRN 2490370.
- Heller, Y. and Mohlin, E. (2015b). Observations on cooperation. Working Paper, Available at SSRN 2558570.
- Hellmann, T. (2013). On the existence and uniqueness of pairwise stable networks. *International Journal of Game Theory*, 42(1):211–237.
- Hellmann, T. and Landwehr, J. (2014). Stable networks in homogeneous societies. Working Paper, Available at SSRN 2477452.
- Hellwig, C. and Veldkamp, L. (2009). Knowing what others know: Coordination motives in information acquisition. *Review of Economic Studies*, 76(1):223–251.
- Herold, F. and Kuzmics, C. (2009). Evolutionary stability of discrimination under observability. *Games and Economic Behavior*, 67(2):542–551.
- Ho, T.-H., Camerer, C., and Weigelt, K. (1998). Iterated dominance and iterated best response in experimental “p-beauty contests”. *American Economic Review*, 88(4):947–969.
- Hofbauer, J. and Sigmund, K. (1998). *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge.
- Horváth, G. (2013). Occupational mismatch and social networks. Working Paper, Available at SSRN 1794082.
- Hwang, H.-S. (1993). Optimal information acquisition for heterogenous duopoly firms. *Journal of Economic Theory*, 59(2):385–402.

- Ibarra, H. (1995). Race, opportunity, and diversity of social circles in managerial networks. *Academy of Management Journal*, 38(3):673–703.
- Iijima, R. and Kamada, Y. (2014). Social distance and network structures. Working Paper.
- Ioannides, Y. M. (1997). Evolution of trading structures. In Arthur, W. B., Durlauf, S. N., and Lane, D. A., editors, *The Economy as an Evolving Complex System II*, pages 129–167. Addison-Wesley, Boston, MA.
- Ioannides, Y. M. and Loury, L. D. (2004). Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature*, 42(4):1056–1093.
- Jackson, M. O. (2005). A survey of network formation models: Stability and efficiency. In Demange, G. and Wooders, M., editors, *Group Formation in Economics: Networks, Clubs, and Coalitions*, pages 11–57. Cambridge University Press, Cambridge.
- Jackson, M. O. (2006). The economics of social networks. In Blundell, R., Newey, W. K., and Persson, T., editors, *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, Volume I*, pages 1–56. Cambridge University Press, Cambridge.
- Jackson, M. O. (2008a). Average distance, diameter, and clustering in social networks with homophily. In Papadimitriou, C. and Zhang, S., editors, *Internet and Network Economics*, pages 4–11. Springer, Berlin Heidelberg.
- Jackson, M. O. (2008b). *Social and Economic Networks*. Princeton University Press, Princeton, NJ.
- Jackson, M. O. (2010). An overview of social networks and economic applications. In Benhabib, J., Bisin, A., and Jackson, M. O., editors, *Handbook of Social Economics, Volume 1A*, pages 511–585. North-Holland, Amsterdam.
- Jackson, M. O. and Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44–74.
- Jansen, J. (2008). Information acquisition and strategic disclosure in oligopoly. *Journal of Economics & Management Strategy*, 17(1):113–148.
- Kalmijn, M. (2006). Educational inequality and family relationships: Influences on contact and proximity. *European Sociological Review*, 22(1):1–16.

-
- Karonski, M., Scheinerman, E. R., and Singer-Cohen, K. B. (1999). On random intersection graphs: The subgraph problem. *Combinatorics, Probability and Computing*, 8(1):131–159.
- Koçkesen, L., Ok, E. A., and Sethi, R. (2000a). Evolution of interdependent preferences in aggregative games. *Games and Economic Behavior*, 31(2):303–310.
- Koçkesen, L., Ok, E. A., and Sethi, R. (2000b). The strategic advantage of negatively interdependent preferences. *Journal of Economic Theory*, 92(2):274–299.
- Köhler, W. (1929). *Gestalt Psychology*. Liveright, New York, NY.
- Kranton, R. and Minehart, D. F. (2001). A theory of buyer-seller networks. *American Economic Review*, 91(3):485–508.
- Laumann, E. O. (1966). *Prestige and Association in an Urban Community: An Analysis of an Urban Stratification System*. Bobbs-Merrill, Indianapolis, IN.
- Laumann, E. O. (1973). *Bonds of Pluralism: The Form and Substance of Urban Social Networks*. John Wiley & Sons, New York, NY.
- Lazarsfeld, P. F. and Merton, R. K. (1954). Friendship as a social process: A substantive and methodological analysis. In Berger, M., editor, *Freedom and Control in Modern Society*, pages 18–66. Van Nostrand, New York, NY.
- Li, L., McKelvey, R. D., and Page, T. (1987). Optimal research for cournot oligopolists. *Journal of Economic Theory*, 42(1):140–166.
- Manea, M. (2011). Bargaining in stationary networks. *American Economic Review*, 101(5):2042–2080.
- Marsden, P. V. (1987). Core discussion networks of Americans. *American Sociological Review*, 52(1):122–131.
- Marsden, P. V. (1988). Homogeneity in confiding relations. *Social Networks*, 10(1):57–76.
- Matêjka, F. and McKay, A. (2012). Simple market equilibria with rationally inattentive consumers. *American Economic Review: Papers and Proceedings 2012*, 102(3):24–29.

- Matêjka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–298.
- MATLAB (2014). *Version 8.3 (R2014a)*. The MathWorks Inc., Natick, MA.
- Mayer, A. and Puller, S. L. (2008). The old boy (and girl) network: Social network formation on university campuses. *Journal of Public Economics*, 92(1):329–347.
- Maynard Smith, J. and Price, G. R. (1973). The logic of animal conflict. *Nature*, 246(5427):15–18.
- McKelvey, R. D., McLennan, A. M., and Turocy, T. L. (2014). Gambit: Software tools for game theory, version 15.0.0. <http://www.gambit-project.org>.
- McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444.
- Mead, G. H. (1934). *Mind, Self, and Society: From the Standpoint of a Social Behaviorist*. University of Chicago Press, Chicago, IL.
- Mengel, F. (2012). Learning across games. *Games and Economic Behavior*, 74(2):601–619.
- Milgram, S. (1967). The small world problem. *Psychology Today*, 2(1):60–67.
- Montgomery, J. D. (1991). Social networks and labor-market outcomes: Toward an economic analysis. *American Economic Review*, 81(5):1408–1418.
- Morris, S. and Shin, H. S. (1998). Unique equilibrium in a model of self-fulfilling currency attacks. *American Economic Review*, 88(3):587–597.
- Myatt, D. P. and Wallace, C. (2012). Endogenous information acquisition in coordination games. *Review of Economic Studies*, 79(1):340–374.
- Myerson, R. B. (1991). *Game Theory: Analysis of Conflict*. Harvard University Press, Cambridge, MA.
- Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review*, 85(5):1313–1326.
- Nash, J. F. (1950a). The bargaining problem. *Econometrica*, 18(2):155–162.

-
- Nash, J. F. (1950b). Equilibrium points in n -person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49.
- Nash, J. F. (1951). Non-cooperative games. *Annals of Mathematics*, 54(2):286–295.
- Nash, J. F. (1953). Two-person cooperative games. *Econometrica*, 21(1):128–140.
- Newman, M. E. (2003). The structure and function of complex networks. *SIAM Review*, 45(2):167–256.
- Newman, M. E. (2006). Random graphs as models of networks. In Bornholdt, S. and Schuster, H. G., editors, *Handbook of Graphs and Networks: From the Genome to the Internet*, pages 35–68. Wiley-VCH, Weinheim.
- O’Donnell, J. F. (2011). Characterisation of stable Manea bargaining networks. Honours thesis, School of Economics, The University of Queensland.
- Ok, E. A. and Vega-Redondo, F. (2001). On the evolution of individualistic preferences: An incomplete information scenario. *Journal of Economic Theory*, 97(2):231–254.
- Pareto, V. (1964). *Cours d’économie politique*. Librairie Droz, Geneva.
- Persico, N. (2000). Information acquisition in auctions. *Econometrica*, 68(1):135–148.
- Piaget, J. (1932). *Le jugement moral chez l’enfant*. Presses Universitaires de France, Paris.
- Polanski, A. and Vega-Redondo, F. (2013). Markets, bargaining, and networks with heterogeneous agents. Working Paper, Available at SSRN 2218605.
- Pruzhansky, V. (2011). Some interesting properties of maximin strategies. *International Journal of Game Theory*, 40(2):351–365.
- Rees, A. (1966). Information networks in labor markets. *American Economic Review*, 56(1):559–566.
- Ritzberger, K. (2002). *Foundations of Non-Cooperative Game Theory*. Oxford University Press, Oxford.
- Robalino, N. and Robson, A. J. (2012). The economic approach to ‘theory of mind’. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1599):2224–2233.

- Robalino, N. and Robson, A. J. (2015). The evolution of strategic sophistication. *American Economic Review*, forthcoming.
- Robson, A. J. and Samuelson, L. (2010). The evolutionary foundations of preferences. In Benhabib, J., Bisin, A., and Jackson, M. O., editors, *Handbook of Social Economics, Volume 1A*, pages 221–310. North-Holland, Amsterdam.
- Rogers, K., Dziobek, I., Hassenstab, J., Wolf, O. T., and Convit, A. (2007). Who cares? Revisiting empathy in Asperger syndrome. *Journal of Autism and Developmental Disorders*, 37(4):709–715.
- Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica*, 50(1):97–109.
- Rubinstein, A. and Wolinsky, A. (1985). Equilibrium in a market with sequential bargaining. *Econometrica*, 53(5):1133–1150.
- Samuelson, L. (2001). Introduction to the evolution of preferences. *Journal of Economic Theory*, 97(2):225–230.
- Schelling, T. C. (1956). An essay on bargaining. *American Economic Review*, 46(3):281–306.
- Selten, R. (1965). Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragerträglichkeit - Teil I: Bestimmung des dynamischen Preisgleichgewichts. *Zeitschrift für die gesamte Staatswissenschaft*, 121(2):301–324.
- Shamay-Tsoory, S. G., Aharon-Peretz, J., and Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, 132(3):617–627.
- Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.
- Sims, C. A. (2006). Rational inattention: Beyond the linear-quadratic case. *American Economic Review*, 96(2):158–163.
- Solan, E. and Yariv, L. (2004). Games with espionage. *Games and Economic Behavior*, 47(1):172–199.
- Stahl, D. O. and Wilson, P. W. (1994). Experimental evidence on players' models of other players. *Journal of Economic Behavior & Organization*, 25(3):309–327.

-
- Stahl, D. O. and Wilson, P. W. (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1):218–254.
- van der Leij, M. and Buhai, S. (2008). A social network analysis of occupational segregation. Working Paper, Available at SSRN 1117949.
- Verbrugge, L. M. (1977). The structure of adult friendship choices. *Social Forces*, 56(2):576–597.
- von Neumann, J. (1928). Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen*, 100(1):295–320.
- von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ.
- Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^* . *Psychometrika*, 61(3):401–425.
- Watts, A. (2001). A dynamic model of network formation. *Games and Economic Behavior*, 34(2):331–341.
- Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442.
- Weibull, J. W. (1995). *Evolutionary Game Theory*. MIT Press, Cambridge, MA.
- Weinstein, J. and Yildiz, M. (2007). A structure theorem for rationalizability with application to robust predictions of refinements. *Econometrica*, 75(2):365–400.
- Wellman, B. (1996). Are personal communities local? A Dumptarian reconsideration. *Social Networks*, 18(4):347–354.
- West, D. B. (2001). *Introduction to Graph Theory*, volume 2. Prentice-Hall, Englewood Cliffs, NJ.
- Zaharieva, A. (2013). Social welfare and wage inequality in search equilibrium with personal contacts. *Labour Economics*, 23(1):107–121.

Short Curriculum Vitae of Florian Gauer

Academic Education

- | | |
|-----------------|---|
| since 10/2012 | Doctoral Student at the International Research Training Group <i>Economic Behavior and Interaction Models</i> (EBIM) at Bielefeld University in collaboration with the University Paris 1 Panthéon-Sorbonne
Member of the <i>Bielefeld Graduate School of Economics and Management</i> (BiGSEM) and the <i>Center for Mathematical Economics</i> (IMW) at Bielefeld University |
| 04/2010–04/2012 | Master of Science Wirtschaftsmathematik (Mathematical Economics) at Bielefeld University |
| 10/2006–02/2010 | Bachelor of Science Wirtschaftsmathematik (Mathematical Economics) at Bielefeld University
Exchange student at the University of Edinburgh for one semester |

Working Papers

- Gauer, F. and Kuzmics, C. (2016). Cognitive Empathy in Conflict Situations. *Center for Mathematical Economics Working Papers*, No. 551.
- Gauer, F. (2015). Strategic Formation of Homogeneous Bargaining Networks. *Center for Mathematical Economics Working Papers*, No. 529.
- Gauer, F. and Landwehr, J. (2015). Continuous Homophily and Clustering in Random Networks. *Center for Mathematical Economics Working Papers*, No. 515.

Conference and Seminar Presentations

- University of Graz, 11/10/2015: Presentation on “Cognitive Empathy in Conflict Situations” in the *Economics Research Seminar*
- University of Münster, 09/08/2015: Presentation on “Strategic Formation of Homogeneous Bargaining Networks” at the *Verein für Socialpolitik Annual Conference 2015*
- University of Cambridge, 07/31/2015: Presentation on “Strategic Formation of Homogeneous Bargaining Networks” at the *15th SAET Conference on Current Trends in Economics*
- University Paris 1 Panthéon-Sorbonne, 02/25/2014: Presentation on “Strategic Formation of Homogeneous Bargaining Networks” in the *Network Research Seminar*