Efficiency and Other-Regarding Preferences in Information and Job-Referral Networks

Antonio Stefano Caria

St Antony's College
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy in Economics

Efficiency and Other-Regarding Preferences in Information and Job-Referral Networks

Antonio Stefano Caria
DPhil in Economics
Michaelmas 2014

Abstract

In this thesis I study how networks are formed and I analyse the strategies that well-connected individuals adopt in public good games on a network.

In chapter one I study an artefactual field experiment in rural India which tests whether farmers can create efficient networks in a repeated link formation game, and whether group categorisation increases the frequency of in-group links and reduces network efficiency. I find that the efficiency of the networks formed in the experiment is significantly lower than the efficiency which could be achieved under selfish, rational play. When information about group membership is disclosed, in-group links are chosen more frequently, while the efficiency of network structure is not significantly affected.

Using a job-referral network experiment in an urban area of Ethiopia, I investigate in chapter two whether individuals create new links with the least connected players in the network. In a first treatment, competition for job-referrals makes it in the player's interest to link with the least connected partners. In this treatment, links to the least connected players are significantly more likely than links to better connected individuals. In a second treatment, connections only affect the welfare of the new partner. Choosing the least connected player minimises inequality and maximises aggregate efficiency. This may motivate other-regarding players. In this treatment, however, links to least connected partners are not significantly more likely than links to other players.

In chapter three I explore the characteristics that individuals value in the people they approach for advice. Using cross-sectional data on cocoa farmers in Ghanaian villages and a matched lottery experiment, I find an association between the difference in the aversion to risk of two farmers and the probability that one farmer is interested in the advice of the other farmer.

In chapter four I study a one-shot public good game in rural India between farmers connected by a star network. Contributions by the centre of the star have a larger impact on aggregate payoffs than contributions by the spoke players. I use the strategy method to study whether the centre of the star contributes more than the average of the spokes. In selected sessions, I disclose participants' expectations about the choices of the centre of star. I find that the centre player contributes just as much as the average of the spokes, and that he is influenced by the expectations that other players hold about his decisions.

Acknowledgment

First and foremost, I would like to thank my supervisors Marcel Fafchamps and Johannes Abeler: without them this work would not have been written. They have been engaged, generous with their time, supportive and, when needed, not afraid to be critical. I hope I will be able to learn from their steadiness, lucidity and intellectual curiosity.

This work was supported by the Economic and Social Research Council [ES/J500112/1]. Funding for the experiments was provided by the Department for International Development (DfID), under the iiG programme, and the International Food Policy Research Institute. I provide more details in each chapter.

All field projects have benefitted from tremendous logistic support. In India, Gaurav Puntambekar and Kabira Namit provided outstanding research assistance. In Ethiopia, Ibrahim Worku Hassen made the fieldwork possible in the first place. I also thank Francis Teal and Andrew Zeitlin for granting me access to the data from the 2008 GCFS dataset. A large number of enumerators and supervisors contributed to the four projects: I would like to thank them all for their great effort.

In Ethiopia, IFPRI ESSPII, and in particular Paul Dorosh, Bart Minten and Alemayehu Seyoum Taffesse supported my research in many ways. I wish them best luck with the important work they are doing.

My gratitude goes to the participants in the experiments and surveys. Their trust, patience, hospitality will be hard to forget.

A number of friends and colleagues have offered many comments on my work and contributed a great deal to it. I owe it to them if the DPhil did not feel like a solitary adventure. I hope we can continue to be intellectual partners for much longer.

I had the luck to be able to count on some great friends in these three years. Their joyful companionship has been invaluable.

Mamma, Papá, Mario, Nonna: questa tesi la dedico a voi!

Contents

List of Figures

List of Tables

Introduction

In this thesis I study the formation of networks and I analyse the strategies of individuals who have a central position in the fabric of social connections.

Individuals interact with others in many ways: they exchange information and opinions, provide advice and material assistance, form partnerships. This has important consequences for the way an economy functions. A long-standing research program in economics, anthropology and sociology highlights how the diffusion of innovations, the selection of personnel, insurance against risk– to name just a few– are influenced by personalised exchange of information, referrals, and resources. These studies emphasise that the overall structure of connections determines the impact of individual actions and shapes aggregate outcomes.

We do not fully understand how networks are formed. There is an established theoretical literature in this area and a number of empirical patterns are well documented– for example, a skewed distribution of connections and the tendency of similar individuals to interact with each other (Goyal, 2007; Jackson, 2010). Evidence on the individual decisions that bring these patterns about, however, is limited (Jackson, 2014). The economic approach suggests that individuals choose connections weighting the costs and benefits of their actions. Benefits can be determined both by self concern and by regard for other players. This regard can take several forms and may be influenced by the nature and framing of the decision (Charness and Rabin, 2002; Falk and Szech, 2013). To explain network formation, we need to understand the objectives that individuals pursue when they connect to others. Can simple models of strategic and other-regarding behaviour predict the connections that individuals choose?

It is difficult to investigate preferences using observational data alone: desired links may not correspond to the links that obtain in equilibrium, and it is hard to control for confounding factors (Fafchamps and Gubert, 2007; Comola and Fafchamps, 2013). Experimental evidence on network formation has also so far been circumscribed (Goeree et al., 2009; Falk and Kosfeld, 2012).

I study link-formation decisions in two experiments. This allows me to analyse separately the decisions individuals take when connections affect their personal payoff and when they affect the payoff of others. Within the experiments, I can also precisely measure efficiency. Furthermore, I study the desired links reported by respondents in a survey. Through a combination of observational data and a matched incentivised game, I investigate whether individuals consider the preferences of their potential partners when they form connections. The results I present contribute to the literature on strategic network formation (Jackson and Wolinsky, 1996; Bala and Goyal, 2000), other-regarding preferences (Fehr and Schmidt, 1999; Charness and Rabin, 2002) and social identity (Akerlof and Kranton, 2000).

In a final experiment, I analyse the public good contributions of the central player in a network (Bramoulle and Kranton, 2007). The results speak to the experimental literature on cooperation (Ledyard, 1995; Chaudhuri, 2011).

The empirical evidence used for this thesis is collected among farmers and urban dwellers in India, Ethiopia and Ghana. These are populations for which the policy prescriptions of network theory are relevant. Reliance on relationships is widespread in developing countries (Fafchamps, 2004) and the theory has shown some of its most promising policy applications in the context of development interventions (Banerjee et al., 2013; BenYishay and Mobarak, 2014). Network theory has been particularly well-received by development economists.

In the rest of this introduction, I outline the specific contributions of each chapter. In the first chapter[1] we use a simple experiment to show that the networks formed by small groups of Indian farmers are inefficient: a change in the pattern of connections

[1] This chapter is joint work with Marcel Fafchamps (Stanford University).

would generate higher total welfare and weakly increase the expected payoff of every player. We further illustrate how initial assignment of players to groups changes the structure of the network, but does not affect its efficiency.

We study a modification of the game of unilateral, one-way-flow link formation proposed by Bala and Goyal (2000). Directed links are added sequentially to an empty network. Each player can only create one link, without requiring the consent of the new partner. We manipulate the direction of the link that players form to separate self and other-regarding behaviour. Crucially, individual incentives are aligned with the achievement of the socially-efficient outcome. In half of the sessions, we publicly disclose information about players' membership in groups created at the beginning of the experiment.

Our empirical illustration of the inefficiency of networks contributes to the theoretical and experimental literature on strategic network formation (Jackson and Wolinsky, 1996; Bala and Goyal, 2000; Falk and Kosfeld, 2012). We show that while many players choose strategies consistent with selfish best response or well-known other-regarding concerns (Charness and Rabin, 2002), a small group of subjects relies on inappropriate heuristics, distorting aggregate outcomes. The insight that strategic and heuristic behaviour coexist can be used in the future to explore empirical departures from standard theoretical predictions about network structure.

We also contribute to identity economics (Akerlof and Kranton, 2000). Following the early work of Tajfel (1981), experiments have shown that when players are categorised in groups they modify their allocation decisions and strategies, and that groups created with arbitrary assignment rules are sufficient to manipulate behaviour. Our findings show that assignment to arbitrary groups affects choice in one more domain: the formation of networks.

Third, our results inform the discussion on homophily- the high relative frequency of links between individuals with similar characteristics, which has been widely documented across social contexts (McPherson et al., 2001; Currarini et al., 2009; Golub and Jackson, 2012). Some of the explanations put forward in the literature suggest that individuals have more opportunities to connect to in-group members, or they derive economic benefits from restricting interaction to the in-group. The treatment effect we report cannot be easily explained by either of these factors, while it is consistent with the existence of norms or preferences for homophily.

In this experiment we are interested to observe networks that are the product of multiple individual decisions. The effect of a new connection on payoffs thus depends on the connections that other players form in subsequent rounds. However, this feature complicates the interpretation of players' choices, as we do not observe their beliefs about the evolution of the network and about the effects of their strategies on other players. In the second chapter, I propose an experimental design that allows cleaner identification of players' motives.

The second chapter reports the results of a link-formation experiment in urban Ethiopia. Participants are assigned a position in an asymmetric job-referral network, determining who can refer whom for a remunerated activity, which I call a 'job' . This is similar to the game studied by Calvo-Armengol (2004). An initial round of jobs is assigned through a lottery. Winners of the lottery perform the task and also refer for the job one of their jobless network partners, chosen at random. In this game, links to a player with the minimum number of connections give the highest chance of being referred for the job. Furthermore, least-connected players have the smallest probability of obtaining a referral for themselves and enjoy the largest increase in expected payoffs from an additional link.

The players' task is to specify, before the job lottery, two additional partners with whom they would like to connect. Only the links indicated by one individual are eventually added to the network, ensuring participants do not have to worry about how other players' decisions affect the structure of the network. I manipulate whether this individual is randomly chosen among those who did not win the job-lottery and would benefit from a referral (treatment one), or among those who got the job and will give a referral to others (treatment two). This allows me to separate self-regarding and other-regarding behaviour.

For treatment one, I find that links with the least-connected players are significantly more frequent than other links. 47 percent of connections are chosen in this way. This confirms that many individuals direct their new links to players who occupy valuable positions in the network, a central tenet in the theory of strategic network formation (Jackson and Wolinsky, 1996; Bala and Goyal, 2000). It is common to assume that the most valuable connections are those with individuals in a central position (Barabási and Albert, 1999). My results show that when links to poorly-connected individuals are preferable, individuals strategically avoid links with socially central partners.

On the contrary, in treatment two, links to the least-connected players are not significantly more frequent than other links. This poses a puzzle. On one hand, players transfer significant amounts in an initial dictator game, which is consistent with the presence of widespread other-regarding preferences like inequality aversion or social-welfare maximisation (Fehr and Schmidt, 1999; Charness and Rabin, 2002). On the other hand, the same individuals are not significantly more likely to choose links that would increase the expected sum of payoffs across players and reduce payoff differences, at no personal monetary cost. Furthermore, the difference between treatment one and two is driven precisely by players who send a positive amount in the dictator game. This finding is difficult to reconcile with simple models of other-regarding preferences. It may be explained by exhaustion of pro-sociality across tasks, or by the presence of norms or social expectations that dictate equal splits in allocations problems and links to central individuals in link-formation tasks.

The value of partners reside as much in their position in the network as in their characteristics. In chapter three I explore the responses of a sample of Ghanaian cocoa farmers to a question on their willingness to seek the agricultural advice of other farmers in their village. A matched lottery experiment allows me to measure farmers' attitudes towards risk. I find an effect of relative risk aversion on willingness to seek advice: the more risk averse the potential advisor compared to the respondent, the less likely is the respondent to be interested in his advice.

This chapter contributes to the observational literature on network formation (Fafchamps and Gubert, 2007). In particular, it provides direct evidence on the characteristics of partners that individuals value. When individuals can refuse connections with others, desired links do not necessarily coincide with actual links. Hence, the results I present complement the large number of studies which analyse actual connections.

This chapter also contributes to the a new literature which shows that individuals consider strategically the preferences of others (Attanasio et al., 2012; Finan and Schechter, 2012). This observation can be used to understand how interventions that target individuals with specific preferences, for example the provision of insurance to the risk averse, can affect the patterns of social interaction of their beneficiaries.

While the first three chapters study the formation of networks, the final chapter analyses the decisions of players embedded in a network.[2] We are particularly interested in how other-regarding individuals act from a central position in the network. In the standard public good game played by a group, subjects are typically prepared to contribute as much as the average contribution of the other players, a strategy known as 'conditional cooperation' (Fischbacher et al., 2001; Chaudhuri, 2011). In our experiment, we study whether the central player in a star network conditionally cooperates with the spoke players.

In an homogeneous group, conditional cooperation is consistent with the objective

[2]This chapter is joint work with Marcel Fafchamps (Stanford University).

of payoff equality. Furthermore, every player is equally effective at increasing aggregate payoff, making differences in contributions hard to justify. In our experiment, on the other hand, connections in the network determine who benefits from the public good contributions of whom as in Bramoulle and Kranton (2007). In this setup, the centre of the star has an influence on aggregate payoffs that is greater than that of the spokes. He also personally benefits from the contributions of all the spokes. Payoff equality requires higher public good provision from the central player.

We find that, on average, players in the centre of the star contribute just as much as the spokes. This extends our understanding of the games where conditional cooperation obtains, contributing to the large literature on the provision of public goods (Ledyard, 1995; Chaudhuri, 2011) and to the study of cooperation in networks (Bramoulle and Kranton, 2007). In terms of methodology, we show how the strategy method can be adapted to study strategic decisions in networks.

We ask participants to report their expectations about the decisions of the player at the centre of the star. In selected sessions, we disclose the average of these expectations before contributions are chosen. We find that players match the disclosed values frequently, and do so more often as the cost of contributions is decreased. These results are consistent with a model of 'guilt aversion', where players dislike to determine a payoff for other players that is lower than what these players expect (Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007). We contribute to this literature by suggesting that a concern for the expectations of others may play a role in motivating individuals who have a special influence on social welfare.

1 Can farmers form efficient information networks? Experimental evidence from rural India

1.1 Introduction

It is well-known that knowledge about new technologies diffuses through social networks. Farmers, for example, share with each other information about the profitability and optimal use of agricultural innovations (Foster and Rosenzweig, 1995; Munshi, 2004; Bandiera and Rasul, 2006; Conley and Udry, 2010; Krishnan and Patnam, 2012). Peer-learning also been documented for new health technologies and financial products (Kremer and Miguel, 2007; Oster and Thornton, 2012; Banerjee et al., 2013). The pattern of this information diffusion is influenced by the structure of networks. For example, when some individuals have no connections, diffusion can be partial. When information has to travel through long chains of links, diffusion can be slow.[1] Efficient information networks maximise the total value of the information circulated, net of the cost of diffusion (Bala and Goyal, 2000).

Evidence on the *efficiency of information networks* is scarce. Little is known, for example, about the efficiency of these networks among farmers. Two recent randomised control trials show that monetary incentives for information agents in rural commu-

This chapter is joint work with Marcel Fafchamps (Stanford University). We are first of all indebted to Gaurav Puntambekar for his outstanding management of the project. We would like to thank Johannes Abeler, Francesco Cecchi, Megenagna Gashaw Guade, Esteban Ortiz Ospina, Cecilia Piantanida, Simon Quinn, Bassel Tarbush, Mateusz Ujma and seminar participants at the 2014 NEUDC conference, the 2014 PACDEV conference, the 2014 conference of the Centre for the Study of African Economies, the 2014 CESifo conference on social economics, and at the CSAE workshop series for useful comments and suggestions. Our gratitude also goes to the the enumerators and supervisors whose contributions made this study possible. We finally would like to thank the study participants for their goodwill and patience. All mistakes remain naturally ours. This research was funded by the UK Department for International Development (DFID) as part of the iiG, a research programme to study how to improve institutions for pro-poor growth in Africa and South-Asia. The views expressed are not necessarily those of DFID.

[1] Network structure also has important consequences for learning when information is noisy and individuals have to aggregate the signals generated by multiple experimenters (Bala and Goyal, 1998).

nities improve the diffusion of information in a cost-effective fashion (BenYishay and Mobarak, 2014; Berg et al., 2013). This *indirectly* suggests that un-incentivised diffusion is suboptimal. *Direct* evidence is largely missing.

Observational assessment of network efficiency is complicated by several factors. First, a census of all individuals and all links is required. Such data is usually hard to obtain. Furthermore, the costs and benefits of each link have to be quantified. These are often not observed by the field researcher. Finally, with observational data alone, it is difficult to attribute inefficiency to the preferences and decision-making rules of individuals, or to the constraints individuals face. In this study, we instead rely on an artifactual laboratory experiment that allows us to quantify the efficiency of small experimental networks, with a great deal of control over the constraints imposed on decision makers (Harrison and List, 2004).

We study whether farmers form efficient networks in a sequential link-formation game. Our motivating example is the diffusion of information. A link, in this example, represents a social interaction where new information is observed. Each farmer can form one link to another player, without requiring the partner's consent. Observation occurs only in one direction. Crucially, when farmer i observes the information of farmer j, he has also access to the information which farmer j has acquired by observing other players. The benefits of a link are thus proportional to the number of players that can be 'reached' directly or indirectly thanks to this link. This set-up is that of unilateral, one-way flow link formation with no decay discussed in Bala and Goyal (2000).

We manipulate the basic game along two dimensions. First, we vary the direction of information exchange. In the first treatment, a player selects *the partner he would like to observe*. In the second treatment, a player chooses *the partner by whom he will be observed*. In both treatments, the cycle network is efficient[2], a Nash equi-

[2]As in Bala and Goyal (2000), we define welfare as the sum of players' payoffs. The efficient network

librium, and generates no inequality in payoffs across agents. Selfish agents playing best response converge to the cycle network after repeated play of the first treatment.[3] Efficiency-minded, other-regarding players converge to the cycle network in the second treatment. By ruling out tradeoffs between efficiency and equilibrium, limiting coordination requirements, and anonymising interaction we give the 'best shot' to the possibility of efficient networks emerging in the game.

Second, we vary whether farmers have knowledge about the group affiliation of the other players. At the beginning of the experiment we randomly assign farmers to groups that have to compete in an unrelated task. In selected experimental sessions, we publicly disclose information about players' group identity.[4] We test whether, as a result, the number of in-group links increases and the efficiency of networks decreases. Social differentiation may discourage links that are desirable from an efficiency point of view. Observational research on networks often points to the importance of homophily- the tendency of similar individuals to interact with each other with disproportionate frequency (McPherson et al., 2001; Currarini et al., 2009; Golub and Jackson, 2012). Homophily can be the result of a norm related to group membership (Akerlof and Kranton, 2000). In our game, restricting links to in-group partners would result in large efficiency losses.

When group identity is not disclosed, we predict that individuals will play simple, intuitive link-formation rules leading to high levels of efficiency. In the first treatment, the selfish best response is to choose the player who reaches, directly or indirectly, the highest number of individuals in the network. In the second treatment, other-regarding

maximises this sum of payoffs. The cycle is the (unique) efficient network in our game. We define a relative measure of efficiency by comparing the sum of expected payoffs determined by a particular network to the sum of expected payoffs which obtains in the cycle network.

[3]Farmers play sequentially. We describe this in detail in the next section and show simulation results indicating that convergence to the cycle network is fast when players follow selfish best response. This is not surprising. Bala and Goyal (2000) study a sequential game with features very similar to ours. Part (a) of their theorem 3.1 applies to a class of payoff functions that includes the one of our game. For these payoff functions, theorem 3.1 shows that the network converges to the cycle with probability one when players play selfish best-response.

[4]Personal identity, on the other hand, is never disclosed.

players who want to maximise the sum of payoffs in the group will target the player who is reached by the highest number of individuals. Players who instead want to maximise the payoff of the least well-off peer will choose the farmer who reaches the smallest number of individuals. We derive these predictions from standard models of strategic network formation augmented to include other-regarding preferences (Bala and Goyal, 2000; Charness and Rabin, 2002).

When group identity is disclosed, we predict individuals will choose in-group links more frequently, possibly at the cost of establishing less efficient networks. Previous research has highlighted how group categorisation generates in-group favoritism (Tajfel, 1981; Brewer, 1999; Akerlof and Kranton, 2010), affects social, risk and time preferences (Benjamin et al., 2010; Chen and Li, 2009; Kranton et al., 2012), influences behaviour in strategic environments (Yamagishi and Kiyonari, 2000; Charness et al., 2007), and modifies performance (Hoff and Pandey, 2006). Akerlof and Kranton (2000) posit that agents receive utility from following prescriptions associated with social categories. Many farmers in our experiment report that restricting links to the in-group is prescribed. Thus, when the preferred partner belongs to the out-group, a farmer faces a tradeoff between efficiency and conformity with the social prescription. We expect that at least some subjects will choose to conform.

In terms of methodology, we take steps against a number of common confounders of experimental inference: low understanding, side payments, wealth effects and experimenter demand effects. We rely on induced, randomised group membership to rule out unobserved covariates that may be correlated with natural groups. As the saliency of induced group membership has been found to influence behaviour in economic experiments (Charness et al., 2007; Eckel and Grossman, 2005), we increase saliency by means of an independent task that sets the two groups in competition.

We run our experiment in the Indian state of Maharashtra. With the many social identities based on caste, religion and class, India offers an appropriate setting to study homophily in social networks (Beteille and Srinivas, 1964; Guha, 2008; Dunning and Nilekani, 2013). Recent work on information agents in rural communities indeed suggests that social distance affects the probability of information diffusion and experimental work has shown that priming natural identities, chiefly caste, affects individual performance and economic outcomes (Hoff and Pandey, 2006; Anderson, 2011; Berg et al., 2013). In India, interest in novel agricultural extension approaches that exploit farmers' dense social network activity is also high.

Our findings can be summarised as follows. First, network efficiency is significantly lower than the level of efficiency which selfish or efficiency-minded players would have achieved playing the rules we outline above. Expected payoffs are, on average, only 65 percent of those in the cycle network. Interestingly, farmers in the second treatment achieve levels of network efficiency similar to those of farmers in the first treatment.

Second, the link-formation rules we derive have considerable predictive power. In the second treatment, for example, 70 percent of decisions are consistent with either of the two rules outlined above. Regression analysis confirms the statistical significance of this result. We also identity two additional link-formation rules which have further predictive power on link-formation decisions: choosing the 'most popular' player in the network, and choosing a player by whom one was chosen in a previous turn. About 65 percent of the decisions that are not consistent with the predicted rules target the 'most popular' player in the network. Simulation analysis suggests that the largest gains in efficiency could be achieved by reducing the proportion of decisions that follow this rule.

Third, when information about group membership is disclosed, the resulting networks have more in-group links, but are not significantly less efficient. These effects can occur simultaneously if farmers (i) always follow their chosen link-formation rule, and (ii) whenever this rule is satisfied by both in-group and out-group partners they prefer to link with an in-group partner. We have evidence consistent with this interpretation. In treatment one, where the effect of group identity is more pronounced, the frequency of links that satisfies the selfish link formation rule is unaffected by disclosure of group identity. However, the proportion of 'selfish' links that are directed to an in-group partner grows by 13.8 percentage points.

Our work relates most directly to the literature on network formation. This has developed theoretically through the seminal contributions of Jackson and Wolinsky (1996) and Bala and Goyal (2000). Experimental work on link formation has been motivated by these models and has explored issues of inequity aversion (Goeree et al., 2009; Van Dolder and Buskens, 2009; Falk and Kosfeld, 2012), coordination (Berninghaus et al., 2006), and whether chosen links are myopic best responses or far-sighted strategies (Callander and Plott, 2005; Conte et al., 2009; Kirchsteiger et al., 2011). In a related experiment, Belot and Fafchamps (2014) compare unilateral partnership formation decisions to dictator game allocations with equivalent payoff consequences. All of these experiments use western subjects, typically university students.

A parallel literature has used observational dyadic data from rural areas of developing countries to explore how specific networks for the sharing of risk, favours, information and labour are formed (Fafchamps and Gubert, 2007; Krishnan and Sciubba, 2009; Karlan et al., 2009; Comola, 2010; Jackson et al., 2012; Santos and Barrett, 2010; Comola and Fafchamps, 2013). Empirical studies of naturally occurring networks typically document some degree of homophily (McPherson et al., 2001). A recent theoretical literature distinguishes between homophily motivated by *preferences*, *opportunities* or *strategic behaviour* (Currarini et al., 2009; Currarini and Menge, 2012; Tarbush and Teytelboym, 2014).[5]

Falk and Kosfeld (2012) study a game of unilateral, one-way-flow link formation that is based on Bala and Goyal (2000) and hence is closely related to ours. The

[5]Individuals may have a desire to match with in-group partners, they may simply be exposed to more potential matches with in-group peers, or they may link with in-group peers because it is in their material interest to do so, for example, in order to avoid sanctions associated with deviations from social norms.

design we propose, however, differs on a number of dimensions: links are added to the network one at a time; players are allowed only one link, so that the only cost of a connection is the opportunity cost of not forming another connection; the game is played by groups of 6. The first two features limit coordination and computation problems and make the game simpler. Falk and Kosfeld (2012) find that efficient networks are achieved in about half of the periods of the game. However, they do not report an average efficiency statistics for the one-way-flow treatment, which makes it difficult to compare their results to ours.[6]

A second literature that we connect to is that on the diffusion of innovations. Foster and Rosenzweig (1995); Munshi (2004); Bandiera and Rasul (2006); Conley and Udry (2010); Krishnan and Patnam (2012) show how technologies diffuse through farmers' networks in India, Mozambique, Ghana and Ethiopia. Duflo et al. (2011), however, cannot find evidence of learning among maize farmer in Kenya. More recently, Centola (2010, 2011); Banerjee et al. (2012) and Banerjee et al. (2013) use experimental techniques to investigate how the structure of the network and the position of the first injection point affect diffusion.

Finally, our study is related to identity economics and a rich literature in economics and social psychology, referenced above, that studies how group categorisation generates in-group bias and modifies behaviour. A recent experiment by Currarini and Menge (2012) shows that group categorisation produces both in-group bias in allocation and a positive willingness to pay for being matched with an in-group player.

We can summarise our contributions as follows. First, we document low levels of network efficiency in unilateral, one-way-flow link formation. This is a stark result in a simple game with limited coordination issues and clear theoretical predictions. The literature has recently started exploring how subjects in rural areas of developing

[6] We can however note that in the last period of our game, the cycle network is achieved in less than 10 percent of the sessions. This is frequency is much lower than in any of the one-way-flow treatments studied by Falk and Kosfeld (2012).

countries achieve lower efficiency in trading experiments than Western players (List, 2004; Bulte et al., 2012). Our findings suggest that networks could be a second domain of widespread inefficiency and cross-cultural difference.

Second, we document heterogeneous link formation strategies that are consistent with a simple models of other-regarding preferences. This can inspire further theoretical study of network formation that explicitly includes other-regarding preferences.

Third, our results document an effect of arbitrary group categorisation on networks. This expands our understanding of the settings in which group categorisation modifies behaviour. It also shows how, beyond the material payoffs of the game, the formation of networks can be influenced by features of the social world. In the experiment, *opportunities* to link with in-group and out-group partners are kept fixed. There are also limited *strategic reasons* to choose in-group partners: personal anonymity makes it difficult to influence interaction with participants after the experiment and, in treatment one– where we find a more pronounced effect of group identity– opportunities for reward or punishment within the game are restricted. Our results are thus consistent with the existence of *norms* or *preferences* to restrict links to the in-group. This could be a particularly fertile area for future experimental research on networks.

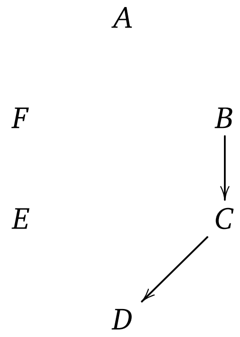
The chapter is organised as follows. Section 1.2 presents the design. Section 1.3 develops predictions and testable hypotheses. Section 1.4 describes the data. Section 1.5 reports the results of the analysis. Section 1.6 concludes.

1.2 Design

The link formation game is played by groups of 6 farmers. One of the farmers is randomly drawn at the end of the experiment to receive a monetary prize.[7] In our

[7]The prize is worth 100 Indian Rupees, or about 5.2 USD at PPP, given an exchange rate at the time of 0.0155 USD per INR and a PPP conversion factors of 10/3 from the 2011 ICP round (http://data.worldbank.org/indicator/PA.NUS.PPPC.RF). This figure would grow if we applied a conversion factor calculated using rural prices only. For comparison, notice also that in 2012-2013 the

Figure 1.1: Example



motivating example, the prize corresponds to a piece of valuable information. Farmers form a network to diffuse this information. The initial network has no links. Each farmer can add to this empty network a link between himself and another player.[8] This decision is unilateral: the consent of the other player is not required. Players who are directly or indirectly linked to the winner of the prize in the final network receive a prize of equivalent value. In other words, once a link is in place, the prize is non-rival and non-excludable. This is the value of a link. Such benefit flows only in one direction. For example, in the network represented in figure 1.1, farmer B reaches farmers C and D, and hence receives the prize if C or D win the prize draw. Player C reaches only player D, while players D, E, F and A do not reach any player.

Play is sequential. The game is divided in two rounds. Each round comprises 6 turns. In every turn, only one player takes a decision. Each player is randomly assigned to one turn per round. Participants are informed of this rule, but do not know the particular order of play which has been drawn for their session. In the first round, players create one link. In the second round, players can rewire their existing link.[9]

Players' decisions are recorded on a network map drawn on a white board visible to

National Rural Employment Guarantee Scheme paid an average daily wage to his workers of 121 INR (http://nrega.nic.in/netnrega/home.aspx).

[8]The instructions for one of the treatments are included in the appendix. The remaining experimental materials can be found here: https://sites.google.com/site/stefanoacaria/lfindia.

[9]In both rounds players have the options not to form any link. This option is rarely used.

all players. The map is updated after every turn. A number of design features ensure sequential updating takes place without breaking anonymity.[10] Furthermore, the pilot revealed that when the network map has more than a few links players find it difficult to calculate the number of direct and indirect connections of each peer. We hence remind the decision maker of the total number of (direct and indirect) connections every other player has in the current network. The counting of connections is done by means of a Java application running on a small laptop operated by the game assistant. After entering a new link, the software produces a table with the number of (direct and indirect) connections of each player in the current network. This number is written next to the respective player ID on the white board, immediately after the network map has been updated with a new link.

The experimental tasks are carried out in the following order:

1. Players randomly draw a card from an urn which assigns them a letter ID and an experimental group.

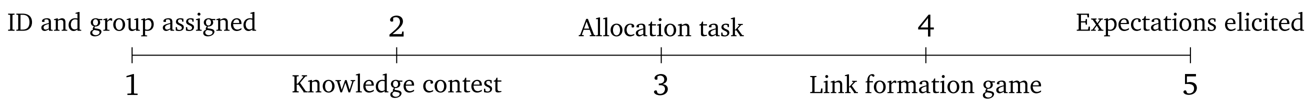
2. Players answer three questions on agricultural knowledge, which are part of an intergroup contest in agricultural knowledge. At the end of the experiment, if all players in a group have answered all questions right, the group receives one point and is applauded by everyone. Points are summed across sessions and participants are informed of the overall ranking between the two groups.[11]

3. Players play a simple allocation task, where they have to divide a fixed sum of

[10]Participants record their decisions on a game sheet. Modified cardboard boxes ensure participants cannot see what other players are choosing. However, the boxes do not prevent players to infer from a peer's body movements whether he is updating his game sheet or not. This threatens anonymity as it is possible to determine which participant has the turn by simply checking who is updating his game sheet at a given point in the game. We solve this problem in the following way. At the beginning of each turn, the game assistant publicly calls the ID of the player who has the turn. After allowing some time to look at the updated network map, the game assistant asks all players to make a circle on their game sheet. The player with the turn circles the ID letter of the player to whom he would like to link, while the players without the turn draw a circle in an empty box provided on the same page of the game sheet. As everybody writes something on their game sheet at the same time farmers cannot infer the identity of the player with the turn by checking who is updating his game sheet at a given point.

[11]Information about the overall ranking is disclosed only at the end of the experiment, that is, after step 5 in figure 1.2. So, whilst the contest creates the feeling of inter-group competition on a second, unrelated domain, it does not affect the beliefs farmers may have about the levels of knowledge and cognitive ability of other farmers in their group.

Figure 1.2: Order of activities in the experiment



money between an in-group and an out-group recipient randomly drawn from the participants in the *following session* of the experiment.

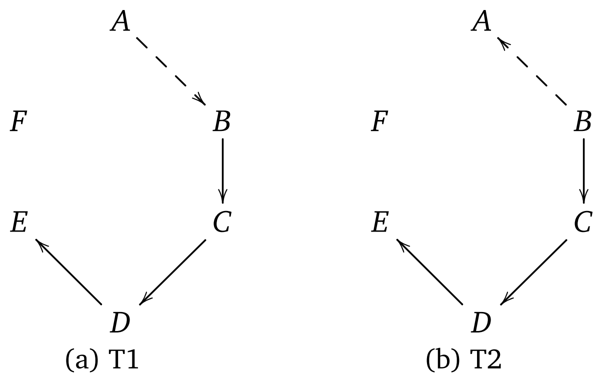
4. Players play the link formation game. They are given the instructions of the game, they answer a number of questions which test their understanding of these instructions[12], and they then play a trial of the game that lasts for seven turns. At the end of the trial, the game assistant randomly draws a participant and shows who would receive the prize if this was the actual game. After the trial, the actual link formation game is played.

5. Players are asked three questions about their expectations and beliefs and are then administered a short questionnaire, which collects information on socio-demographic variables and asks participants to explain the motivation behind their decisions in the game. At the end of this fifth phase, participants are informed of which team won the contest in agricultural knowledge and of the number of points each team has collected across sessions.

We rely on a between-subject design. We vary the direction of the flow of benefits associated with a link. In Treatment 1 (henceforth T1), players form links that let them *reach other individuals*. This means that if player A chooses player B, then player A will receive the monetary prize whenever B wins the prize, but not vice versa. In Treatment 2 (henceforth T2), links *let other players reach the player who proposes the link*. If A chooses a link with B, then B will receive the monetary prize whenever A wins it, but not vice versa. Figure 1.3 illustrates. Following our theoretical predictions

[12] The game assistant checks the answers and is instructed to give further explanations if more than one player makes more than one mistake. Hence these can be considered as a lower bound on the level of understanding of players.

Figure 1.3: Links in the two treatments



in the next section, as the network is updated, players are reminded of the number of individuals which each player reaches. In T2, players are also reminded of the number of individuals who reach a particular player. For example, in panel (b) of figure 1.3, player D reaches one player (player E). While two players (B and C) reach player D. In T1, players are only reminded about the fact that D reaches one player. In T2, they are reminded both about the fact that D reaches one player and the fact that two players reach D.[13]

We also vary the information about peer group membership available to players during the link formation game. This is cross-cut with T1 and T2. In a first set of treatments, which we call T1no and T2no, individuals have no information about peers' group affiliation. Hence their link formation decisions are by design unrelated to the groups formed at the beginning of the experiment. In a second set of treatments, called T1id and T2id, group identity is common knowledge, as players belonging to different groups are identified with different symbols on the public network map on the whiteboard.[14] We hence run four treatments, as shown in table 1.1.

Instructions are framed in terms of a salient example from the local context. The link formation game is presented as a game where one farmer will receive a valuable

[13]We give precise definitions of these concepts in the following section.
[14]Mango group players are identified with a circle. Pineapple group players are identified with a triangle.

Table 1.1: Summary of treatments

	No identity	Identity
Links take prize	T1no	T1id
Links give prize	T2no	T2id

piece of information about a new agricultural technology. The network determines who receives help from the farmer with the valuable information. In T1 the choice of a link is presented in terms of choosing who to approach for help to access the valuable information. In T2 the choice is about which other player one wants to help in case one accesses the valuable information. The groups are called the mango and the pineapple group. In the explanation they are associated with the producer groups which farmers typically form in the areas of the study.

In our design, group membership is randomly allocated. The original experiments in social psychology, on the other hand, rely on groups which are formed on the basis of trivial preferences.[15] While preference-based matching has the potential to increase the saliency of group membership, it also has two disadvantages. First, players' characteristics may be correlated to what the researchers considers as orthogonal preferences. Second, even if the chosen set of preferences is truly orthogonal, some players may believe otherwise. For example, a player may (erroneously) think that people with a certain preference in art or sport are smarter. In both cases, the effects of common knowledge of social identity would be confounded by those associated with correlated categories and beliefs.

We hence opt for a design which relies on random assignment to social groups and increases the saliency of group identity by means of the contest in agricultural knowledge. This task combines four desirable features: (i) it is linked to the overall framing of the experiment, (ii) it creates a feeling of competition between the two

[15]Notice however that in-group bias in allocation tasks is found also when group membership is determined by the flip of a coin (Tajfel, 1981)

groups on a domain, that of agricultural knowledge, which is distinct to the domain of monetary outcomes of the experiment, (iii) the relative position of the two groups in this second domain is only revealed after the link formation game has been played, (iv) every player can have a strong marginal impact on the group's outcome: if a player fails to answer one question correctly, the whole group fails to gain the point for that session. The idea of using contests to increase the salience of group identity has been successfully used before in experimental studies (Eckel and Grossman, 2005).

To ensure comparability and minimize noise during play, we follow a number of established practices in the lab-in-the-field literature. These include extensive piloting, simple standardized instructions that are read out to participants, double translation of all written material, and reliance on physical randomization devices (Barr and Genicot, 2008; Viceisza, 2012).

Possible Confounders

We take a number of steps against common confounders of experimental analyses.

Low understanding. We test players' understanding before the game starts. Subjects in T1no and T1id are asked 8 understanding questions, while subjects in T2no and T2id are asked 7 understanding questions. The questions test for understanding of the network map and of the incentives that result from the rules of the game. After the questions are asked, enumerators briefly check the answers and give further explanations on the points where players made mistakes. Hence these answers give a lower bound to the level of understanding of players in the game.

Figure 1.19 in the appendix presents the cumulative distribution of mistakes. In both T1 and T2 more than 50 percent of players made at most one mistake, and about 80 percent of players made at most two mistakes.

To further increase understanding, we also run a trial round of the link formation game before the main game is played.

Side-payments. Personal identity is not disclosed in the game and payments are disbursed privately. This decreases the possibility of side-payments. In particular, it decreases the possibility that links will be targeted towards individuals from whom side payments can be extracted more easily.

Wealth effects. Both the allocation task and the link-formation game are incentivised with monetary payments. In the allocation task individuals choose how to split a sum of money between two farmers in a future session of the experiment. The allocation decision does not affect the wealth of the decision maker. It also does not affect the wealth of the other farmers in his session. This rules out unintended influences across the two tasks created by endogenous shocks to players' wealth.

Experimenter demand effects. These arise when subjects, in an attempt to please the experimenter, respond to implicit cues embedded in the experimental design (Zizzo, 2010). For example, the fact that we disclose information about group identities may suggest to players that we expect them to use this information somehow. To minimise such concerns, we rely on a between-subjects design. These designs are thought to be less vulnerable to the demand effects critique (Zizzo, 2010). Furthermore, we refrain to give knowledge about players' experimental group identities in the instruction phase and in the trial round, to avoid making unintended suggestions about how we expect players to use the group membership information.

The visual reminder of the number of connections of each player can be a second source of experimenter demand effects. It could be argued that this feature biases the results in the direction of efficiency, as it increases the saliency of network statistics related to efficiency-enhancing strategies. Our aim in including this feature was to exclude the possibility that lack of familiarity with the graphical representation of the network would be driving departures from efficiency. Hence this design features is meant to “give the best shot" to the possibility of efficient networks. In the light of this design feature, our finding that network efficiency is significantly below potential becomes, if anything, more compelling.

1.3 Predictions

Our objective is to study the efficiency of the experimental networks formed by farmers. We hypothesise that farmers will choose their links on the basis of the structure of the network in predictable ways. In particular, we expect that farmers in T1 will play selfish best response, while farmers in T2 will either try to maximise the sum of the payoff of all players in the session, or the payoff of the least well-off player. We first present the ‘link-formation rules’ that follow from these preferences. Then, we simulate link-formation games where individuals follow the proposed rules and study the efficiency of the resulting networks. We show that when all farmers play selfish best response in T1, or when they maximise the sum of all payoffs in T2, the structure of the network converges to the cycle with high frequency within two rounds of the game.

Link-formation rules

Throughout the analysis we repeatedly use two concepts: network reach and in-reach. We define the *reach* of farmer j as the number of players whom farmer j observes directly or indirectly. The *in-reach* of farmer j is the number of players who directly or indirectly observe farmer j. The expected payoff of farmer j is a linear function of his reach in the final network. If we normalise the value of the prize to 1, the expected payoff of farmer j is simply given by:

$$\pi_j = \frac{reach_j + 1}{6} \tag{1.1}$$

Farmer j’s in-reach, on the other hand, determines the number of players who indirectly observe the information that is observed by farmer j. It is a measure of how far the information available to farmer j travels in the network. We present formal notation and definitions of these concepts in the appendix.

Following much of the existing literature, we assume myopic behaviour: a farmer considers the network which obtains after his link is added as the final network of the game. This rules out dynamic strategies based on threats, rewards, or signals. Recent research shows that the strategies played in experimental network formation games are often consistent with myopic best response (Conte et al., 2009).

In T1, a new link by player i affects his reach in the network, and hence his expected payoff. Only one link is permitted. Before this link is formed, player i has a reach of 0. A new link to player j allows player i to reach all the farmer whom player j reaches. Picking the partner with the highest reach maximises the reach of player i and, hence, his expected payoff.[16] The link-formation rule of a selfish player in T1 will hence be:

Rule 1. *Form a link with the player with the maximum reach.*

For this first rule and for all the rules that follow, we assume that in case of a tie between two or more potential partners, farmer i picks one of these partners at random. Furthermore, when we study whether player i's decisions conform to any of the link-formation rules, we exclude links to player i from the computation of the reach and in-reach of his potential partners. This is because, from player i's perspective, these links are redundant: they do not allow him to observe more information in T1, or to spread his own information further in T2.

In T2, new links to do not affect the reach of the player that forms them. A purely selfish player would be indifferent between forming and not forming a link in this treatment. If he forms a link, he would be indifferent about its consequences on the

[16] In the appendix we show this formally.

welfare of other players. However, a large body of evidence in experimental economics shows that individuals care about the payoffs of the other players in systematic, heterogenous ways (Charness and Rabin, 2002; Andreoni and Miller, 2002).[17] Following the literature on other-regarding preferences, we assume that players have a utility function that weights concerns for the player's own payoff and the payoff of all other players:

$$u_i = \pi_i + \gamma f(\pi_{-i}) \tag{1.2}$$

where we assume that there are n individuals in the set of all players N, and $\pi_{-i} = \{\pi_1, \pi_2, .., \pi_{i-1}, \pi_{i+1}, ..., \pi_n\}$. To advance further we have to make some assumptions about the shape of function f. We can explore two archetypal candidates. The first is:

$$u_i = \pi_i + \gamma \sum_{j \in N \setminus i} \pi_j \tag{1.3}$$

Utility function 1.3 expresses a concern for aggregate welfare. Charness and Rabin (2002) argue this is the model of social preferences with the highest predictive power for dictator game allocations. When a farmer forms a new link with player j in T2, he increases the reach of player j and of all the players who observe player j. Intuitively, the effect on aggregate welfare of a new link is proportional to the number of individuals who observe player j. This is player j's in-reach. In T2, the link that maximises the sum of individual payoffs should thus be a link with the player with the maximum in-reach.[18] We predict that a fraction of players in T2 have other-regarding preferences expressed by 1.3 and will thus play according to the following link-formation rule:

Rule 2. *Form a link with the player with the maximum in-reach.*

[17] Player i's strategy in T1 also has an impact on the payoffs of the individuals who reach player i. In future work, we will extend this section to include an the analysis of how other-regarding preferences may affect behaviour in T1.

[18] In the appendix we show this formally and explain one qualification that applies to links that create a small cycle.

A second possibility is that players care about the welfare of the player who is least well-off in the network. The literature in empirical social choice has documented this type of concern (Yaari and Bar-Hillel, 1984), which we can express using the following max-min utility function:

$$u_i = \pi_i + \gamma \min_{j \in N \setminus i} \pi_j \tag{1.4}$$

Utility function 1.4 is akin to the Rawlsian social welfare function which is a staple of social choice theory. The function is maximised by choosing the player with the lowest reach, which is the player with the lowest expected payoff. We predict that a fraction of players in T2 have other-regarding preferences expressed by 1.4 and will thus play according to the following link formation rule:

Rule 3. *Form a link with the player with the minimum reach.*

Notice that, depending on the structure of the network, the sets of players who satisfy rules 2 and 3 can be disjoint or overlapping. Figure 1.15 in the appendix shows an example where the two sets are disjoint: F has the minimum reach while A,B,C, and D all have the maximum in-reach.

A third model of social preferences is that of inequality aversion (Fehr and Schmidt, 1999). Under inequality aversion, a player feels guilt towards players with a lower expected payoff and envy towards players with a higher expected payoff. An inequality averse player in the first turn of a T2 session prefers not to form any link, as this would cause him to feel envy towards the player who benefits from the link. This prediction is virtually always falsified in our pilot and main data. We thus do not explore the predictions of the model of inequality aversion any further.

On the basis of the discussion above, we make the following prediction regarding individual decisions.

Prediction 1.1. *In T1 players form links with partners with the maximum reach. In T2 players form links either with partners with the maximum in-reach or with partners with the minimum reach.*

For ease of exposition we will sometimes refer to rule 1 and rule 2 as the 'efficiency-minded rules', as both of these rules follow from a desire to maximise payoff (either one's own, or that of the rest of the group). We will also refer to rule 3 as the 'Rawlsian rule', as it reflects the max-min logic of the Rawlsian social welfare function.

Network efficiency

We measure welfare as the sum of individual expected payoffs. The cycle network, where each player wins the prize for sure, maximises this sum and is the unique efficient network structure in our game. To compare the cycle network to other networks, we define a continuous measure of efficiency by taking the ratio between the average reach of players in a network and average reach of players in the cycle.

$$\text{Efficiency}_g = \frac{\frac{1}{n}\sum_{i=1}^{n} \text{reach}_i}{5} \tag{1.5}$$

The cycle network has efficiency 1 under this measure. All other possible networks have a level of efficiency that falls in the interval $[0,1)$. Our definition of efficiency rises monotonically with the sum of the expected payoffs in a network.

We simulate link-formation games where players follow the link-formation rules outlined above and we study the overall efficiency of the resulting networks. Our first set of simulations shows that when all players follow rule 1 in T1, average efficiency is about 96 percent. Figure 1.4 gives an example of how play in accordance with rule 1 achieves the cycle network within 2 rounds. Once the cycle network is reached, under rule 1 no player wants to rewire his link.

In a very small number of cases the process does not converge to the cycle. This is because players randomise between candidates of equal value, without consideration to the future order of play. This sometimes results in a situation where the player who can form the cycle network by re-wiring his link has already played his second turn. If we allow more rounds, the likelihood of this occurring in every round becomes very small. For example, in three rounds rule 1 achieves 99 percent efficiency.

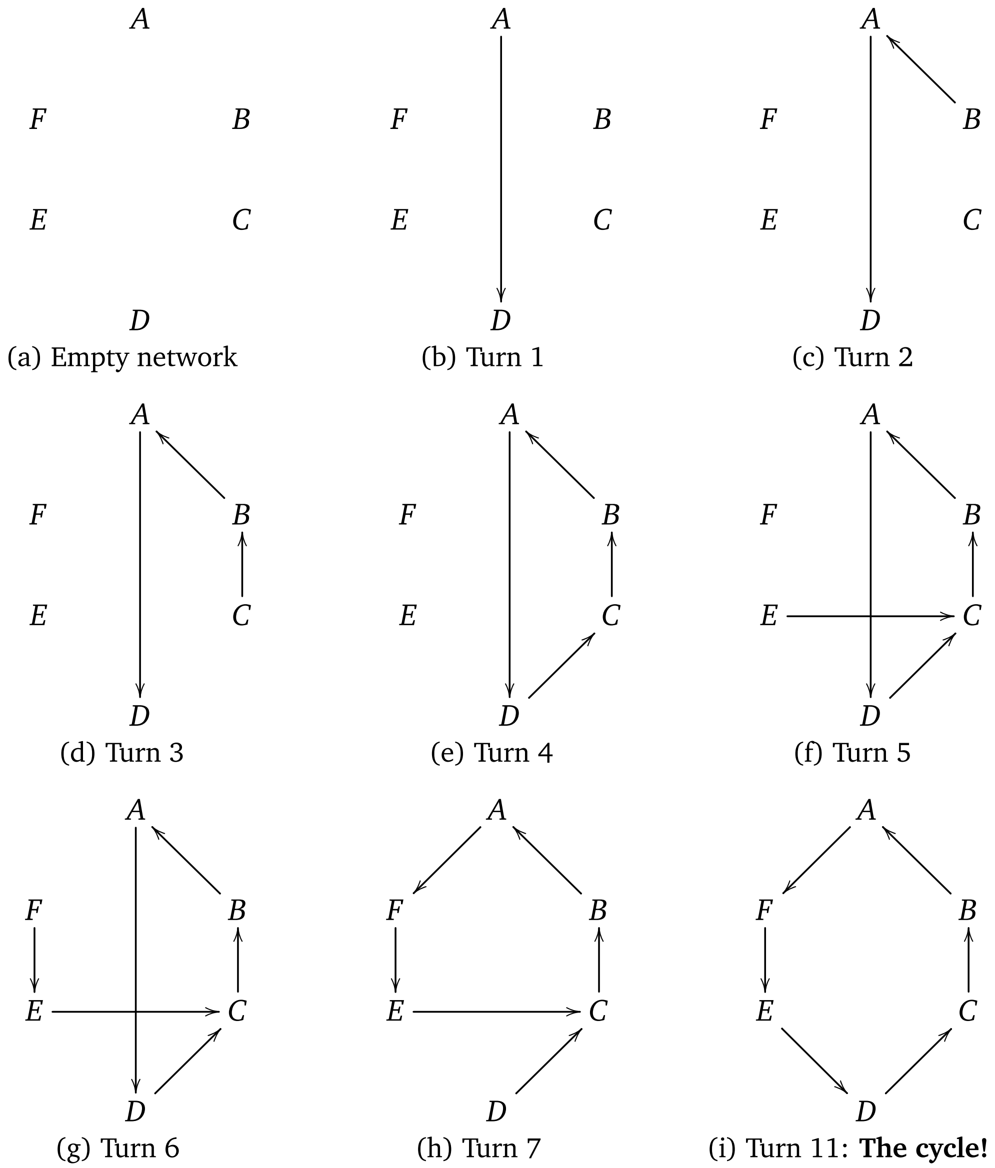
Our second set of simulations shows that when all players play according to rule 2 in T2 average efficiency is also about 96 percent.[19] When all players in T2 play according to rule 3, on the other hand, network efficiency is 67 percent. Figure 1.5 reports kernel density estimates and average efficiency for simulated sessions where play is in accordance, respectively, with rule 2, rule 3, and with a random link formation process. The random link formation process achieves average efficiency of about 52 percent.

We also study efficiency in sessions where a mix of rules is played. We simulate sessions where a fraction p of decisions follow rule 3, and a fraction 1-p of decisions follow rule 2. Results show that efficiency decreases monotonically with p in the interval between 96 and 67 percent. Figure 1.16 shows this graphically. We can thus formulate the following prediction on session-level efficiency.

Prediction 1.2. *Network efficiency in T1no is close to 96 percent. In T2no, it is between 96 and 67 percent.*

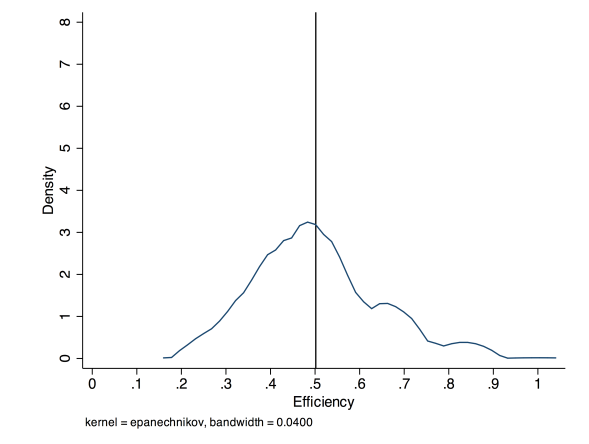
[19]This is not surprising, as rule 2 generates link-formation processes that are symmetrical with the respect to rule 1 in T1.

Figure 1.4: Network evolution under rule 1 in T1

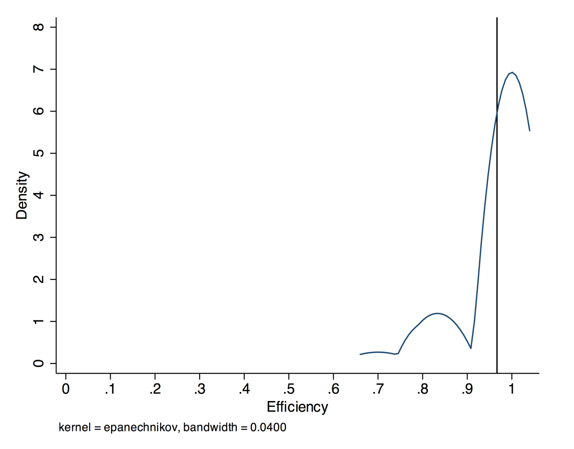


For ease of presentation, the order of play is assumed to be the order of the alphabet. All players in this simulation play according rule 1. Turns 7-11 are in the second round, where players rewire their existing link. Turns 8-10 are omitted because no rewiring takes place.

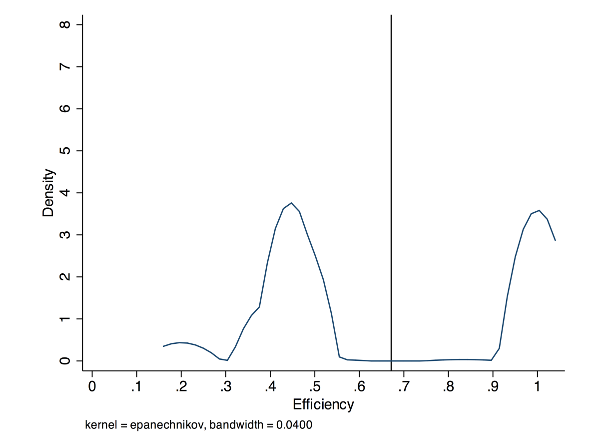
Figure 1.5: Average efficiency under different link-formation rules in T2



(a) Random



(b) Rule 2: Maximum in-reach



(c) Rule 3: Minimum reach

Each panel reports kernel density estimates of the distribution of the average reach, after 12 turns of play, for 500 simulated sessions. The vertical line indicates the mean value over all simulations. The rule used in each set of simulations is indicated below the panel.

Figure 1.17 in the appendix shows the evolution of efficiency under the different link-formation rules. In the second round, efficiency has no trend when all players play random or rule 3. However, when players play rule 2, average efficiency monotonically increases every turn.

The effect of group identity

In order to make predictions about behaviour when group identity is disclosed, we follow the seminal paper of Akerlof and Kranton (2000) and introduce a positive effect on utility which comes from following a social prescription:

$$u_i = \pi_i + \gamma f(\pi_{-i}) + P_i \tag{1.6}$$

P_i is equal to a positive constant c if player i follows the social prescription. In our game, for example, farmers may get positive utility whenever they use their link to connect with an in-group partner. This may describe the satisfaction arising from following a norm which states that links should be restricted to in-group partners. Whenever an in-group link generates an additional positive effect on utility of c we say that the individual is subject to a **norm of homophily**. Self-reports from our players are consistent with the existence of such norm. In a questionnaire administered after the game, 51 percent of players agree with the statement: 'In a game like this, one should only link to a player of his own group'. Furthermore, about 70 percent of players expect at least 3 of the other 5 individuals in the session to agree with the statement.

What will be the effect of disclosing group identity in our game? Suppose farmer i follows rule 2 in T2. For any positive value of c, whenever there are both in-group and out-group players who have maximum in-reach, farmer i will form a link with one of the in-group players. Before disclosure of group identity, he would have chosen randomly among the players with maximum in-reach. After disclosure, he can target his link to an in-group partner. The *frequency* of in-group links increases.

Now consider the case where there are no in-group players with maximum in-reach. If c is small, the positive utility from following the social norm will not compensate the loss in utility from failing to maximise the social objective. In this case, farmer i will act in the same way as he would have when group identity was not disclosed: he will form a link with an out-group player.

If c is high, on the other hand, the in-group player with highest reach within the set of in-group players may be preferred to the out-group player with the overall highest in-reach. In this case, after disclosure of group identity, farmer i will sometimes choose links that have a weaker effect on aggregate welfare. The *efficiency* of the network decreases.

The higher c, the larger the difference in in-reach a farmer is prepared to tolerate in order to conform to the social norm. In a set of simulations reported in figure 1.18 in the appendix, we show that when individuals play rule 2 and tolerate a difference in in-reach of 2 units, average network efficiency is 53 percent. When players play rule 3 and tolerate a difference in reach of 2 units, average efficiency is about 40 percent.[20] These considerations motivate a final prediction:

Prediction 1.3. *Disclosure of group identity generates networks characterised by (i) more in-group links and, depending on the magnitude of c, (ii) lower efficiency.*

[20] As a limit case, suppose farmers in T1 would never link to an out-group peer. If the play rule 1 among in-group partners, the network will converge to two small cycles with 3 players each. The average reach for this network structure is 2, corresponding to 40 percent efficiency.

Analysis

We analyse treatment effects using non-parametric two-sided Wilcoxon rank sum tests over session-level outcomes. We focus in particular on efficiency and the number of in-group links in the final network. The Wilcoxon rank sum test is a test of the null that the outcomes of the two treatments are drawn from same distribution. The alternative hypothesis is that either outcome is *stochastically greater*[21] than the other.

Further, we study individual decisions with dyadic regression analysis. In particular, we use models of the following form:

$$\text{link}_{ijr} = \alpha + \textbf{Network Position}_{jr}\beta + \mathbf{D}_{ij}\gamma + \delta \text{round}_r + e_{ijr} \tag{1.7}$$

The unit of observation is all i-j dyads in each session s. We observe each dyad once for each of the two rounds r. link_{ijr} is a dummy which takes value 1 if player i has chosen to establish a link with player j in round r. The matrix 'Network Position' contains variables which describe the network position of player j before player i's decision in round r. For T2, these include a dummy for having the minimum reach, and a dummy for having the maximum in-reach. For T1, they include a dummy for having the maximum reach. As a check, in T1, we also include a dummy for having the minimum value of in-reach. For robustness, we will also run specifications where we include the actual values of reach and in-reach.

To control for correlations between our variable of interests and the fixed positions of the players in the network map, we introduce a dummy variable for each possible pairing of map positions.[22] The matrix D_{ij} contains these variables. Furthermore, we

[21]For two populations A and B, A is stochastically greater than B if $Pr(a > b) > \frac{1}{2}$, where a and b are observations from population A and B, respectively. The two-sided Wilcoxon rank sum test sets the null of $Pr(a > b) = \frac{1}{2}$ against the alternative hypothesis that $Pr(a > b) \neq \frac{1}{2}$. The two-sided test is more conservative than the one-sided test.

[22]For example, see figure 1.4. From the perspective of player A, while B's position in the network is evolving, B remains A's closest neighbour in the visual representation of the set of players. This may make

control for round specific effects.

Model (1.7) will be estimated using OLS, correcting standard errors for arbitrary correlation at the session level. We can plausibly assume that there is no correlation between errors terms involving individuals in different sessions of the experiment. However, as explained, individuals are only allowed one link. This generates a correlation between error terms involving similar individuals within a session. For example, since a link to j precludes a link to k, $E[e_{ijr}e_{ikr}] \neq 0$. This inference problem is typical in dyadic regression analysis (Fafchamps and Gubert, 2007). We correct for within-session correlation in error terms using cluster-robust standard errors for inference.

Previous studies have shown that when the number of independent groups of observations is low, the cluster correction delivers downwardly biased standard errors (Cameron et al., 2008). Thus, when we run regressions with less than 40 clusters, we apply the wild bootstrap correction to p-values proposed by Cameron et al. (2008).

1.4 Data

We run our field experiment in the Indian state of Maharashtra. We randomly sample from a census list of all villages in 4 'talukas' (sub-districts) of the Pune and Satara districts.[23] Villages in these subdstricts are situated approximately 1,30 to 3h hours away from Pune. This is a similar distance to the district capital as that of the villages selected in the study of Banerjee et al. (2013). To reflect the large heterogoeneity in geographic conditions in this area of India, we choose two subdistricts which mostly comprise mountainous areas, and two subdistricts in the agricultural plains.

player A more likely to choose B than more distant players when A makes mistakes. To reach the cycle network, Player A may also choose an immediate neighbour as part of a coordination strategy which relies on physical proximity. A similar possibility is explored in Callander and Plott (2005). Alternatively, some positions in the map, for example A's position, may be visually more salient. If the position in the map is correlated with the network position of the player, regression analysis would suffer from omitted variable bias. We hence include position dummies for all possible directed dyads (AB, AC, .. , BA, BC, ..) to ensure that the effect of network position which we study in regression model (1.7) is not confounded by correlations with the initial position in the network map.

[23]We exclude large towns from the sampling universe.

We select study participants through door-to-door random sampling. Before reaching the village, our team is shown a Google Earth map of the village. On alternating days, the teams start sampling from the periphery of the village or from the center of the village.[24] We invite all male adult farmers who are encountered in the door-to-door visit until we have enough farmers to fill in all planned sessions.

Data collection took place between September and October 2013. In total, we run 81 sessions with 486 subjects. We run 20 sessions of T1no, T1id and T2id, and 21 sessions of T2no. In three of the sessions one participant left before the beginning of the link formation stage. This leaves us with 483 subjects, which correspond to 4800 dyads.[25] Table 1.2 summarises the number of observations we have for each treatment.

Table 1.2: Number of observations by treatment

Treatment	**Sessions**	**Players**	**Dyads**
T1no	20	120	1200
T1id	20	119	1180
T2no	21	126	1260
T2id	20	118	1160
Total	81	483	4800

At the end of the game, participants compile a short questionnaire. We hence have a small set of covariates.[26] Average age is 43 years. 95 percent of participants are Hindu, 72 percent do not belong to a scheduled caste, tribe or an other backward caste (OBC), 28 percent of them have completed high school. We also find that average total land holdings are about 4 hectares and average land cultivated is 3.6 hectares. On average, farmers report sharing information about agriculture on a regular basis with 11 other

[24]We identity the centre by asking village dwellers. This is typically a small square in front of the village temple.
[25]Each person in sessions with 6 individuals creates 10 dyads (5 per round). Each person in sessions with 5 individuals creates 8 dyads (4 per round).
[26]When participants fail to answer a question or report an illegible script, we code a missing value. This explains the changing number of observations in table 1.3.

farmers.

Table 1.3: Summary statistics: individual covariates

Variable	**Obs**	**Mean**	**Std. Dev.**	Min	Max
Age	479	43.41	12.96	22	85
Hindu	457	.95	.22	0	1
Non backward caste	433	.72	.45	0	1
Completed High School	466	.28	.45	0	1
Land Owned	475	4.07	4.67	.1	50
Land Cultivated	470	3.6	4.18	.1	45
Information network size	428	10.9	8.94	1	60

From session 9 onwards,[27] we ask each farmers whether he knows each of the other 5 participants and on how many days of the last months he has had a conversation with them. The density of the within-session networks we record is very high: 87 percent of participants know every other farmer in their session. On average, farmers speak on 13.5 days in a month with each of the participant they know.

Table 1.4: Summary statistics: session networks

Variable	**Obs**	**Mean**	**Std. Dev.**	Min	Max
Out-degree	438	4.78	.76	0	5
Average days spoken with known peers	430	13.44	9.56	0	30

Degree refers to the reported number of other participants that a player knows. For each known farmer j, we ask farmer i on how many days of the last month he has spoken to farmer j. We compute the average of this variable across all farmers j for each farmer i. In the second row of the table, we average this variable over all farmers i. 8 farmers do not know anybody in the network, so we do not compute this variable for them.

In tables 1.6 to 1.9 in the appendix, we present some regressions that test for covariates balance across treatments. We cannot find any statistically significant difference in average characteristics across treatments.

1.5 Results

We organise our discussion around four key results.

[27] This means that we ask this question to 438 individuals in 73 sessions

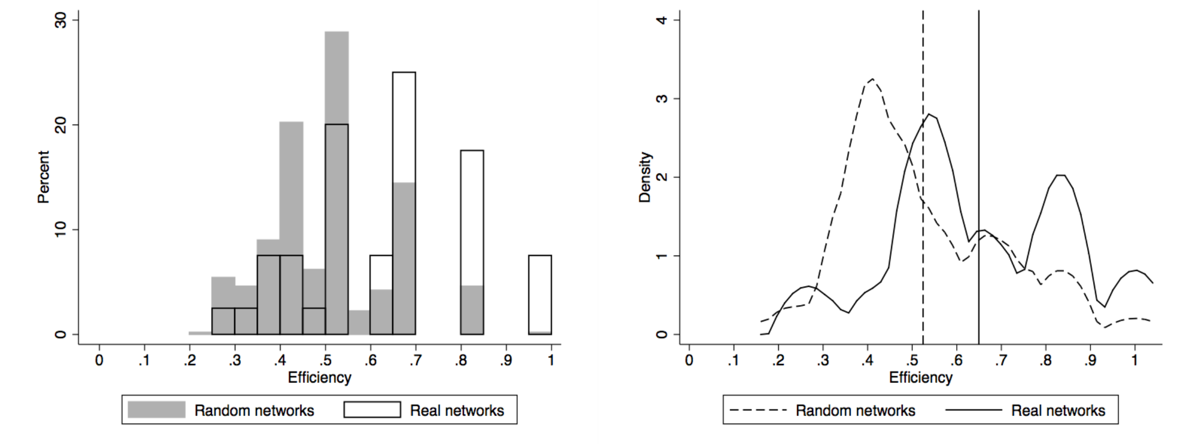
We first investigate overall efficiency. Table 1.10 summarises treatment-level averages of reach and efficiency for the final network of the game. We pool all sessions with no knowledge of group identity together and compare the distribution of average session efficiency to two simulated benchmarks: the distribution of average session efficiency which would obtain if individuals chose their links at random, and the distribution under 'efficiency-minded' link formation (rule 1 in T1, rule 2 in T2). We obtain the following result, which is represented graphically in figure 1.6:

Result 1.1. *Network efficiency in T1no and T2no is 65 percent. This is 31 percentage points below average efficiency under 'efficiency-minded' link-formation (rule 1 in T1, rule 2 in T2), and 13 points above average efficiency under random network formation. Both differences are statistically significant.*

The efficiency of the experimental networks is 31 percentage points below the average level achieved by the 'efficiency-minded' link-formation rules. A Wilcoxon rank-sum the tests confirms that the difference between the distribution of network efficiency in our data and the simulated distribution is statistically significant at the 1 percent level ($Z = 12.08$, $p < .001$). On the other hand, the efficiency of the experimental networks is higher, by a significant 13 percentage points, than average efficiency achieved with random play ($Z = 4.62$, $p < .001$).

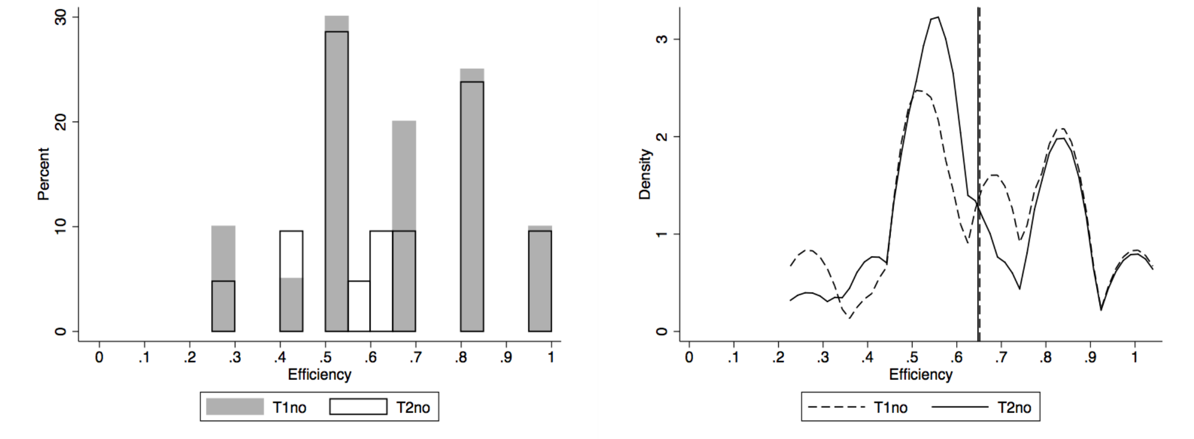
The direction of the flow of benefits associated with the links does not affect average efficiency. Hence the result above is not driven a by lower efficiency in the T2 treatment. Average efficiency across the T1no and T2no treatments is in fact very similar. A Wilcoxon rank sum test cannot reject the null that the outcomes of the two treatments are drawn from the same distribution ($Z = -.11$, $p = .91$). Figure 1.7 presents this result graphically. We predicted that efficiency in T2 would vary in the range between 57 and 96 percent, and that efficiency in T1 would be 96 percent. The

Figure 1.6: Efficiency in no-identity sessions and in random networks



(a) Histogram (b) Kernel density estimate and means

Figure 1.7: Efficiency in T1no and T2no sessions



(a) Histogram (b) Kernel density estimate and means

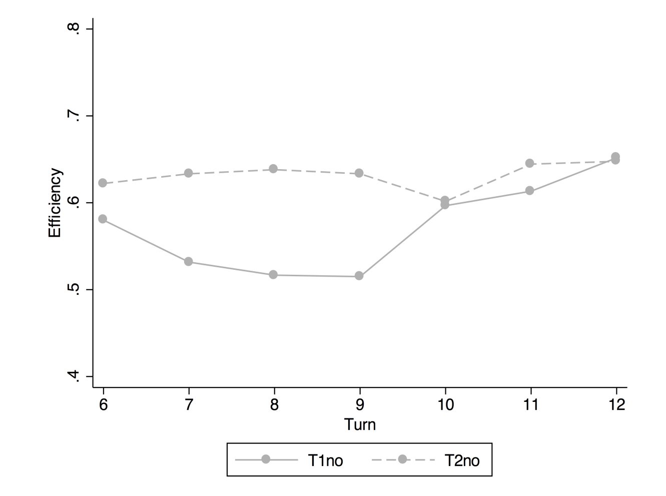
prediction for T1 is clearly rejected.

Result 1.2. *Efficiency in T2no sessions is not significantly different from efficiency in T1no sessions.*

It is important to note that low efficiency is not an artefact of truncation at 12 turns: efficiency has no monotonic updward trend in either T1no or T2no, and efficiency at turn 12 is only a few percentage points higher than at turn 6. Figure 1.8 illustrates.

Falk and Kosfeld (2012), on the other hand, document strong learning dynamics and positive efficiency trends in their experiment.

Figure 1.8: Time series of effciency in T1no and T2no, turns 6-12



Efficiency is lower than what any of the rules we specified would have achieved. Further, it is not different across the two treatments. How do these results come about? We can answer this question estimating dyadic regression model (1.7). Results are reported in table 1.5.

As hypothesised, we find that in T1no ties are directed towards players with the maximum reach. In T2no, on the other hand, ties are directed towards players with the maximum in-reach and the minimum reach. The effects are highly statistically significant and of a meaningful magnitude. In T1no, player i is 13 percentage points more likely to choose a player with the maximum reach. This amounts to a 40 percent increase over the probability of choosing a player who does not have the maximum reach. In T2no, player i is 11 percentage points more likely to choose a player with maximum in-reach and 7 percentage points more likely to pick a player with minimum reach. A Wald test cannot reject the equality of these two coefficients.

Table 1.5: Dyadic linear probability model (1.7)

	(1)	(2)	(3)	(4)
Panel a				
max reach$_j$	.132 (.001)***	.130 (.001)***		
min in-reach$_j$	.018 (.461)	.016 (.627)		
max in-reach$_j$			.111 (.006)***	.120 (.002)***
min reach$_j$			.073 (.078)*	.066 (.142)
Const.	.323 (.002)***	.367 (.002)***	.192 (.002)***	.218 (.028)**
Panel b				
max reach$_j$ = min in-reach$_j$	10.34 (.004)***	10.81 (.004)***		
max in-reach$_j$ = min reach$_j$			0.57 (.459)	1.12 (.304)
Obs.	1200	910	1260	940
Sample	T1no	T1no	T2no	T2no
Cluster N	20	20	21	21
Controls		✔		✔

Dyadic OLS regression. Dependent variable is a dummy which takes a value of one if i chose to establish a link with j. Each regression contains controls for the round and for each possible pairing of map positions. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Regressions in columns 2 and 4 include controls for age, land owned, land cultivated, number of contacts in real information networks, number of mistakes in the initial understanding questions and dummies for having completed secondary education, for being Hindu, and for not belonging to a backward caste. Standard errors are corrected for clustering at session level. P-values obtained with wild bootstrap-t procedure reported in parentheses. Panel b reports the F statistics (and p value in parenthesis) for a Wald test of the equality of coefficients.

We confirm the robustness of these results by running a specification that substitutes the dummies with the values of reach and in-reach. This allows players to make mistakes, while requiring larger mistakes to be less likely than smaller mistakes. Table 1.12 in the appendix reports the estimates. Results are significant and of a larger magnitude. In T1no, for example, player i is 22 percentage points more likely to choose a player with a reach of 4 than a player with a reach of 0. Given that a player with 0 reach has a probability of being chosen of 27.1 percent, this amounts to an increase by 81 percent.

We summarise this analysis in the following result, which supports prediction 1.1:

Result 1.3. *In T1no, links to farmers who have the maximum reach in the network are significantly more likely to be formed than other links. In T2no, links to farmers who have the maximum in-reach and to farmers who have the minimum reach are significantly more likely to be formed than other links.*

Table 1.12 shows a further significant effect: in T1no player i is more likely to establish a link with a player with a lower in-reach. A caveat is in order, as, in the previous specification, when we include a dummy for whether an individual has the minimum in-reach we report a positive, but small and insignificant coefficient. This suggests that the effect of in-reach in T1 is probably not substantial. This result is also difficult to explain within our theoretical framework. One possibility is that links carry social value for the person receiving the link proposal. Individuals who choose peers with a low in-reach in T1no could thus be targeting the players who have accumulated the minimum social value in the game so far. We cannot provide a direct test for this interpretation.[28] We have however some qualitative evidence in support of it. In the post-play questionnaire farmers are asked the following question: "Do you think that choosing a farmer from your own group is a way of showing respect to him?". 51 percent of farmers answer yes to this question. This is consistent with the view that links carry social value, but represents by no means a full-fledged test.

We define an additional link-formation rule to describe this behaviour:

Rule 4. *Form a link with the player with the minimum in-reach.*

From now on, we will refer to rule 3 and rule 4 jointly as the 'Rawlsian' rules

While result 1.3 is in line with prediction 1.1, not all decisions are consistent with the archetypal rules we have proposed. This becomes apparent when we look at the relative frequency of decisions consistent with the various rules. In T1no, 51 percent of decisions are consistent with rule 1 and 63 percent with rule 4. In T2no, 56 percent of decisions are consistent with rule 2 and 68 percent with rule 3.

[28]Furthermore, strictly speaking, this motive would lead to a rule targeting the player with the minimum in-degree, unless indirect connections also carry social value.

This exercise, however, poses two problems. First, there are often multiple individuals who satisfy a particular rule. Hence, rules satisfied by a larger number of candidates are selected more frequently when individuals choose randomly or make random mistakes. This makes it difficult to interpret frequencies and to compare different rules. To address this, we calculate the probability of observing a decision consistent with a particular rule when farmers choose links at random. We then calculate a confidence interval around the frequency with which we observe decisions consistent with the same rule in the data. Finally, we check whether the probability of choosing such rule under random play lies below the confidence interval. If so, we are observing a rule being chosen significantly more often than under random play.

Second, the sets of individuals satisfying different rules often overlap. In a line network, for example, the first individual has both the maximum reach and the minimum in-reach. The last individual, on the other hand, has the maximum in-reach and minimum reach. This complicates comparison across rules. To study the extent of the problem we investigate the frequency of overlaps. For each turn, we check whether the sets of potential partners who satisfy the 'efficiency-minded' and 'Rawlsian' rules are disjoint, partially overlapping, or fully overlapping.[29] Results are presented in panel (a) of figure 1.21 in the appendix. Overlaps are very frequent. We will hence repeat the analysis restricting the data to turns where the best response sets are not fully overlapping.

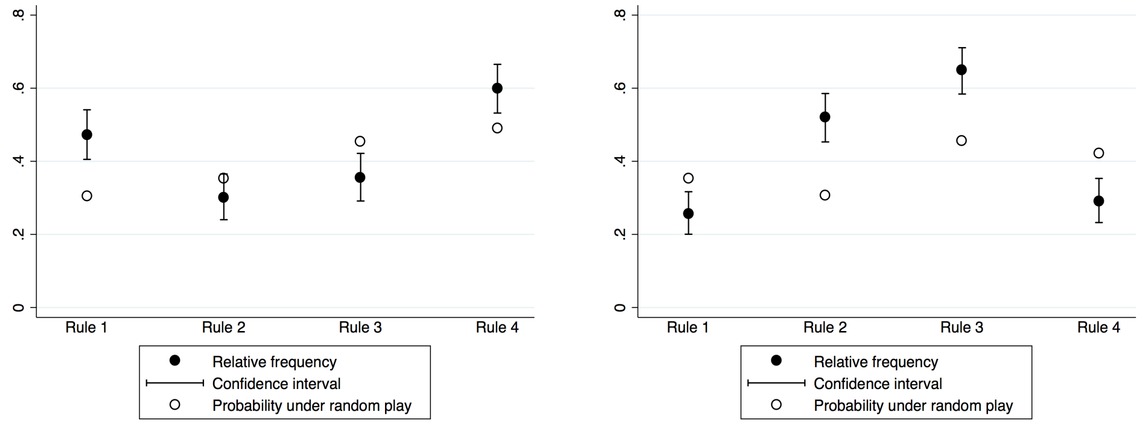
Figure 1.9 presents the analysis for the whole sample. In T1no, both decisions consistent with rule 1 and decisions consistent with rule 4 are observed significantly more often than under random play. While decisions consistent with rule 4 are more

[29]Consider turn t when farmer i has to play. Let BR_t^1, BR_t^2, BR_t^3 and BR_t^4 be the sets of players who, from the point of view of farmer i, satisfy link-formation rules 1,2,3 and 4, respectively. For T1, we focus on BR_t^1 and BR_t^4 and define three mutually exclusive cases:

1. *Fully overlapping*: $BR_t^1 \cap BR_t^4 = BR_t^1 = BR_t^4$.
2. *Disjoint*: $BR_t^1 \cap BR_t^4 = \emptyset$.
3. *Partially overlapping*: not disjoint and not fully overlapping.

For T2, we focus on BR_t^2 and BR_t^3 and similarly define the three cases.

Figure 1.9: Relative frequency of decisions consistent with the hypothesised rules



(a) T1no (b) T2no

frequent, they also have a higher probability of occurring under random play. In T2no, rules 2 and 3 are also observed significantly more often than under random play. These results are basically unaffected if we restrict the analysis to turns where the 'efficiency-minded' and 'Rawlsian' best response sets are not fully overlapping. As a further robustness check, we also consider the two decisions taken by each player jointly and find that pairs of decisions consistent with a single archetypal rule occur more often than under random play. Figures 1.23 and 1.24 in the appendix illustrates.

The frequency of 'efficiency-minded' and 'Rawlsian' links is similar in the T1no and T2no treatments. This explains why these two treatments have similar levels of network efficiency. In both T1no and T2no, about 70 percent of decisions are consistent at least one of the archetypal rules. Panel (b) of figure 1.21 shows this. The majority of these decisions are consistent with both rules, while about 16 percent of links in both treatments satisfy only the 'Rawlsian' rule.

What about the remaining 30 percent of links that are not consistent with either rule? We explore two possible additional link-formation rules:

Rule 5. *Choose a link to the player who has been chosen by most other players in the* *current network*[30]

Rule 6. *Choose a link to the player who has chosen you in a previous round* [31]

Regression analysis reported in table 1.13 in the appendix suggests that, in general, rules 5 and 6 do not significantly predict link-formation decisions. Nevertheless, links consistent with rule 5 are observed frequently: 66 percent of decisions that do not follow the 'efficiency-minded' or the 'Rawlsian' rule target the 'most popular' player in the network instead. Reciprocal links are not as common: they occur only in 18 percent of decisions that not consistent with the archetypal rules. Figures 1.22 illustrates.

Simulation analysis shows that the largest efficiency gains can be achieved by reducing the proportion of links that are targeted to the 'most popular' player, as opposed to reducing the proportion of 'Rawlsian' links. We simulate a link formation process where 54 percent of decisions are consistent with rule 1, 16 percent with the rule 4 and the remaining 30 percent with the 'most popular' player rule.[32] We then switch increasing proportions of decisions assigned to follow rule 4 to rule 1, keeping the proportion of rule 5 decisions fixed. We repeat the same exercise for rule 5: we switch increasing proportions of decisions assigned to follow rule 5 to rule 1, keeping the proportion of rule 4 decisions fixed. The results are stark: switching all rule 5 decisions to the rule 1 delivers an efficiency gain of 25 percentage points, while an equivalent reduction of 'Rawlsian' decisions results only in a 5 percentage points increase.[33] Figure 1.10 illustrates. In figure 1.25 in the appendix we present the same simulation, with a

[30]We refer to this player as the 'most popular' player. In T1, this corresponds to the player with the maximum in-degree. This is the player that is *directly observed* by the highest number of farmers. In T2 to the player with the maximum degree. This is the player that is *directly observing* the highest number of farmers. Past links that have been rewired are not included in the computations.

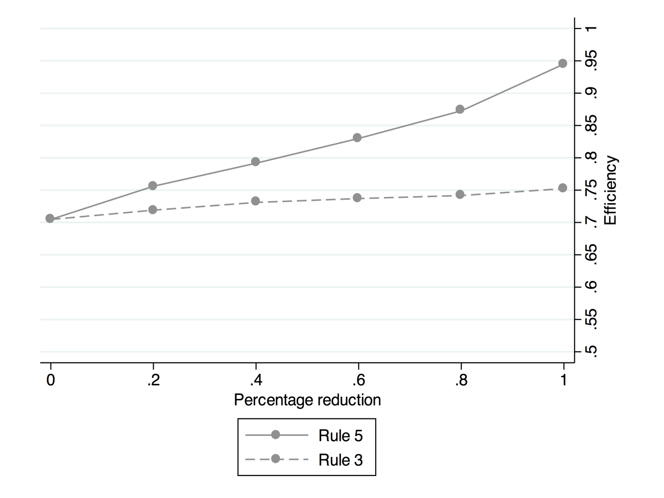
[31]We refer to this as the 'reciprocal' link-formation rule.

[32]This reflects the decisions in our data, with two simplifying assumptions: (i) all decisions that are consistent with both rule 1 and rule 4 are assumed to follow the rule 1, (ii) all decisions that are not consistent with the archetypal rules are assumed to follow rule 5.

[33]We also know from the simulations reported in figure 1.16 what would happen if switch all rule 5 decisions to rule 3. This thought experiment corresponds to a simulation where 46 percent of decisions follow rule 3 and 54 percent follow rule 1. Figure 1.16 shows that network efficiency in such scenario would be above 90 percent, which corresponds to 20 percentage points gain.

different assumption about the baseline proportion of decisions following the various rules. Qualitatively, results are not affected.

Figure 1.10: Efficiency simulations



Note. In the baseline simulation 54 percent of decisions follow rule 1, 16 percent follow rule 4, and 30 percent follow rule 5. Each point in the graph represents average efficiency over 100 repetitions of the link formation game.

We next turn to the the effect of social identity. We start by showing two pieces of evidence which suggest that our group assignment procedure creates salient groups and that subjects believe that a norm prescribing restriction of links to the in-group applies to our link formation game.

Figure 1.26 in the appendix shows results from the initial allocation task where a player has to divide a sum of money between an in-group and an out-group partner in the following session. The modal allocation in this task is skewed towards the in-group partner. Overall 54 percent of individuals show such in-group bias, while 30 percent choose equal allocations. This shows that the saliency of our experimental groups is sufficient to modify individuals' allocations. Second, we investigate whether individuals perceive that a norm of homophily applies to behaviour in our game. In all treatments, we ask participants at the end of the game whether they think that a player in this game 'should' only link to in-group peers.[34] 57 percent of players answer

[34]Players in T1no and T2no also answer this question, albeit information about group affiliations was not

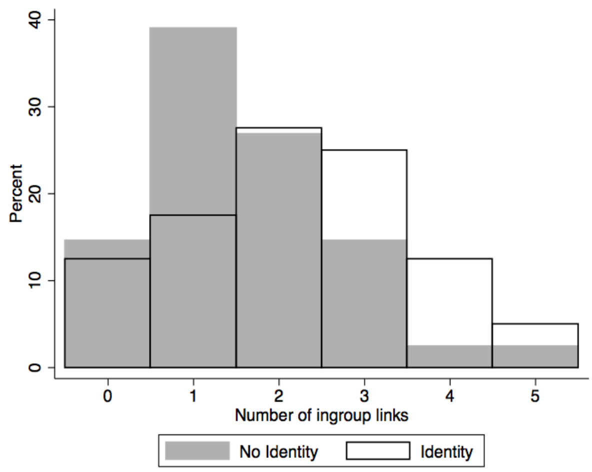
yes to this question. Furthermore, about 70 percent of players expect at least 3 of the other players (the majority) to answer yes. Table 1.11 and figure 1.27 document this. Somewhat in contrast to this, only 38 percent of players expect the last player of the game to choose an in-group link.

Our main result on the identity treatments is the following:

Result 1.4. *In treatments where group identity is disclosed in-group links occur more frequently than in treatment with no knowledge of group identity, while network efficiency does not decrease.*

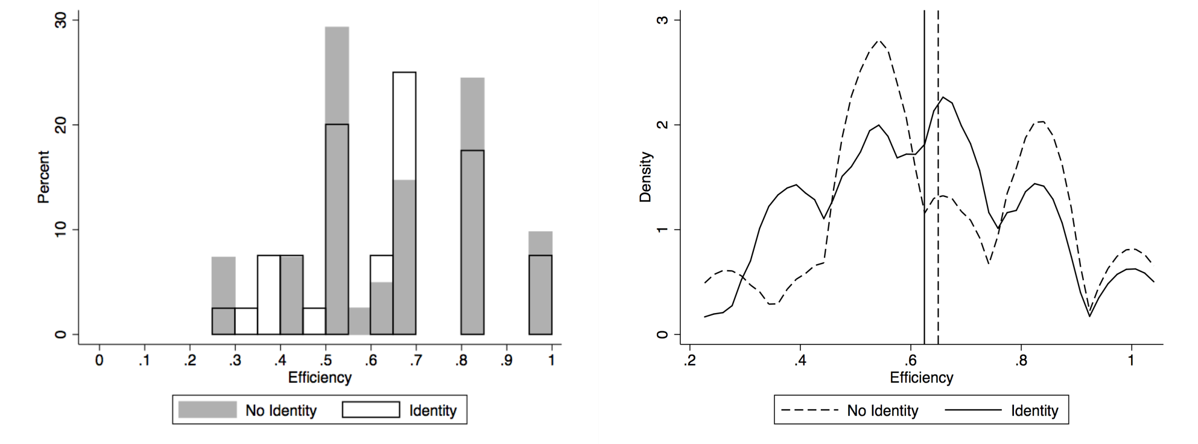
This result confirms the first part of prediction 1.3. First, in-group links increase. Figure 1.11 shows a histogram of the number of in-group links in the final network for sessions where group identity is disclosed and session where it is not. The distribution clearly shifts to the right. A Wilcoxon rank-sum test confirms this difference is significant at the 5 percent level (Z= 2.23, p= .02).

Figure 1.11: In-group links in identity and no-identity treatments



disclosed in these treatments during the link formation game.

Figure 1.12: Efficiency in identity and no-identity sessions



(a) Histogram (b) Kernel density estimate and means

Note: Only links in the final network are considered. 'No identity' sessions include T1no and T2no. 'Identity sessions' include T1id and T2id.

Second, we cannot detect a systematic effect of disclosing players' group identity on session level efficiency. Mean efficiency decreases to 58 percent in T1 and essentially stays put in T2. A Wilcoxon rank-sum test cannot reject the equality of the distributions (Z= -0.51, p= .61). This is documented graphically in figure 1.12.

In order to investigate how disclosure of group identity affects link formation, we run linear probability models of the following form:

$$x_{dis} = \alpha + \beta_1 \text{Identity Session}_s + e_{dis} \quad (1.8)$$

$$x_{dis} = \alpha + \beta_1 \text{Identity Session}_s + \beta_2 z_{is} + \beta_3 \left(z_{is} * \text{Identity Session}_s \right) + e_{dis} \quad (1.9)$$

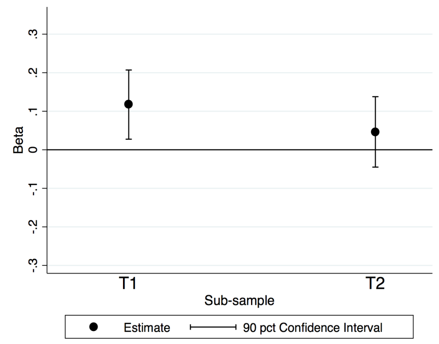
x_{dis} is an indicator variable which takes the value of 1 if decision d by player i in session s has a certain characteristic. We perform the analysis with three definitions of x_{dis}: whether decision d is a link towards an in-group player, whether decision d is consistent with the 'efficiency-minded' rule, and whether decision d is consistent with the 'Rawlsian' rule.[35] In model (1.9), we also study whether the effect of being in an identity sections is stronger for certain types of players, for example, players who have allocated more to the in-group partner in the initial allocation task. Standard errors are clustered at the session level. Figure 1.13 shows graphically the coefficient estimates, while full regression tables are available in the appendix.

In-group links increase in both treatments. However, the effect is significant only for T1. In this condition, links to an in-group player are about 11 percentage points more likely once group identity is disclosed. This corresponds to a 40 percent increase in the probability of an in-group link. For T2 treatments the effect drops to 5 percentage points and is not significant.

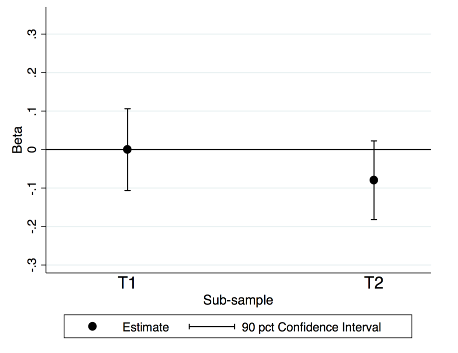
The frequency of 'efficiency-minded' or 'Rawlsian' links is unaffected by the disclosure of group identity in T1. In section 3, we argued that a farmer who derives a limited, but positive benefit from following the social norm may not be prepared to sacrifice his objective in order to conform to the norm. However, when there are both in-group and out-group partners who satisfy his preferred link-formation rule, he will strictly prefer a link to an in-group partner. In this scenario, disclosure of group identity will be associated to: (i) an increase in in-group links, (ii) no change in the relative frequency of 'efficiency-minded' and 'Rawlsian' links, and (iii) an increase in the proportion of, say, 'efficiency-minded' links that are directed towards an in-group partner. We have already presented evidence consistent with effects (i) and (ii). We test for effect (iii) by restricting the sample to 'efficiency-minded' decisions in T1 and T2 and running again model (1.8), with the in-group link dummy as dependent variable. We estimate a significant coefficient of 13.8 for T1. This can be interpreted as the

[35] As explained above, the 'efficiency-minded' rules are rule 1 for T1 and rule 2 for T2. The 'Rawlsian' rules are rule 4 for T1, and rule 3 for T2.

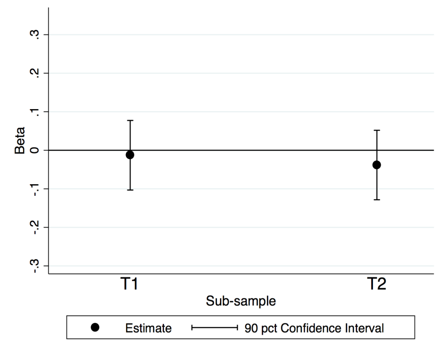
Figure 1.13: Linear probability model (1.8): coefficient estimates



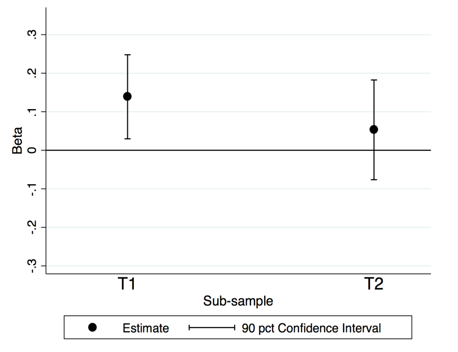
(a) x_{is}=: in-group link



(b) x_{is}= efficiency-minded link



(c) x_{is}= Rawlsian link



(d) x_{is}= in-group link, restricted sample.

Coefficients estimates from linear probability model (1.8). The dependent variable is indicated below each graph. The regression in graph (d) is run over a sample restricted to include only 'efficiency-minded' links. Standard error are clustered at session level. Full regression results are reported in tables 1.14, 1.15, 1.16, and 1.17.

percentage point increase in the probability that an 'efficiency-minded' link is directed to an in-group farmer.

Decisions consistent with the two archetypal rules are, on the other hand, observed less frequently in T2. Links towards players with maximum in-reach drop by 9 percentage points (a 17 percent fall with respect to the baseline probability of such links in T2no). The effect is however only significant at the 15 percent level. Links towards players with minimum reach also decrease by an insignificant 4 percentage points.

These results suggest that disclosure of group identity *may* operate through different mechanisms in the two treatments. We are unable to shed more light on these mechanisms through estimation of model (1.9): the effect of group identity disclosure on the likelihood of choosing an in-group link is not stronger for individuals who show in-group bias in the allocation task, who agree with the norm of homophily, or who expect more peers to agree with the norm of homophily.[36]

We have not said much about understanding so far. The last set of results we report shows that the likelihood of choosing a link consistent with any of the rules we have discussed is generally not correlated with the number of correct answers players give in the initial understanding questions. We show this using the following regression model:

$$\begin{aligned} x_{dis} =& \alpha + \beta_1 \text{Identity Session}_s \\ & + \beta_2 \left(understanding_{is} * \text{Identity Session=}0_s \right) \\ & + \beta_3 \left(understanding_{is} * \text{Identity Session=}1_s \right) + e_{dis} \end{aligned} \tag{1.10}$$

$understanding_{is}$ is the z-score of the number of correct answers players give in the initial understanding questions. β_2 captures whether high-understanding subjects are more likely to choose links consistent with strategy x in sessions where group identity is not disclosed. β_3 measures whether high-understanding subjects are more likely to choose links consistent with strategy x in sessions where group identity is disclosed. Tables 1.20 and reports results of separate estimations of model (1.10) for T1 and T2. The only significant result is that in sessions where group identity is disclosed, players with a higher understanding z-score are less likely to link to the player who has been chosen most frequently. The coefficient is only significant at the 10 percent level and small in magnitude: a standard deviation increase in understanding is associated with a decrease in the likelihood of a link towards the 'most popular' player of 4 percentage point, corresponding to 10 percent reduction in this probability.

[36] In future work, we plan to attempt estimation of a structural mixture model of link selection, which allows for different link formation objectives in the population.

1.6 Conclusion

Social networks play an important role in the diffusion of innovations such as new agricultural technologies, health schemes or financial products. Theoretical models show that in games of unilateral, one way flow link formation, myopic, selfish individuals can converge to efficient networks after repeated play (Bala and Goyal, 2000). We offer the first experimental test of this prediction for a non-western population- a sample of farming communities in rural India. This is a policy-relevant setting: interest for new, cost-effective intervention designs that promote the diffusion of agricultural technology is currently high in India. We make a second contribution to the literature by exploring how a pervasive feature of the social world, group categorisation, affects the way networks are formed.

We find that the efficiency of the networks formed in our experiment is significantly and substantially lower than the level of efficiency which myopic selfish play would have achieved. While many farmers choose welfare-enhancing links, large efficiency losses come about because a minority group of farmers chooses to link with the 'most popular' farmer in the network. When information about group membership is disclosed, more in-group links are formed, but networks do not become significantly less efficient.

Inefficiently structured networks can limit the diffusion of information about new technologies and hence their adoption. This creates a rationale for development policies that support the diffusion of socially beneficial innovations. Interventions can bypass social structures altogether and rely instead on modern technologies. Recent trials show that agronomic information transmitted via SMS, phone lines and voice messages can be effective at increasing yields, and discouraging the use of inefficient pesticides (Cole and Fernando, 2012; Casaburi et al., 2014). Alternatively, interventions can try to strengthen peer-to-peer transmission by incentivising farmers to share information (BenYishay and Mobarak, 2014) or by fostering the creation of new links

(Vasilaky and Leonard, 2013).

Future work can use artefactual link-formation field experiments to inform the development of diffusion policy in two ways. First, it can further explore link-formation heuristics in specific settings. Program design should ensure compatibility with those heuristics. For example, where individuals preferentially attach to the 'most popular' farmers, peer-to-peer extension programs can rely on a few, prominent injection points. Where links are reciprocal and less centralised, different models may be required. Second, experimental designs can help develop our understanding of how social features impact network formation and potentially limit peer-to-peer diffusion. Further study of the effects of group categorisation using natural groups is required. Within-group status differentiation offers a second important avenue for exploration.

Appendix

A.1 Formal derivation of the rules

A.1.1 Notation

We define some basic notation following Goyal (2007). Let N=(1,2,..,n) be the set of players. In T1, each player i chooses a (pure) strategy $g_i = (g_{i1}, g_{i2}, ..., g_{ii-1}, g_{ii+1}, ..., g_{in})$[37], which is a vector of directed links $g_{ij} \in \{0,1\}$. In T2, on the other hand, every player chooses a strategy $g_i = (g_{1i}, g_{2i}, ..., g_{i-1i}, g_{i+1i}, ..., g_{ni})$, a vector of directed links $g_{ji} \in \{0,1\}$. Let Γ_i be the set of possible values of g_i.[38] $\Gamma = \Gamma_1 \times \Gamma_2 \times ... \times \Gamma_n$ is the set of all possible combinations of player strategies. The vector of player strategies $g = (g_1, g_2, ..., g_n)$, drawn from Γ, can be represented as a directed network. $g + ij$ is the network obtained from adding the link $g_{ij} = 1$ to network g.

In our game player i receives the prize if he is the winner of the prize lottery, or if he is connected to the winner via a *path of links*. A path from player i to player j is a set of links such that: $g_{iy} = g_{yw} = ... = g_{zj} = 1$. A direct link is a path of length 1. The notation $i \to^g j$ indicates that in network g there is a path from i to j. If the path $i \to^g j$ exists, we say that play i *reaches* player j network g. In this case, player i is assigned the prize whenever player j is assigned the prize. A path $i \to^g j$, on the other hand, has no implication on whether player j is assigned the prize when player i is assigned the prize.

We need to introduce two crucial concepts for our analysis. First, let $N_j(g) = \{k \in N | j \to^g k\}$ and $\mu_j(g) = |N_j(g)|$. $\mu_j(g)$ represents the number of players whom player j reaches in network g. Sometimes we want to exclude from the count the path from player j to player i. Let $N_{ji}(g) = \{k \in N \setminus i | j \to^g k\}$ and $\mu_{ji}(g) = |N_{ji}(g)|$. $\mu_{ji}(g)$ is ***the number of players whom player j reaches*** in network g, *excluding player i*. We call

[37] Link from player i to player i are ruled out.

[38] In both T1 and T2, we impose that at most one link can be equal to 1. Thus, there are n possible values of of g_i: n-1 possible links plus the strategy of establishing no links at all.

$\mu_{ji}(g)$ the **reach** of player j in network g.

Second, let $N_{-j}(g) = \{k \in N | k \to^g j\}$ and $\mu_{-j}(g) = |N_{-j}(g)|$. $\mu_{-j}(g)$ represents the number of players who reach player j in network g. Again, we sometimes need to exclude the path from player i to player j. Let $N_{-ji}(g) = \{k \in N \setminus i | k \to^g j\}$ and $\mu_{-ji}(g) = |N_{-ji}(g)|$. $\mu_{-ji}(g)$ is ***the number of players who reach player j*** in network g, *excluding player i*. We call $\mu_{-ji}(g)$ the **in- reach** of player j in network g.

The notions of μ_j and μ_{-j} should not be confused with the most common notions of out-degree and in-degree, which represent the number of direct links of a player in the network.[39]

Network g determines an expected payoff $\pi_j(g)$ for each player. This is simply calculated as the value of the prize, which we normalise to 1, times the probability of winning the prize, which is equal to the fraction of players accessed by player j:

$$\pi_j(g) = \frac{1 + \mu_j(g)}{n} \tag{1.11}$$

A.1.2 Rule 1

We assume that in T1 player i chooses strategy g_i to maximise expected payoff. That is, he has to pick the partner j such that:

$$\max_j \ \pi_i(g + ij) \tag{1.12}$$

Notice that when i has the turn $N_i(g) = \{\emptyset\}$ and $\mu_i(g) = 0$. In the first round of the game g_i has all of its entries equal to zero. In the second round, the decision maker

[39]The formal definitions of out-degree and in-degree are as follows. Let $N_i^d(g) = \{j \in N | g_{ij} = 1\}$ be the *set* of players with whom player i is directly linked. $\mu_i^d = |N_i^d(g)|$ is the *number* of players to whom with whom player i is directly linked. This is the *out-degree* of player i. $N_{-i}^d(g = \{j \in N | g_{ji} = 1\}$, on the other hand, is the set of players j such that $g_{ji} = 1$. $\mu_{-i}^d = |N_{-i}^d(g)|$ is the *in-degree*: the number of players who have a direct link to i.

has to consider the game as if g_i had only zero entries, as the link specified in the first round is removed once he declares his second-round strategy.

Proposition 1.1. *Player i maximises $\pi_i(g+ij)$ by choosing the partner with the maximum value of $\mu_{ji}(g)$ in the network g.*

Proof. Rewrite $\pi_i(g+ij)$ as: $\frac{1+\mu_i(g+ij)}{n}$. Notice that, as $\mu_i(g)=0$, $\mu_i(g+ij)=1+\mu_{ji}(g)$. Thus $\pi_i(g+ij)=\frac{2+\mu_{ji}(g)}{n}$, which is monotonically increasing in $\mu_{ji}(g)$. □

A.1.3 Rule 2

Let player i's preferences in T2 be described by utility function (1.3). The gain in utility obtained from a g_{ji} link can be summarised as follows:

$$\begin{aligned} u_i(g+ji)-u_i(g) &= \pi_i(g+ji)-\pi_i(g)+\gamma\sum_{k\in N\setminus i}\pi_k(g+ji)-\pi_k(g) \\ &= \gamma\sum_{k\in N\setminus i}\pi_k(g+ji)-\pi_k(g) \\ &= \gamma f(g+ji) \end{aligned} \tag{1.13}$$

where $f(g+ji)=\sum_{k\in N\setminus i}\pi_k(g+ji)-\pi_k(g)$. In T2 a link as no effect on player i's reach, nor on his effected payoff. However, player i has a preference for links that maximise the increase in the expected payoff of the other players. $f(g+ji)$ measures the increase in expected payoff for the other players. To study how $f(g+ji)$ is related to player j's position in the network, there are two cases we need to consider.

Case 1: $j\notin N_i(g)$. Here $f(g+ji)$ can be simply expressed as:

$$\left(\mu_i(g)+1\right)\left(\mu_{-ji}(g)+1\right) \tag{1.14}$$

To derive (1.14), we first need to show the following property of networks in T2.

Lemma 1. *In T2, when it is player i's turn to play, if $j \notin N_i(g)$, no player in $j \cup N_{-j}(g)$ reaches a player in $i \cup N_i(g)$.*

Proof of lemma. We refer to $g_{ij} = 1$ as an 'outgoing' link for player i and $g_{ji} = 1$ as an 'incoming' link. In T2, players can have have multiple 'outgoing' links, but can have *at most* one 'incoming' link. When it is player i's turn to play, player i has no 'incoming' links. He may have one or more 'outgoing' links, in which case $N_i(g)$ is nonempty. Every individual k in $N_i(g)$ has exactly one 'incoming' link. This 'incoming' link is either a link with i ($g_{ik} = 1$), or it is a link with a third player z in $N_i(g)$ ($i \to^g z$). Thus, no player in the set $i \cup N_i(g)$ has an 'incoming' link with a player outside the set $i \cup N_i(g)$. If a player j is not in $N_i(g)$, he cannot reach any player in $i \cup N_i(g)$.

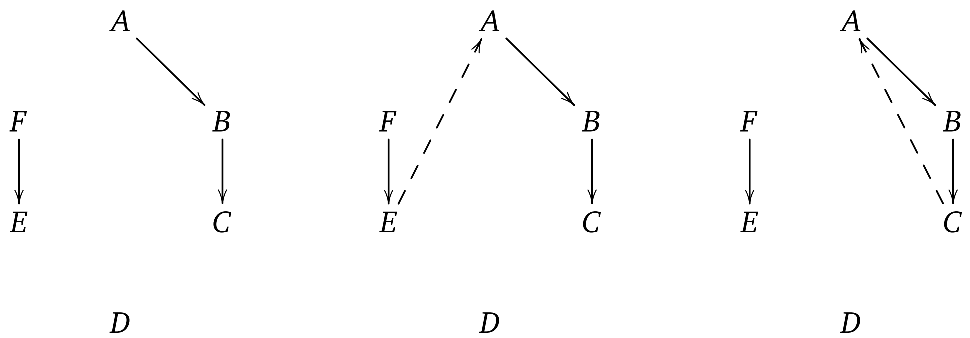
Furthermore for $j \notin N_i(g)$, $i \cup N_i(g) \cap N_{-j}(g) = \emptyset$. Player i does not reach player j. Further, No player reached by player i reaches player j. If some player reached by player i reaches player j, then player j would be part of $N_i(g)$, which contradicts $j \notin N_i(g)$. This concludes the proof. □

If $j \notin N_i(g)$, and player i sets $g_{ji} = 1$, player j now reaches player i and all players in $N_i(g)$. Lemma 1 tells us that none of the players in $N_i(g)$ was previously reached by player j. In other words, each individual in $N_i(g)$ allows player j to observe new information. Player j's expected payoff increases by $\mu_i(g) + 1$.

When player i sets $g_{ji} = 1$, all individuals in $N_{-j}(g)$ now reach player i and all players in $N_i(g)$. Lemma 1 tells us that none of the players in $N_i(g)$ was previously reached by any player in $N_{-j}(g)$. The expected payoff of each player in $N_{-j}(g)$ also goes up by $\mu_i(g)+1$. There are $\mu_{-j}(g)$ players in $N_{-j}(g)$. So the sum of expected payoff among players increases by $\left(\mu_i(g)+1\right)\left(\mu_{-j}(g)+1\right)$. As $j \notin N_i(g)$, $\mu_{-j}(g)=\mu_{-ji}(g)$. We can express the increase in expected payoff also as: $\left(\mu_i(g)+1\right)\left(\mu_{-ji}(g)+1\right)$.

Case 2: $j \in N_i(g)$. Now $g_{ji} = 1$ link creates a cycle. The effect on $f(g+ji)$ is smaller than that of a link to a player outside of $N_i(g)$ who has the same in-reach. This is because some of the information that player j shares with the players who reach him is redundant. Figure 1.14 shows this with an example. In the example, player A has a reaches player C through player B. If player A links to player C, player C now reaches both A and B, and hence μ_C increases by 2. μ_B, however, increases by 1 only, as player B already observes player C. In other words, some of the information given to player B is redundant.[40]

Figure 1.14: Effect of a g_{ji} link



(a) Example (b) Links creates no cycle (c) Links creates cycle

The simple heuristic of choosing the player with the maximum in-reach approximates well the more complicated rule which would calculate $f(g+ji)$ for every possible partner j and pick the partner with the highest value of $f(g+ji)$.

[40]If $N_i(g)$ is a line, when player i creates a link to a player $j \in N_i(g)$, the first player in the $i \to^g j$ path, call him k, experiences an increase in $\mu_k(g)$ of 1, the second player an increase of 2, and so on.. until we to get to player j, who gets an increase in $\mu_j(g)$ of $\mu_{-ji}(g)+1$. Thus we can express $f(g+ji)$ as

$$\begin{aligned} f(g+ij) &= \sum_{n=1}^{\mu_{-ji}+1} n \\ &= \frac{(\mu_{-ji}+1)(\mu_{-ji}+2)}{2} \end{aligned} \tag{1.15}$$

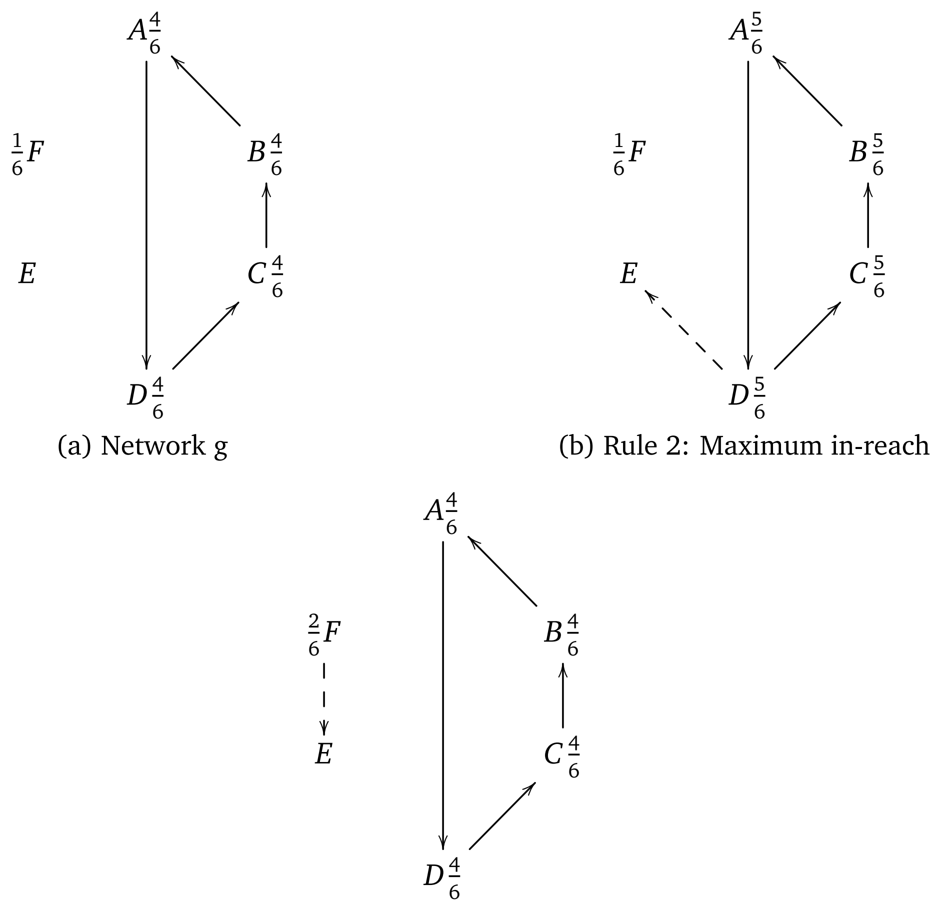
If $N_i(g)$ is a tree, this expression would become more complicated.

1. When $N_i(g) = \emptyset$, $f(g + ji)$ monotonically increases in j's in-reach. In this case, rule 2 maximises $f(g + ji)$.
2. When $N_i(g) \neq \emptyset$ and no player with maximum in-reach is in the $N_i(g)$ set, rule 2 maximises $f(g + ji)$.
3. When $N_i(g) \neq \emptyset$ and at least some of the players in the set of individuals with maximum in-reach are in $N_i(g)$, rule 2 sometimes fails to maximise $f(g + ji)$. The latter occurs when a link to a player outside of $N_i(g)$ who does not have maximum in-reach has a larger effect on $f(g+ji)$ than a link to a player in $N_i(g)$ with maximum in-reach. Rule 2 would suggest the wrong partner here.

To minimise the likelihood of the third scenario, we formulate rule 2 in terms of $\mu_{-ji}(g)$, as opposed to $\mu_{-ji}(g)$. That is, we do not consider player i in the count of player j's in-reach. This only penalises players in $N_i(g)$, and makes it less likely that rule 2 selects them when players outside of $N_i(g)$ with a higher potential impact on $f(g + ji)$ are available.

A.2 Figures

Figure 1.15: Example of a network



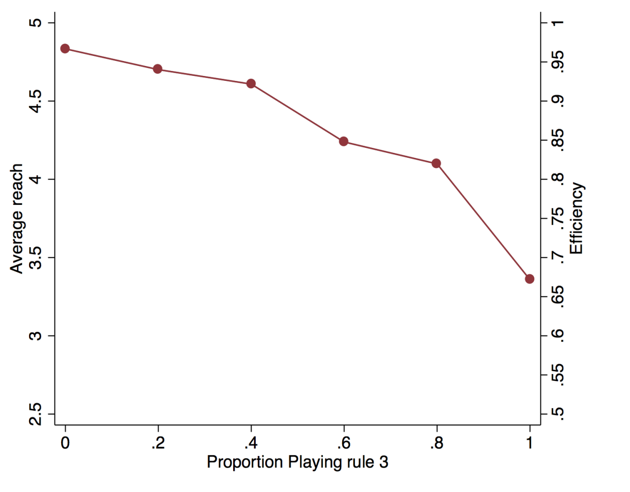
(a) Network g

(b) Rule 2: Maximum in-reach

(c) Rule 3: Minimum reach

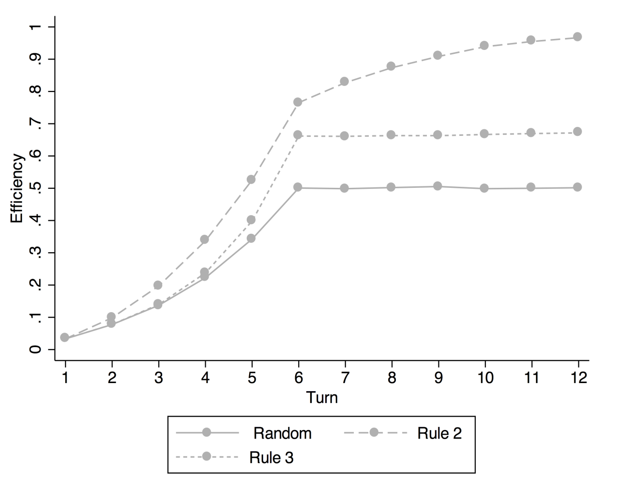
E has the turn. The probability of winning the prize is reported next to each player. Panel (b) and (c) show network $g + ji$, where the new link is chosen following link-formation rule 2 or link-formation rule 3.

Figure 1.16: Link-formation process with mixed rules



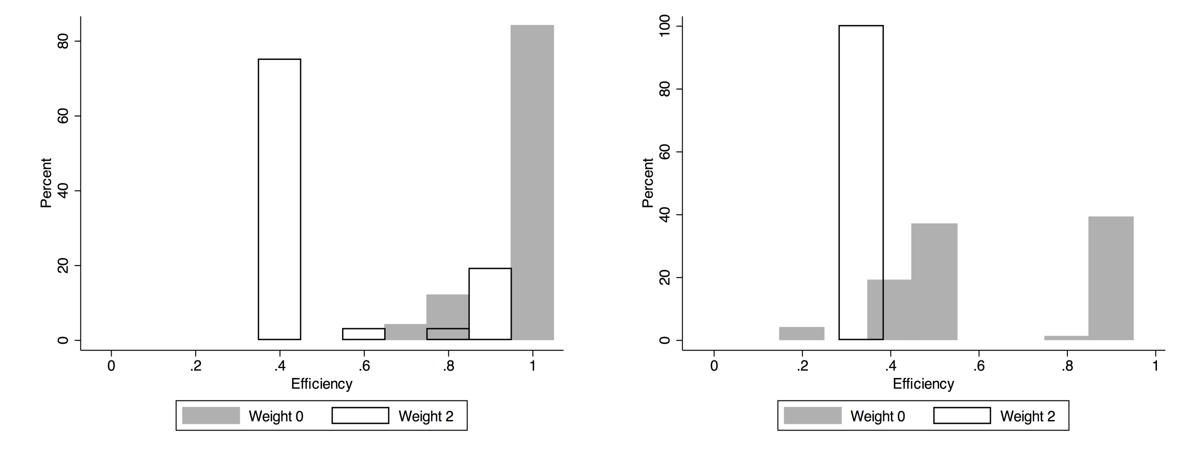
The figure reports simulations where rule 3 is played with probability p and rule 2 with probability 1-p. 500 simulation for each level of p.

Figure 1.17: Simulated time series of average reach



Each rule is simulated 500 times.

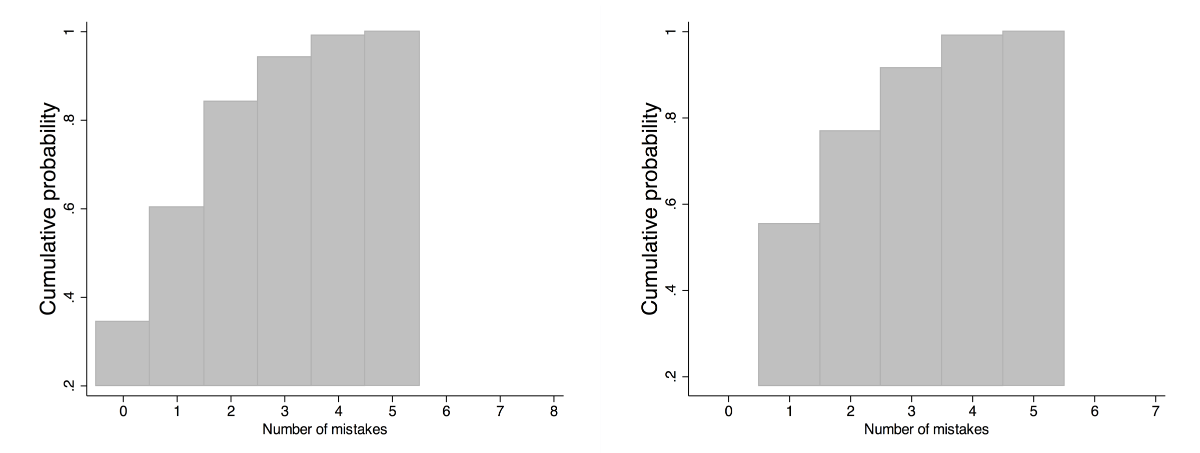
Figure 1.18: Simulated effect of group identity concerns on network structure



(a) Rule 2 (b) Rule 3

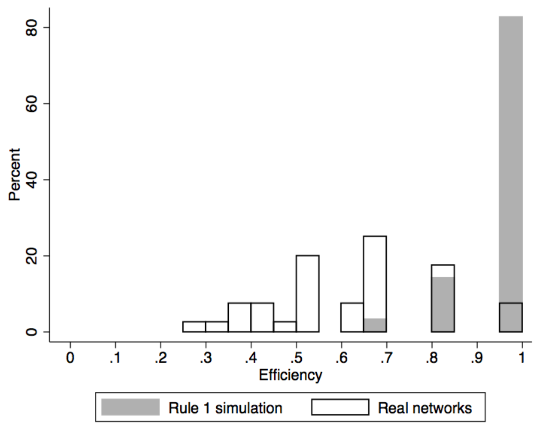
Weight 0 simulations show efficiency when all players play rule 2 (panel a) or rule 3 (panel b). Weight 2 simulations show efficiency when players value a link to an in-group player as much as 2 units of in-reach (panel a) or two units of reach (panel b). For each case we report results summarising 100 simulations.

Figure 1.19: Cumulative distribution of mistakes in understanding questions

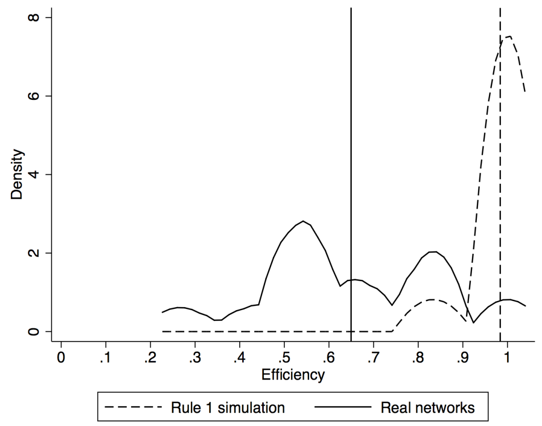


(a) T1 (b) T2

Figure 1.20: Efficiency in no-identity sessions and under rule 1

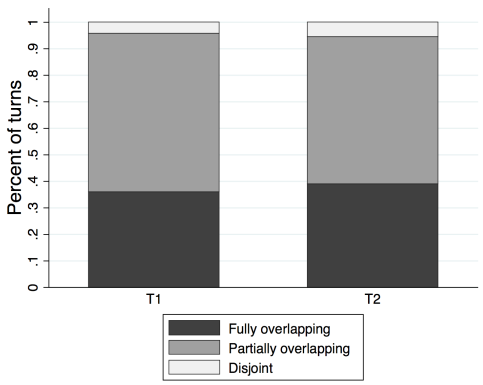


(a) Histogram

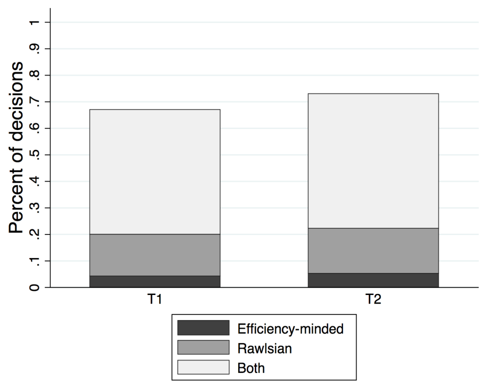


(b) Kernel density estimate and means

Figure 1.21: Overlap in decisions and in choice sets

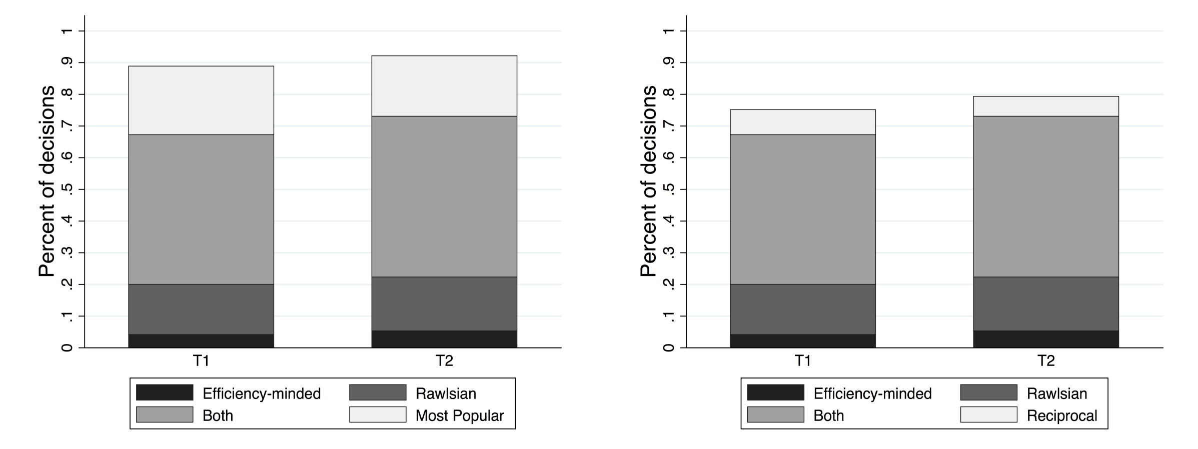


(a) Best response sets



(b) Decisions

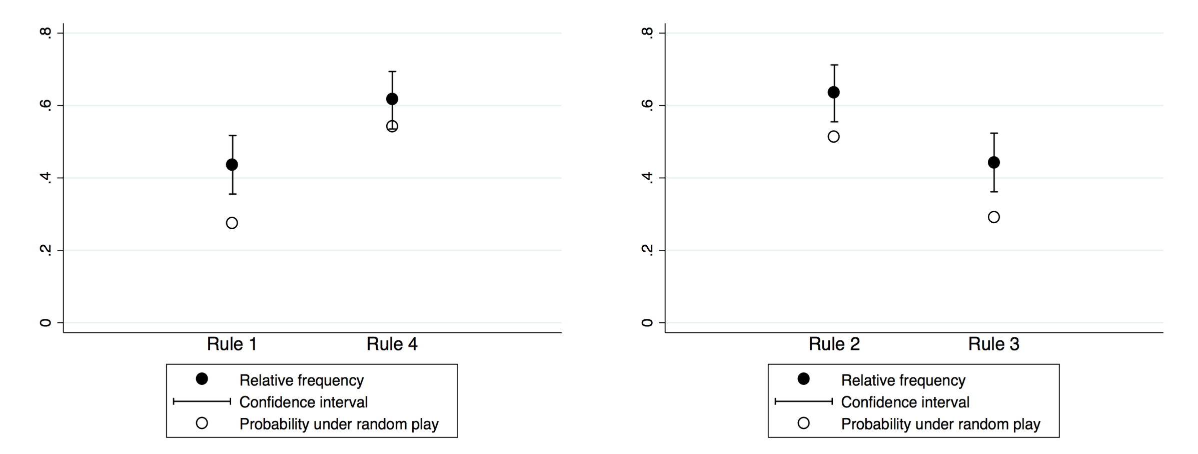
Figure 1.22: What explains links that are not consistent with the archetypal rules?



(a) Links to the 'most popular' player (b) 'Reciprocal' links

The category 'most popular' shows the relative frequency of decisions consistent with rule 5 and not consistent with the 'efficiency-maximising' and 'Rawlsian' rules. The category 'reciprocal' shows relative frequency of decisions consistent with rule 6 and not consistent with the 'efficiency-maximising' and 'Rawlsian' rules. Only data for sessions with no knowledge of group identity is shown.

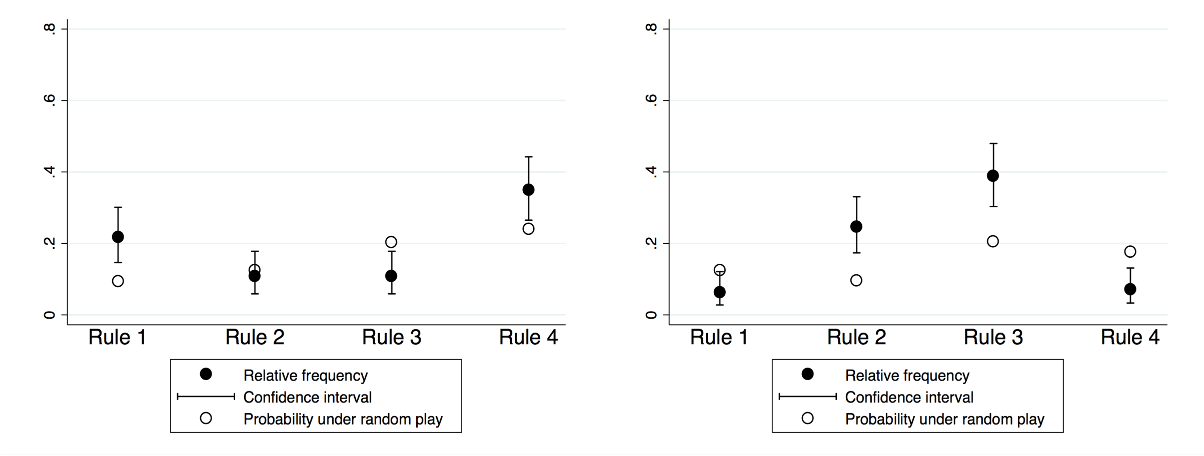
Figure 1.23: Relative frequency of decisions consistent with the hypothesised rules. Best response sets not fully overlapping



(a) T1 (b) T2

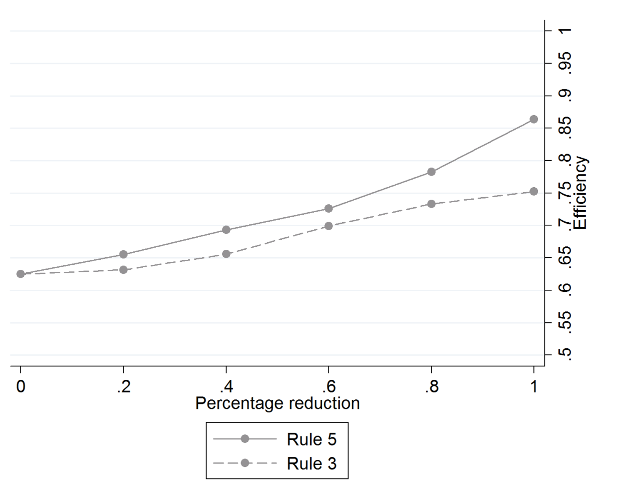
T1 sessions: rounds in which the set of individuals with maximum reach is not perfectly overlapping with the set of individuals with minimum in-reach. T2 sessions: rounds in which the set of individuals with maximum in-reach is not perfectly overlapping with the set of individuals with minimum reach.

Figure 1.24: Relative frequency of individuals who play twice consistently with the hypothesised rules



(a) T1 (b) T2

Figure 1.25: Efficiency simulation



In the baseline simulation 5 percent of decisions follow rule 1, 65 percent follow rule 4, and 30 percent follow rule 5.

Figure 1.26: Distribution of coin allocations to in-group partner

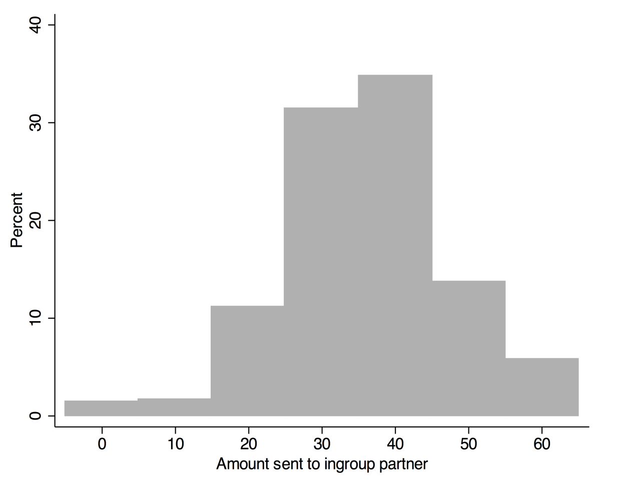
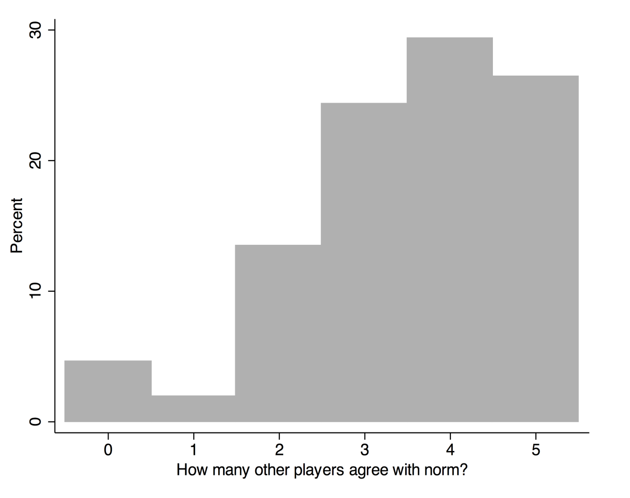


Figure 1.27: 'How many of the other 5 players in the session do you think answered YES to the previous question?' Distribution of expectations



A.3 Tables

Table 1.6: Balance test: identity sessions

	Age	Edu	UpperCaste	LandOwned	LandCult	NetSize
	(1)	(2)	(3)	(4)	(5)	(6)
Identity	-.194 (1.764)	.029 (.056)	-.087 (.067)	.063 (.517)	.101 (.468)	-.201 (1.100)
Obs.	479	466	433	475	470	428

OLS regressions. The dependent variable is indicated in the row's name. Upper caste is a variable that takes value of 1 if respondent is not from a schedule caste, a scheduled tribe or an Other Backward Caste. Network size is the self reported number of peers with whom the farmer exchanges advice on agricultural matters. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors clustered at the session level reported in parentheses.

Table 1.7: Balance test: T2 sessions

	Age	Edu	UpperCaste	LandOwned	LandCult	NetSize
	(1)	(2)	(3)	(4)	(5)	(6)
T2	-1.582 (1.761)	-.028 (.056)	-.052 (.068)	-.085 (.514)	-.049 (.465)	1.293 (1.089)
Obs.	479	466	433	475	470	428

OLS regressions. The dependent variable is indicated in the row's name. Upper caste is a variable that takes value of 1 if respondent is not from a schedule caste, a scheduled tribe or an Other Backward Caste. Network size is the self reported number of peers with whom the farmer exchanges advice on agricultural matters. . Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors clustered at the session level reported in parentheses.

Table 1.8: Balance test: identity sessions in T1

	Age	Edu	UpperCaste	LandOwned	LandCult	NetSize	Und
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Identity	-2.378 (2.544)	.091 (.081)	-.040 (.098)	.077 (.737)	.141 (.657)	.111 (1.102)	-.267 (.178)
Obs.	235	232	215	234	231	211	240

OLS regressions. The dependent variable is indicated in the row's name. Upper caste is a variable that takes value of 1 if respondent is not from a schedule caste, a scheduled tribe or an Other Backward Caste. Network size is the self reported number of peers with whom the farmer exchanges advice on agricultural matters. Und is the number of mistakes in the initial 7 understanding questions. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors clustered at the session level reported in parentheses.

Table 1.9: Balance test: identity sessions in T2

	Age	Edu	UpperCaste	LandOwned	LandCult	NetSize	Und
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Identity	1.879 (2.400)	-.033 (.076)	-.135 (.093)	.046 (.733)	.061 (.673)	-.482 (1.877)	-.224 (.197)
Obs.	244	234	218	241	239	217	246

OLS regressions. The dependent variable is indicated in the row's name. Upper caste is a variable that takes value of 1 if respondent is not from a schedule caste, a scheduled tribe or an Other Backward Caste. Network size is the self reported number of peers with whom the farmer exchanges advice on agricultural matters. Und is the number of mistakes in the initial 8 understanding questions. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors clustered at the session level reported in parentheses.

Table 1.10: Reach and efficiency in final networks

Treatment	Average reach	Efficiency
T1no	3.258	.652
T1id	2.895	.582
T2no	3.238	.648
T2id	3.333	.666
Total	3.167	.637

Table 1.11: Summary statistics of allocation, expectations and norms

Variable	Obs	Mean	Std. Dev.	Min	Max
Amount allocated to in-group partner	483	36.066	11.659	0	60
In-group bias in allocation	483	.542	.499	0	1
Agrees with norm of homophily	483	.571	.495	0	1
No. other players expected to agree with the norm	478	3.513	1.309	0	5
Expects last link to be an in-group link	402	.385	.487	0	1

'Amount allocated to in-group partner' is the number of Rupees, out of 60, allocated to the in-group partner in the initial allocation task. 'In-group bias in allocation' is a dummy equal to 1 if the player has allocated more than half of the endowment to the in-group partner in the allocation task. 'Agrees with norm of homophily' is a dummy equal to 1 if the player answered yes to the question 'In the link formation game you have just played, do you think a player should only link to a peer of his own group?'. 'No. other players expected to agree with the norm' is the answer to the question 'How many of the other 5 players in the session do you think answered YES to the previous question?'. There is 1 missing value. We also set to missing answers that are greater than 5. 'Expects last link to be an in-group link' is a dummy equal to 1 if the respondent expects the player with the last turn choose an in-group link. This variable excludes the 81 players who have the last turn in the session.

Table 1.12: Dyadic linear probability model (1.7)

	(1)	(2)	(3)	(4)
reach_j	.054	.055	-.042	-.039
	(.001)***	(.001)***	(.001)***	(.014)**
in-reach_j	-.033	-.031	.052	.051
	(.009)***	(.092)*	(.001)***	(.002)***
Const.	.373	.395	.264	.279
	(.002)***	(.006)***	(.002)***	(.094)*
Obs.	1200	910	1260	940
Cluster N	20	20	21	21
Sample	T1no	T1no	T2no	T2no
Controls		✔		✔

Dyadic OLS regression. Dependent variable is a dummy which takes a value of one if i chose to establish a link with j. Each regression contains controls for the round and for each possible pairing of map positions. Regressions in columns 2 and 4 include controls for age, land owned, land cultivated, number of contacts in real information networks, number of mistakes in the initial understanding questions and dummies for having completed secondary education, for being Hindu, and for not belonging to a backward caste. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level. P-values obtained with wild bootstrap-t procedure reported in parentheses.

Table 1.13: Dyadic linear probability model (1.7)

	(1)	(2)
max reach$_j$	.121 (.012)**	
min in-reach$_j$	-.0004 (.931)	
max in-reach$_j$		.122 (.022)**
min reach$_j$		.059 (.202)
Reciprocal$_{ij}$	-.093 (.006)***	-.010 (.399)
Most popular$_j$	-.031 (.308)	-.011 (.685)
Const.	.402 (.001)***	.225 (.036)**
Obs.	910	940
Sample	T1no	T2no
Cluster N	20	21
Controls	✔	✔

Dyadic OLS regression. Dependent variable is a dummy which takes a value of one if i chose to establish a link with j. Each regression contains controls for the round and for each possible pairing of map positions. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level. P-values obtained with wild bootstrap-t procedure reported in parentheses. Panel b reports the F statistics (and p value in parenthesis) for a Wald test of the equality of coefficients.

Table 1.14: Linear probability model (1.8): in-group links

	T1	T2
Identity session	.117 (.055)**	.046 (.056)
Const.	.282 (.037)***	.268 (.033)***
Obs.	438	447
Sample	T1	T2
Cluster N	40	41

Linear Probability Model. Dependent variable takes the value of 1 if link d in session s is towards an in-group partner. First turn, first round decisions are dropped. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors clustered at the session level reported in parentheses.

Table 1.15: Linear probability model (1.8): efficiency-minded links

	T1	T2
Identity session	-.0003 (.065)	-.080 (.062)
Const.	.473 (.044)***	.519 (.045)***
Obs.	438	447
Sample	T1	T2
Cluster N	40	41

Linear Probability Model. Dependent variable takes the value of 1 if link d in session s is towards a partner with the maximum out-degree (in T1) and maximum in-degree (in T2). First turn, first round decisions are dropped. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors clustered at the session level reported in parentheses.

Table 1.16: Linear probability model (1.8): Rawlsian links

	T1	T2
Identity session	-.013 (.055)	-.038 (.055)
Const.	.600 (.033)***	.649 (.036)***
Obs.	438	447
Sample	T1	T2
Cluster N	40	41

Linear Probability Model. Dependent variable takes the value of 1 if link d in session s is towards a partner with the minimum in-degree (in T1) and minimum out-degree (in T2). First turn, first round decisions are dropped. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors clustered at the session level reported in parentheses.

Table 1.17: Linear probability model (1.8): in-group links
Restricted sample

	T1	T2
Identity session	.139 (.066)**	.053 (.079)
Const.	.298 (.048)***	.242 (.051)***
Obs.	207	215
Sample	T1	T2
Cluster N	40	41

Linear Probability Model. Dependent variable takes the value of 1 if link d in session s is towards an in-group partner. Sample restricted to 'efficiency-minded' links. First turn, first round decisions are dropped. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors clustered at the session level reported in parentheses.

Table 1.18: Linear probability model (1.9): in-group links T1 treatment

	(1)	(2)	(3)	(4)
Identity session$_s$	.095 (.073)	.133 (.077)*	-.014 (.115)	.191 (.073)***
Bias in allocation task$_i$	.002 (.084)			
Bias$_i$* Identity session$_s$	.044 (.124)			
Homophily Norm$_i$		.055 (.080)		
Norm$_i$*Identity session$_s$		-.027 (.097)		
Homophily norm expectation$_i$			-.014 (.024)	
Norm expectation$_i$*Identity session$_s$			.040 (.032)	
Ingroup link expectation$_i$				.074 (.072)
Link expectation$_i$*Identity session$_s$				-.133 (.107)
Obs.	438	437	435	371
Sample	T1	T1	T1	T1
Cluster N	40	40	40	40

Linear Probability Model. Dependent variable takes the value of 1 if link d in session s is towards an in-group partner. "Bias in allocation task" is a dummy equal to one if the player has allocated more than half of the dictator endowment to the in-group partner. "Homophily norm" is a dummy equal to one if the player has agreed with the statement of the norm of homophily. "Homophily norm expectation" captures the number of other players that the individual expects to agree with the norm of homophily. "in-group link expectation" is a dummy equal to one if the player expects the last player to choose an in-group link. First turn, first round decisions are dropped. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors clustered at the session level reported in parentheses.

Table 1.19: Linear probability model (1.9): in-group links T2 treatment

	(1)	(2)	(3)	(4)
Identity session$_s$	.014 (.080)	.001 (.073)	-.080 (.133)	.024 (.060)
Bias in allocation task$_i$	-.015 (.074)			
Bias$_i$*Identity session$_s$	.054 (.111)			
Homophily norm$_i$		-.011 (.082)		
Norm$_i$*Identity session$_s$		.083 (.103)		
Homophily norm expectation$_i$			-.032 (.021)	
Norm expectation$_i$*Identity session$_s$			.034 (.040)	
Ingroup link expectation$_i$				.018 (.064)
Link expectation$_i$*Identity session$_s$				.178 (.100)*
Obs.	447	447	440	371
Sample	T2	T2	T2	T2
Cluster N	41	41	41	41

Linear Probability Model. Dependent variable takes the value of 1 if link d in session s is towards an in-group partner. “Bias in allocation task" is a dummy equal to one if the player has allocated more than half of the dictator endowment to the in-group partner. “Homophily norm" is a dummy equal to one if the player has agreed with the statement of the norm of homophily. “Homophily norm expectation" captures the number of other players that the individual expects to agree with the norm of homophily. “in-group link expectation" is a dummy equal to one if the player expects the last player to choose an in-group link. First turn, first round decisions are dropped. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors clustered at the session level reported in parentheses.

Table 1.20: Linear probability model: understanding in T1

	Rule 1	Rule 4	Reciprocal	Most Popular	Ingroup
	(1)	(2)	(3)	(4)	(5)
Identity session$_s$	-.008 (.065)	-.030 (.069)	-.086 (.043)**	.120 (.052)**	.105 (.085)
Understanding$_i$* Identity session$_s$=0	.016 (.034)	-.003 (.041)	-.035 (.023)	.032 (.029)	-.010 (.034)
Understanding$_i$* Identity session$_s$=1	.031 (.026)	.017 (.029)	.027 (.017)	-.043 (.024)*	-.009 (.034)
Const.	.491 (.053)***	.633 (.057)***	.153 (.039)***	.363 (.037)***	.304 (.062)***
Obs.	478	478	478	478	478
Sample	T1	T1	T1	T1	T1
Culster N.	40	40	40	40	40

Linear Probability Model. Dependent variable takes the value of 1 if a link is consistent with the link-formation rule indicated in the heading. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors clustered at the session level reported in parentheses.

Table 1.21: Linear probability model: understanding in T2

	Rule 2	Rule 3	Reciprocal	Most Popular	Ingroup
	(1)	(2)	(3)	(4)	(5)
Identity session$_s$	-.072	-.030	-.004	-.001	.053
	(.059)	(.051)	(.031)	(.041)	(.055)
Understanding$_i$* Identity session$_s$=0	.0009	.023	.017	-.022	.008
	(.036)	(.034)	(.019)	(.029)	(.031)
Understanding$_i$*Identity session$_s$=1	.042	.024	-.010	.007	.031
	(.040)	(.029)	(.018)	(.024)	(.041)
Const.	.559	.675	.088	.417	.264
	(.044)***	(.035)***	(.023)***	(.028)***	(.032)***
Obs.	488	488	488	488	488
Sample	T2	T2	T2	T2	T2
Cluster N	41	41	41	41	41

Linear Probability Model. Dependent variable takes the value of 1 if a link is consistent with the link-formation rule indicated in the heading. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors clustered at the session level reported in parentheses.

Instructions

Good evening! Thank you for joining us. Tonight you will be taking part in a experiment that will last about 1,30 h.

During the experiment you are not allowed to talk. Also, please switch off your mobile phones. You will have a chance to ask questions after I finish each part of this explanation. But while the experiment is on, we ask you please not to talk. You are given a paper copy of the instructions I am about to read. You can refer to that at any point in the game.

You are of course allowed to leave the game whenever you want. However, payments will be made only at the end of the whole game. Hence if you choose to leave, you will only collect the show-up fee and will not receive any further payment.

In this experiment, you have the chance to win real money. How much money you win depends on the way you play the game, and on chance.

In front of you, you will find a game sheet and a copy of these instructions.

The game sheet is where all decisions in this game will be recorded.

Tonight your activities will be divided in 5 tasks. I will explain each task in turn.

Before we start, please fill in your name, surname, and telephone number in the appropriate boxes at the beginning of your game sheet.

Also, before we start, you will be assigned to a group. There are two groups: the mangoes and Pineapples. We will assign people to each group randomly. You can think of these two groups as producer association: one group produces mangoes and the other group produces pineapples.

We have also assigned a letter of the alphabet for each player. During the game, you will be identified by this letter. You are not allowed to inform anyone in the room of which letter you have been assigned. If you are found doing so, you will not be allowed to continue the game and will not collect any winnings apart from the show-up fee.

On your game sheet, we have informed you of which group you belong to. We have also made a note of the letter you have been assigned.

Task 1

The Mango and Pineapple group will now take part in a contest of agricultural knowledge.

I will ask three questions. Each one of you will have to answer them on their game sheet.

If everyone in the group gets the three questions right, the team will win one point. You will be informed at the end of the game of whether your team won the point. There will be applause from all participants to celebrate each team that wins the point.

Also, if your group wins a point, this point will be added to the points scored by other farmers in your group in previous days. At the end of the game, you will also be informed of which team has collected the highest number of points so far in the contest of agricultural knowledge.

(QUESTIONS OMITTED FROM THIS VERSION)

Task 2

In this task, you have to divide a sum of money- 30 INR- between two other players in the following session of this experiment. Their names and identities will remain unknown to you. Player 1 will be a player assigned to the Mango group, and Player 2 a player assigned to the Pineapple group.

Please find in front of you six fake plastic coins. On your sheet there is one box that corresponds to the Mango player, and one box that corresponds to the Pineapple player. To divide the money between the two players, please divide the coins between the two boxes.

Notice that the box on the left corresponds to the Mango player. If you put more coins on this box, you are giving more money to the Mango player. If you put more coins on the box on the right, you are giving more money to the Pineapple player.

Please tick the box next the allocation you would like to make.

Task 3

I will now explain to you the rules of the help game. In this game one of the participants will be randomly drawn to receive a prize. Imagine this was something of real value, for example, the name of the best fertilizer to grow a certain crop.

Your task in this game will be to choose a farmer from whom you would like to receive some help. This is like in real life, where other farmers often help you to manage well your farm by sharing with you valuable information about new agricultural technologies.

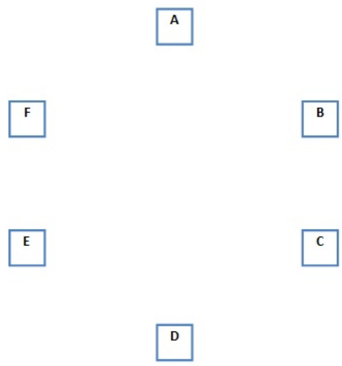
Helping another farmer means sharing the valuable information with him. So if the farmer who is assigned the prize helps a second farmer, the second farmer will also win the prize. This will continue like a chain. If the second farmer is helping a third farmer, the third farmer will also win the prize. And so on.

In the game, each participant is allowed to request help only from one other participant. He will however be indirectly helped by all the players who are helping his direct friend.

At the end of the game, both the person who is assigned the prize and those who are helped directly or indirectly by him will be paid 100 Rupees.

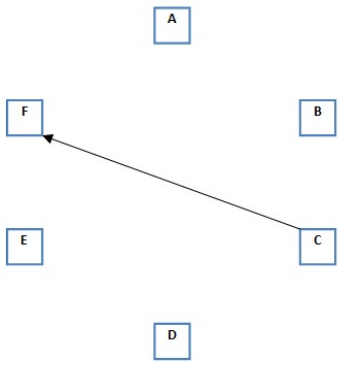
Let me give you some examples about the way the network works. In your session there are 6 people. Each player is identified with an ID letter: A, B, C, D, E, F. Notice that in the network map players who belong to the Mango group are represented with a circle. Players who belong to the Pineapple group are represented with a triangle.

Network map 1



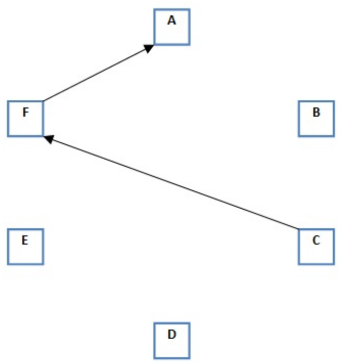
When a player helps another player, we draw an arrow between them. Notice that when the arrow points from player C to player F, for example, this means the information flows from player C to player F. This means that player C is helping player F while player F is not helping player C.

Network map 2



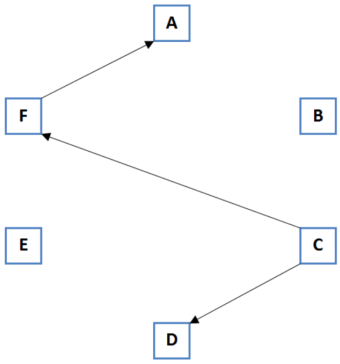
Consider the case presented in network map 3 where player C is helping player F and player F is helping player A. In this case player A receives help from both F and C. If F or C get the fertilizer information, A also gets it and wins the prize. If on the other hand, D gets the prize, A will not get the prize.

Network map 3



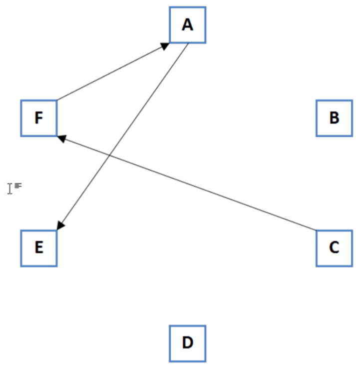
The easiest way to check which players are helping player A is to follow the arrows with your finger. For example, in network 4 below player A receives help from player F and C. Notice he does not receive help from D.

Network map 4



In network 5 below A is receiving help from C and F. Notice he does not receive help from E.

Network map 5



When it is your turn to play, we will remind you of the number of indirect sources of help, which every player would give you if you choose to link with them. For example, the number 3 next to C means that C is receiving help from 3 players. Hence, if you request help from player C, you will also indirectly receive help from three more players.

You can now ask any questions you may have.

I will now ask you some questions to make sure you have understood the rules of this game. Please go to the second page of your game sheet.

Task 4

Move to page X of your game sheet. Now each one of you will have the opportunity to request help from one player.

Let me hand over to you an empty network map where all decisions in this task will be recorded.

You will choose sequentially. There will be 2 rounds of play. For each round, there will be 6 turns of play: every participant will have a chance to play for a turn in each round. So you will have a chance to play two times during this task. The order of play within a round will be chosen randomly and you will not be informed about it beforehand.

During each turn only one of you will choose. Between turns I will collect the game sheets, update the network sheets, and give both of them back to you. You will know that your turn to choose in a given round has come because I will tick the "your turn" box on your game sheet before I give you back the game sheet for that turn.

If it is your turn you will find that we have written a number next to box for each other player in the network map. This reminds you the number of additional players whose help you will receive if you request help from a particular player. For example, the number 3 next to C means that if you request help from player C, you will also indirectly receive help from three more players.

I will ask you to look at the updated network for a minute. After that I will ask you to circle the letter of the player you would like to observe in your game sheet. To do so, I will say "Now write down your choice". If it is not your turn to play, when you hear that, please make a circle around one of the Xs at the bottom of the page.

Notice you cannot choose to request help from yourself. You can however choose not request help from anybody. To do that, just make a circle around one of the Xs at the bottom of the page.

As I have explained, you are only allowed to request help from one person. If in the second round you choose a person who is different from the one you chose in the first round, you will not be able to get direct help from the first person anymore. For example, say that in the first round you choose D. If you then choose E in the second round, you lose the chance of receiving direct help from D.

After two rounds of play I will stop the game. I will then draw one of you who will get the prize. Whoever is helped directly or indirectly by this individual will also get the prize. The prize is worth 100 Rupees.

Notice that the higher the number of people who help you indirectly, the higher is your chance to win the prize.

Also remember that when you request help from a person, this does not imply that he will receive your help. Hence in this game, you cannot choose who you help.

Please now ask any questions you may have.

We will now play a practice round of the game. In this round, everyone will get to make a choice.

Task 5 (LINK FORMATION GAME)

Now we are going to play the real game. Whoever wins the prize in this round will be paid 100 INR.

Finally, remember that players who belong to the Mango group are represented with a circle. Players who belong to the Pineapple group are represented with a triangle.

Task 6 (AFTER LINK FORMATION GAME)

Before we draw the prizes, and before we show the decision of the last player, we will ask you to guess whom the last player of this game requested help from.

Now, please answer the following question: “Do you think in a help network like the one you have just created a player should always request help from people belonging to his own group?” Please tick the YES or the NO box below this question in your game sheet.

Now, we would like to ask you how many of the other 5 players you think have answered YES to the previous question. That is, how many players think that one should always request help from people belonging to one’s group. Please write down the answer on your game sheet.

2 Choosing connections: Evidence from a link-formation experiment in urban Ethiopia

2.1 Introduction

In this chapter I study a link-formation experiment carried out in urban Ethiopia. I embed participants in an asymmetric job-referral network and ask them to connect to two additional players. A lottery determines whether each participant is allowed to complete a remunerated laboratory task– which I call a 'job' as in Beaman and Magruder (2012). Additionally, players who get this opportunity refer another player for the job, chosen at random among their 'unemployed' neighbours in the job-referral network. The network hence determines who can refer whom for the job. In this setting, similar to Calvo-Armengol (2004), a player is more likely to be referred for the job by a neighbour with fewer connections. In addition, the least-connected players in the network have the lowest expected payoff and the largest marginal benefit from an additional link. I ask the following question: do participants in this game form their new links with the least-connected players in the network?

Empirical analyses of social networks often reveal a wide distribution of links across individuals (Goyal, 2007; Jackson, 2010). Observational data confirms this pattern for

I am indebted to Marcel Fafchamps, Ibrahim Hassen Worku, Bart Minten, and Alemayhu Seyoum Taffesse for their support during this project. I would like to thank Johannes Abeler, Abigail Barr, Marcel Fafchamps, Edoardo Gallo, Megenagna Gashaw Guade, Derek Headey, Jeremy Magruder, Bart Minten, Esteban Ortiz Ospina, Cecilia Piantanida, Simon Quinn, Christopher Roth, Pieter Seernels, Alemayehu Seyoum Taffesse, and seminar participants at the 2013 SEEDEC Conference in Bergen, Norway, the Ethiopian Development Research Institute (Addis Ababa), the Centre for Experimental Social Sciences (Oxford), and the 2013 conference of the Centre for the Study of African Economies for useful comments and suggestions. My gratitude also goes to the Ethiopian Development Research Institute, the Kombolcha City administration, and the enumerators and supervisors whose contributions made this study possible. I finally would like to thank the participants and respondents for their goodwill and patience. All mistakes remain naturally mine. I acknowledge financial support from the International Food Research Policy Institute.

the local 'job-contact network' of the subjects in this experiment.[1] This is the network along which individuals share referrals and information about vacancies (Calvo-Armengol, 2004). A large body of evidence suggests that access to referrals and information influences outcomes in the labour market (Granovetter, 1995; Ioannides and Loury, 2004; Topa, 2011).[2] Individuals who are poorly-connected in job-contact networks can thus suffer from a substantial disadvantage in job search. In developing countries, this is aggravated by the frequent lack of efficient matching mechanisms, high search costs, and absent or insufficient unemployment insurance (J-PAL, 2013).

In the initial network of the experiment participants have an unequal number of connections. New links will target the least-connected players and reduce inequality in connections if players (i) understand the incentives that arise from the structure of the network and (ii) choose links to maximise expected payoff or to pursue other-regarding objectives such as inequality aversion or social-welfare maximisation (Fehr and Schmidt, 1999; Charness and Rabin, 2002).

Manipulation of the rule determining when new links are added to the network allows me to separately analyse self-regarding and other-regarding link-formation behaviour. In the first experimental condition, which I call a 'treatment', I unilaterally

[1] See figures 2.8 and 2.9 in the appendix. In these figures, I report an undirected link if respondent i has reported any exchange of information or referrals from respondent i to respondent j, or from respondent j to respondent i. In figure 2.8 I map the links in the job-contact network of a residential block in my sample. Blue squares represent individuals and lines denote links. The isolated squares on the left of the picture represent the individuals who have never exchanged information nor referrals with any other sampled resident of their block. I define the number of links that an individual has in the network as his 'degree'. In figure 2.9 I plot the distribution of respondents' degree in the job-contact networks of their residential block, for both knowledge networks and job contact networks, and under different conventions on how to treat a link when the reports of i and j do not coincide. The modal degree in the job-contact networks of the residential blocks I study is zero, while some individuals have a large number of connections. This finding is robust to changes in the definition of a link. More information about the data used to produce these figures is given in section 2.4. However an important caveat to notice here is that connections outside the residential block are not captured in the data.

[2] Researchers have collected a large body of descriptive and econometric evidence on these mechanisms. See Topa (2011) for a recent review. In Ethiopia, referral hirings are common (Mano et al., 2010), and the use of social networks in job search is widespread and has a positive impact on the exit rate from unemployment (Serneels, 2007). The data confirms the importance of social interactions: 41 percent of employed individuals in the sample have heard of their current occupation through their social contacts and 29 percent have received an explicit referral.

activate the links chosen by a single, randomly drawn 'unemployed' participant. Under this rule new links improve the probability that the player proposing the link is referred for the job when he is unemployed, but have no effect on the opportunities of the new neighbour. Expected payoff is maximised by forming links with the least-connected players. In the second treatment, I update the network with the links chosen by a single, randomly drawn 'employed' participant. Here links increase the new neighbour's probability of being referred for the job, while leaving the opportunities of the player who proposes the link unaffected. Forming links to the least-connected players maximises the sum of players' payoffs and minimises differences in the number of connections.

I run the experiment with young adult dwellers from a mid-sized town in Ethiopia characterised by an extensive use of job-contact networks. Individuals who play the game together always come from the same neighbourhood, often from the same residential block. While I keep decisions private throughout the experiment, in randomly selected sessions I disclose the personal identity of the players before new links are chosen. This allows me to study the influence of real connections and characteristics on link-formation decisions. Further, the choice of a subject pool that is connected by a dense job-contact network strengthens the external validity of the experimental design. Following the taxonomy of Harrison and List (2004) this study can be defined as an *artefactual field experiment*.

I find broad support for selfish network formation. In the first treatment where decisions affect the payoff of the player who proposes the link, subjects are significantly more likely to form new ties with the least-connected players than with the other individuals in the network. Disclosure of personal identity does not change this pattern. Overall, 47 percent of all proposed new links are directed towards players with the lowest initial number of connections. This reduces the dispersion of links: the distribution of connections that emerges dominates- in the sense of second-order stochastic dominance- the distribution of connections produced by a link-formation strategy that picks each player with the same probability, irrespective of network position.

82 percent of individuals send a positive amount in a dictator game played before the link-formation experiment. This is consistent with widespread other-regarding preferences. However, in the second treatment where links affect the payoff of the new neighbour, players are not significantly more likely to link with least-connected players. Surprisingly, individuals who share a positive fraction of their dictator endowment drive this effect. A potential explanation is that giving in the dictator game captures the propensity of players to follow norms or to conform to the expectations of others, and that what is expected of players is to share endowments equally and to form links with well-connected players. While I present some evidence consistent with this interpretation, the data does not allow me to formulate a direct test.

Finally, when personal identity is disclosed, subjects are more likely to choose the individuals they know in real life. The effect has a large magnitude and occurs irrespectively of who benefits from the link. Among the individuals they know, subjects are more likely to choose those who are unemployed and those who are well-connected in real job-contact networks. However, these last two effects are relatively small in magnitude and not significant at conventional levels.

In this study I make three contributions to the literature. First, I show that individuals strategically choose links on the basis of the structure of the network. This is a central tenet of models of strategic network formation (Jackson and Wolinsky, 1996; Bala and Goyal, 2000; Calvo-Armengol, 2004). It is also an important theme in the empirical analysis of the formation of networks for the sharing of risk, favours, information and labour (Fafchamps and Gubert, 2007; Krishnan and Sciubba, 2009; Karlan et al., 2009; Comola, 2010; Jackson et al., 2012; Santos and Barrett, 2010; Attanasio et al., 2012). A number of papers use field data to explicitly study how decisions are affected by the structure of the network. For example, Karlan et al. (2009) show that the propensity to exchange loans increases with the maximum network flow between two individuals, while Comola (2010) documents that indirect contacts matter for the formation of risk-sharing links. Both studies offer observational tests of the prediction that individuals play strategically with respect to network structure. This chapter offers a clean experimental test of that prediction.

Second, I find that when the welfare of other players is at stake, new links are often not directed to the least-connected players. This result speaks to the literature on social interactions in the labour market, which has explored the importance of the quality of job-seekers' networks (Bayer et al., 2008; Magruder, 2010; Cingano and Rosolia, 2012; Beaman, 2012; Schmutte, 2015).

More broadly, this result suggests that models of other-regarding preferences with high predictive power in simple allocation decisions can perform poorly in different domains– in this case, a link-formation task. This will be of interest to scholars studying other-regarding preferences and norms in Sub-Saharan Africa (Barr and Stein, 2008; Jakiela, 2011; Miller Moya et al., 2011; Mueller, 2012). It is consistent with the evidence presented in Belot and Fafchamps (2014), who highlight a weakening of altruism in large groups, and should motivate researchers to investigate social norms that arise in specific domains and richer models of preferences incorporating expectations (Bicchieri, 2006; Charness and Dufwenberg, 2006).

Third, on the methodology side, I address typical endogeneity concerns that arise in the study of networks through random link assignment. Furthermore, I exclude by design coordination issues and non-equilibrium reasoning. In this sense my work complements a small set of repeated link formation games which have been attempted in the literature (Callander and Plott, 2005; Conte et al., 2009; Goeree et al., 2009; Falk and Kosfeld, 2012) and the experiment of chapter one. In these games, the experimenter has imperfect control over the beliefs that individuals have about the links that other players will create. Identification of the motives behind players' decisions requires particular assumptions about the form of these beliefs. Such assumptions are not necessary to analyse behaviour in the experiment presented in this chapter, as players are certain about the structure of the network on which they are adding their new links. This allows me to provide convincing evidence on the effect of network structure on link formation.

The next section presents the experimental design. Section 2.3 puts forward a number of predictions from a model of network formation with self and other-regarding concerns. Section 2.4 presents the data and section 2.5 the results. Section 2.6 concludes.

2.2 Design

In the experiment, subjects are assigned a position in a pre-determined network and are asked to specify two additional links they would like to form in this network.[3] The way in which I use the preferences for new links to update the network is central to the design and is explained in detail below. After choosing his new links, each subject plays an independent lottery. Winners of the lottery have to perform a task in the lab, for which they will be remunerated. I call this task a 'job' and define players as 'employed' or 'unemployed' depending on whether they have won the lottery that allows them to perform the job. The remuneration for the job is called a 'wage'.

Each employed player refers for the job one unemployed player with whom he is linked in the network. The network thus determines who can refer whom for the job. If a job holder is connected with more than one unemployed player, the referral is given to only one of these players, chosen at random. A referred player is always hired for the job and is paid the same wage as the winners of the original lottery. This game is similar to the game analysed in Calvo-Armengol (2004).[4]

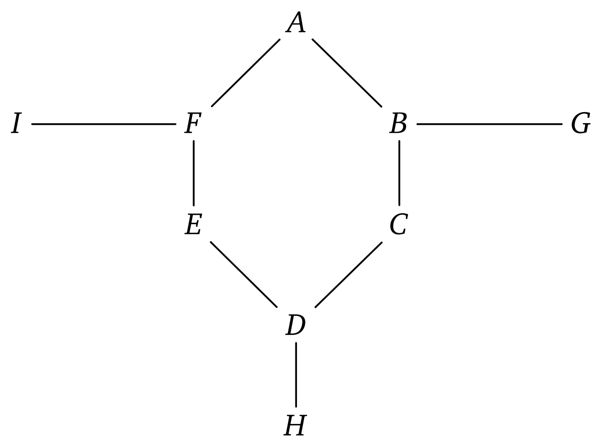
[3]The instructions for one of the treatments are included in the appendix. The remaining experimental materials can be found here: https://sites.google.com/site/stefanoacaria/lfethiopia

[4]The main difference is that in this setting players hold a fixed endowment of connections and can unilaterally add two links, at no cost. In Calvo-Armengol (2004), on other hand, links require mutual consent and impose a cost on both parties. Further, the network-updating rules I describe below are unique to the game presented in this chapter.

For each session, I recruit nine subjects to play the game. Participants typically reside in the same neighbourhood, often in the same block, and hence in many cases know each other. I exploit information about the real-life connections and characteristics of players in the analysis.

The initial network that connects the nine participants is presented in figure 2.1 below. Links are undirected, so if player i is connected to player j, player j is also connected to player i. Individuals are identified with ID letters. Two players connected by a link are referred to as 'friends'. I call the number of friends of a player his 'degree centrality', or just 'degree'. In the network of figure 2.1 there are three degree-one players, three degree-two players and three degree-three players. No player is linked to all players with a given degree. Hence, when choosing his new links, every player can select a partner with degree one, two or three.

Figure 2.1: The experimental network



The experiment proceeds as follows. First, each subject draws from a urn an ID letter, which will remain private throughout the experiment. Second, each subject plays a standard dictator game with an anonymous partner in the same session. Third, the game assistant gives instructions for the link-formation game and disclose subjects' positions in the network, which correspond to the ID letter they have drawn. The game assistant then test subjects' understanding of the network structure and the incentives of the link-formation game by means of a simple questionnaire. Then, all subjects are asked to specify the two links they would like to add to the network.[5] The analysis focuses on this decision. After subjects take the link-formation decisions, they play the job lottery. The network is then updated according to a rule which varies with treatment. The updated network is used to determine the referrals made by employed players.

At the end of the experiment, all winners of the job lottery are asked to perform the job and can then collect their earnings. These include a show-up fee, allocations from the dictator game, and the wage for the job.[6].

Participants who have been referred for the job collect the show-up fee and the dictator game allocation, and are invited to come back the next day in order to perform the job and collect the wage. Finally, participants who did not win the lottery and were not referred for the job collect the show-up fee and dictator game earnings and leave. All payments are given privately.

Treatments vary the network-updating rule in order to identify different motives behind link formation. In a first set of treatments, which I call T1 treatments, I update the network with the two links specified by a single, randomly drawn unemployed player. This rule achieves two things. First, links are implemented only when the player is unemployed and hence is not in a position to make a referral. Hence, in T1, the new links do not affect who the player can give a referral to. This minimises other-regarding considerations.[7] Second, the fact that I implement the links of a single

[5] If they do not specify a link, or if the write R, a link is randomly drawn for them.

[6] The value of the show-up fee was 0.5 USD, the total amount to be divided in the dictator game was about 1.1 USD and the wage 2.2 USD.

[7] Other-regarding considerations are not wholly removed for sophisticated players who care about the welfare of the friends of their chosen new partner. A new link to player j decreases the chances that player j's current friends are going to get a referral. While all links will impose a negative externality on two-step-away players, an intrinsically-motivated- and quite sophisticated!- player may prefer to impose such externality on

player removes strategic thinking (Crawford et al., 2013). Subjects do not have to speculate about what others will do: if their choice is implemented, it will be the only modification to the existing network.[8]

In a second set of treatments, the T2 treatments, I update the network with the links of a single, randomly drawn employed player. Links are now implemented only when the player is in the position to give a referral. They determine who will get the player's referral and cannot be used to change the probability of getting a referral for oneself when one is unemployed. Other-regarding motives become salient in this treatment, while self-regarding considerations are switched off.

Treatment T1 and T2 are played both with anonymous identities (T1a, T2a) and with identities that are common knowledge (T1n, T2n). In the second case, players are first asked to communicate their name in front of the group. Names are then written next to the respective node in the network map. Each participant is given a copy of the map.[9]

Non-anonymous treatments give a more arduous test to the prediction that individuals prefer to link with degree-one players. While degree-centrality is clearly salient in the anonymous treatments, players may focus on a number of other characteristics in the non-anonymous case.For example, they may choose links which reinforce previous bonds with people they know, or they may choose to avoid individuals of specific so-

the better-off players in the experiment. In the network I impose, this gives again a reason to link with a degree-one player, as their only tie is a 'well off' degree-three player. While I cannot exclude with certainty that this type of reasoning occurred in the experiment, I offer two pieces of evidence against it. First, when given the opportunity to directly benefit the least well off partners in a different treatment, players do no show a systematic desire to do so. It is not very plausible then that other-regarding considerations towards one-step-away partners do not drive behaviour, while other-regarding considerations towards two-step-away partners do. Second, I interact decisions in this baseline treatment with the amount sent in the dictator game and show that this interaction is not significant.

[8]Players with level-one rationality, for example, will worry that other subjects are also going to choose to link with degree-one players and that hence degree-one players will actually be quite central in the final network. Players with level-two rationality are going to best respond to level-one players, and so on. The procedure I implement rules out these considerations.

[9]In anonymous treatments, each participants is also given a copy of the network map that only identifies nodes with their ID letter.

cial categories. Importantly, however, decisions remain private: at the end of the game participants are told whether they received a referral or not, but are not informed by whom they received the referral, nor they are shown how the network has been updated. This makes it harder for players in non-anonymous treatments to require side payments from each other at the end of the experiment.

In the standard protocol subjects are explicitly informed about the incentives that arise from the structure of the link-formation game.[10] Before taking their link-formation decisions, players' understanding of the network map and of the incentives is tested with five questions.[11] If more than one participant makes more than one mistake, the lab assistant is instructed to go through the explanation one more time. Such explanations and tests are important to ensure participants understand the consequences of the decisions they are making. However, one may worry that if participants have a desire to please the experimenter (Levitt and List, 2007; Zizzo, 2010), they may be primed by my questions to behave in the way they think I want them to behave. The last treatment- T1a2- is thus devised in order to reduce priming effects. In T1a2, I give participants the same explanation of the rules of the link-formation game, but omit any discussion of the incentives that these rules produce.[12] Furthermore, I postpone the test of understanding until the end of the link-formation game. In this way I reduce both cognitive and social priming (Zizzo, 2010). Furthermore, the T1a2 treatment allows me to test to what extent individuals independently grasp the incentives that derive from the asymmetric structure of the network and the rival referral opportunities. Table 2.1 below summarises the treatments and their characteristics.

Subjects are presented the network structure through the network map reproduced in figure 2.1. Recent research on network cognition has uncovered a projection bias

[10]Subjects in T1 treatments are told that a low-degree player is more likely to give them a referral than a high-degree player. Subjects in the T2 treatments are told that a low-degree player is both the least-likely player to get a referral for himself, and is the one whose chance of getting a referral increases the most with an additional link.

[11]These are presented in the appendix

[12]That is, I do not explain to participants that in this game a high-degree player is less likely to provide them with a referral. Players can of course infer this from the rules of the game, and many of them do.

Table 2.1: Summary of treatments

	Links affect outcomes of...	Anonymity	Explanation of incentives
T1a	self	✔	✔
T1a2	self	✔	
T1n	self		✔
T2a	other players	✔	✔
T2n	other players		✔

which is relevant to the my design: subjects overestimate the degree of players characterised by a degree lower than themselves (Dessi et al., 2012). If that is the case among the participants in this experiment, the role of centrality may become less salient, simply because high-degree subjects overestimate the degree of the other players. I take steps to limit this problem. First, the simple network structure in my design mitigates this concern. Second, the degree of each node is specified next to the ID letter. Third, the first three questions in the understanding questionnaire test whether subjects can infer degree-centrality from the map. Incorrect responses to these questions are extremely infrequent.

At the beginning of the experiment subjects play a standard dictator game (Camerer, 2003). Players split 20 Ethiopian Birr with an anonymous partner in the same session. Each subject is assigned to two different pairs: he plays the sender in first pair and the receiver in the second pair. Players are aware that they are not playing twice with the same partner. Allocation decisions are private. Subjects are informed of the amount they have received only at the end of the experiment.

Dictator games offer a simple way to measure social preferences (Camerer, 2003). These are typically categorised in four standard groups (Charness and Rabin, 2002): selfish, competitive, inequality averse, and social-welfare maximising. Selfish individuals care only about their own payoff. Competitive individuals maximise the difference between their payoff and that of the other player. Inequality-averse individuals minimise payoff differences between themselves and the other players. An inequality-averse person hence gives in the dictator game only when he has a endowment that is larger than that of the other player. Finally, social-welfare maximisers care positively about the receivers' payoffs regardless of the relative size of initial allocations.

The dictator is endowed with ETB 20, whereas the receiver has no initial endowment. The amount given in this game hence reflects the strength of the sender's concern for the receiver's payoff when the initial allocation favours the sender. Giving in the dictator game is hence consistent both with inequality aversion and with a general preference for social-welfare maximisation. For the purposes of this experiment, I do not need to distinguish between these two types of preferences, as they motivate identical behaviour in the link-formation game.[13] I hence consider giving in the dictator game as a general measure of pro-social preferences.

To ensure comparability and minimise noise factors during play, I follow a number of established practices in the lab-in-the-field literature. These include extensive piloting, simple standardised instructions that are read out to participants, double translation of all written material, and reliance on physical randomisation devices (Barr and Genicot, 2008; Viceisza, 2012).

2.3 Predictions

Player i's problem is to choose with whom he should form his two new links. In this section, I study player i's problem formally.

Let $\pi_i(g)$ be player i's chance of being employed given network g, $p \in (0,1)$ be the likelihood of being initially selected to perform the job and $Q_i(g)$ be player i's chance of being referred for the job by at least one player in network g. Furthermore, let $N_i(g)$ be the set of direct contacts of player i in network g, and $n_i(g)$ be the degree of player i. $N_i^c(g)$ represents the set of individuals with whom player does not have a link. $g+ik$

[13]With one exception, explored in the next section.

is the original network augmented by a link between player i and player k.

In T1, links have a direct effect on the probability of being referred for the job. A selfish player will choose his new links so as to maximise his chance of obtaining a referral when unemployed. Formally, for each link, he solves the following problem:

$$\max_{ik \in N_i^c(g)} \pi_i(g+ik) = p + (1-p)Q_i(g+ik) \tag{2.1}$$

$Q_i(g+ik)$ can be expressed as the inverse of the probability of receiving no referral:

$$Q_i(g+ik) = 1 - \prod_{j \in N_i(g+ik)} (1 - pq(n_j(g+ik)) \tag{2.2}$$

Conditional on player k being employed, $q(n_k(g))$ gives the probability that player k will refer player i out of his $n_k(g)$ friends. This probability can be expressed as:

$$q(n_k(g)) = \sum_{z=0}^{n_k(g)-1} \binom{n_k(g)-1}{z} \frac{p^{n_k(g)-1-z}(1-p)^z}{z+1} \tag{2.3}$$

$$= \frac{1-p^{n_k(g)}}{(1-p)n_k(g)} \tag{2.4}$$

After some algebra, (2.4) follows from (2.3).[14] When player k is employed, the probability he will refer player i is given by (i) the probability that player i will be chosen out of the average number of unemployed friends of player k- $\frac{1}{(1-p)n_k(g)}$ times (ii) the probability that at least one friend is unemployed- $1-p^{n_k(g)}$. I can show algebraically that (2.4) decreases in $n_k(g)$:[15]

Proposition 2.1. $\frac{\partial q(n_k(g))}{\partial n_k(g)} < 0$ *for* $p \in (0,1)$ *and* $n_k(g) \geq 1$

[14] See appendix for all proofs.
[15] This result is essentially the first part of Remark 1 in Calvo-Armengol (2004).

The result is intuitive: two-links-away partners are competitors for referral opportunities. A partner with higher degree centrality is less likely to provide player i with a referral. Player i will thus prefer to link with the individual in $N_i^c(g)$ with the lowest degree. The monetary gains from following the optimal strategy are meaningful. Conditional on being un-employed, a degree-one player who establishes two links with two degree-one players has a 72 percent probability of being referred for a job. This probability drops to 58 percent if he chooses to link with two degree-three players. This is a substantial, 14 percentage points difference.[16] I thus formulate the first prediction:

Prediction 2.1. *Subjects in T1 treatments form new links with degree-one players.*

This prediction is based on the the assumption that new links are created in order to maximise the expected payoff from the game. On the other hand, if subjects create links following the heuristic of 'preferential attachment' to high-centrality players, they will form links with degree-three subjects even if it is not in their material interest to do so.

Let use now analyse player i's problem in T2 treatments. Here new links affect the chances of receiving a referral of the other players, and do not affect the referral chances of player i. In the game, the chances of receiving a referral when unemployed increase monotonically with the number of links. Degree-one player are hence those who have the smallest overall chance of employment. Furthermore, the marginal benefit of a new link is decreasing in the number of existing links. Degree-one players are hence also those who stand to gain the most from an additional link. More formally:[17]

Proposition 2.2. $\frac{\partial Q_i(n_k(g))}{\partial n_i(g)} > 0$ *and* $\frac{\partial^2 Q_i(n_k(g))}{\partial n_i(g)^2} < 0$

[16]Subjects in T1a and T1 treatments are informed of these probabilities. Subjects in T1a2 are not.
[17]The second part of Remark 2 in Calvo-Armengol (2004) makes a point similar to the first part of proposition 2 here.

There is now an established literature in economics which explores other-regarding preferences. In a frequently cited paper, Charness and Rabin (2002) categorize the standard types.[18] Players who are social-welfare maximisers or inequality-averse would send money in a dictator game which starts with unequal endowments.

For illustration, let me assume other-regarding preferences of the following form:[19]

$$u_i(g) = \pi_i(g) + \frac{\gamma}{n-1} \sum_{j \in N} \pi_j(g) \tag{2.5}$$

γ is the altruism parameter. In T2 when player i chooses a link to player k, he will increase player k's welfare, while decreasing the welfare of the $n_i(g)$ current contacts of i, who are now facing one more competitor for player i's referral. I can hence decompose the effect of the new link on player i's utility in three elements:

$$\begin{aligned} u_i(g+ik) - u_i(g) = & \underbrace{\pi_i(g+ik) - \pi_i(g)}_{(a)} \\ & + \underbrace{\frac{\gamma}{n-1} \pi_k(g+ik) - \pi_k(g)}_{(b)} \\ & + \underbrace{\frac{\gamma}{n-1} \sum_{j \in N_i(g)} \pi_j(g+ik) - \pi_j(g)}_{(c)} \end{aligned} \tag{2.6}$$

The first element, reflecting the change in the payoff of player i, is equal to zero in T2. The third element captures the negative externality on i's current links. Such

[18]Inequality-averse individuals feel envy towards those with higher payoff and guilt towards those with lower payoff. Social-welfare maximisers care positively about the welfare of the other person, independently on relative payoffs. Competitive individuals maximise the difference between oneself and the other players. Selfish individuals maximise their personal payoff.

[19]These preferences reflect a simple form of altruism. They are akin to the social-welfare preferences for a game with many players with $\delta = 0$ discussed in Appendix 1 of Charness and Rabin (2002).

externality is present no matter whom player i decides to link to and does not depend on $n_k(g)$.[20]

The second term is what motivates player i to link with the least connected individuals. Proposition 2 shows that the marginal payoff player k gets from a link with player i decreases in player k's degree. Choosing a degree-one partner means establishing the link with the player that will benefit from the connection the most. This maximises the value of the second term in (2.6). In the T2 condition, subjects with preferences given by (2.5) will link with a degree-one player.

A similar argument can be made for inequality-averse players (Fehr and Schmidt, 1999). In this game, link-formation affects inequality. Degree-three and degree-two inequality-averse individuals will link to degree-one players because this both avoids increases in disadvantageous inequality, and minimises advantageous inequality. Inequality-averse degree-one individuals, on the other hand, will increase disadvantageous inequality with any of the available links. Connections to degree-one individuals increase the number of individuals with more connections than player i, but decrease the mean difference in connections. Depending on the form of inequality aversion, degree-one individuals may hence prefer to link to degree-one peers or to more connected peers.

Participants may have an additional reason to link to degree-one players: the fact they are the least well off in the game. Several questionnaire studies in empirical social choice have reported that other-regarding individuals attach special weight to the well-being of the least well off (Yaari and Bar-Hillel, 1984; Gaertner and Schokkaert, 2011). These considerations would strengthen the desire to link with degree-one individuals.

[20]Notice that if the player specifies no link, two links are assigned to him anyways. Also notice that unemployed players can receive multiple referrals, in which case one of the referrals is simply lost. This ensures that the degree of player k has no influence on the size of the third element in (2.6). On the other hand, if referrals to players who have already been referred for a job were reassigned, player k's probability of being referred for a job would have influenced the size of the third term.

Prediction 2.2. *Subjects in T2 treatments form new links with degree-one players.*

This effect is driven by pro-social, other-regarding individuals. Hence I expect a positive correlation between giving in the dictator game, which is meant to capture such preferences, and the choice to link with a degree-one player. Notice that while giving in the dictator game is costly, the choice to include a degree-one player does not involve any payoff cost. This minimises concerns about wealth effects generating a spurious negative correlation between the two games (Andreoni and Miller, 2002).

Prediction 2.3. *Subjects who have sent a positive amount in the dictator game are more likely to link with a degree-one player in the T2 treatments. Dictator game giving is uncorrelated with play in T1 treatments.*

The game is played by subjects who come from the same neighbourhood and, in many instances, from the same block. Theories of directed altruism predict that individuals in the T2n treatment will act more altruistically towards acquaintances (Goeree et al., 2010; Leider et al., 2009; Ligon and Schechter, 2012). I do expect this effect to be at work in the T2n treatment, where players are aware of the identities of those who receive their referral. I do not expect, however, directed altruism to play a role in the choices under the T1n treatment, where links to a friend do not produce any material benefit for the friend and can be suboptimal for player i if his friends happen to be degree two or degree three.

Prediction 2.4. *Subjects in the T2n treatment are more likely to form links with those players they know in real life. Decisions of subjects in the T1n treatments will not be affected by acquaintance with other players.*

I will analyse the data using dyadic regression analysis. In particular, I will test predictions 2.1 and 2.2 using models of the following form:

$$link_{ij} = \alpha + \beta_1 c2_j + \beta_2 c3_j + e_{ij} \tag{2.7}$$

The unit of observation is all initially unlinked, directed dyads.[21] $link_{ij}$ is a dummy which takes a value of one if player i has chosen to establish a new link with player j. $c2_j$ and $c3_j$ are dummy variables indicating individuals with degree centrality of two and three, respectively. The coefficients on $c2_j$ and $c3_j$ will provide the basic test for predictions 2.1 and 2.2. If β_1 and β_2 are negative and significant I will have evidence that in this experimental network players value links to less connected individuals more than links to high-degree players.

I run separate analyses for the data from the T1 and T2 treatments, as these elicit different decision mechanisms. I use dummy variables interacted with $c2_j$ and $c3_j$ to investigate the effects of the disclosure of anonymity, the removal of explanations, and player i's characteristics such as understanding or giving in the dictator game:

$$link_{ij} = \alpha + \beta_1 c2_j + \beta_2 c3_j + \delta_1 t_i + \delta_2 (t_i * c2_j) + \delta_3 (t_i * c3_j) + u_{ij} \tag{2.8}$$

Models (2.7) and (2.8) will be estimated using OLS, correcting standard errors for arbitrary correlation at the session level. Previous studies have shown that when the number of independent groups of observations is low, which is often defined as less than 42, hypothesis tests based on clustered standard errors over-reject the null. In my case, regression analysis is based on 30 sessions of the T1 treatment and 20 sessions of the T2 treatment and is hence characterized by a low number of clusters. In a widely cited paper Cameron et al. (2008) show that the wild bootstrap-t method can be used to achieve accurate inference even with few clusters. This method simulates a

[21] This means that the matrix is not the full $n(n-1)$ square matrix. I call the dyads directed because they express directed willingness to link. Notice that the actual dyads in the network are undirected

distribution of the test statistics which can be used for hypothesis testing in conjunction with the original test statistics. I apply the wild-bootstrap-t method throughout the analysis. As recommended by Cameron et al. (2008), I impose the null hypothesis of no effect and use Rademacher weights for resampling. Regression tables will report p-values obtained from the bootstrapped distribution of the test statistics.

2.4 Data

The fieldwork for this project took place between September and October 2012 in the city of Kombolcha, on the main road between the Ethiopian capital Addis Ababa and Mekelle, in the South Wollo province of Amhara region. According to the 2007 census, the city has a population of about 59,000. Background qualitative fieldwork at the onset of the project and descriptives from the survey data reveal that reliance on job-contact networks in Kombolcha is indeed extensive. Further, in recent years, Kombolcha has benefited from a number of industrial developments. The city can now count on an expanded textile factory, a metalwork factory, a large brewery, as well as smaller firms working on the processing of leather and seeds. This expansion of the formal sector makes Kombolcha an ideal place to study job-referral networks in Ethiopia. As constraints on the number of available jobs are progressively relaxed, asymmetric job-referral networks can exclude poorly connected individuals and groups from the new economic opportunities.

The sampling strategy is based on the following steps. First, based on qualitative fieldwork and discussions with local officials, I select three low-to-middle income urban, residential neighbourhoods.[22] I identify all residential blocks of houses in each neighbourhood using the Google-Earth map of the city and randomly sample 19 of these blocks. In each block, I list and interview all individuals in the age group 20-40 resident in the block at the time of fieldwork. I invite all interviewed individuals to take part in the experiment. Out of 518 individuals that are interviewed, 447 take part

[22] I exclude the other three neighbourhoods either because they include rural agglomerates or because they have few dwellers.

in the experiment.[23]

The sampling strategy enables me to capture block level networks. I have reason to believe this is a focal domain of interaction. There is substantial support for this assumption in the empirical literature.[24] In my sample, I observe that 21 percent of pairs of individuals in a block who know each other exchange information or referrals.[25]

To collect network information, I present individuals with a list of all people in the 20-40 age group residing in the block and ask them to identify the people they know.[26] For each person they know, respondents answer questions regarding the strength of the tie and various dimensions of social interaction: asking, giving and receiving job information; giving and receiving referrals; borrowing and lending; gift exchanges. Subjects that know each other have spoken, on average, on 12 days of the previous 30 days (see figure 2.10 in the appendix for the full distribution of this variable). In about 65 percent of cases, respondent i defined respondent j as a 'worship-place acquaintance', in 12 percent of cases as a member of the same family and in 14 percent of cases as a close friend.

[23]Table 2.11 in the appendix shows that there are few statistically significant differences in observable characteristics between individuals who take part in the experiment and individuals who do not. Selected individuals are less likely to be Muslim and tend to report a higher number of links in the neighbourhood. When I replace self-reported links with the number of times an individual has been mentioned by others as a friend, the network degree variable becomes balanced across the two groups.

[24]Marmaros and Sacerdote (2006), for example, show that geographical proximity is an important determinant of friendship among adolescents in the US. Their empirical strategy is particularly credible as it relies on random assignment to dorms. In a developing country setting, Karlan et al. (2009) report that 59 percent of observed ties among Peruvian shanty-town dwellers are between neighbours, while Fafchamps and Gubert (2007) document that geographic proximity is a strong predictor of risk sharing behaviour. Looking more specifically at labour market settings, Bayer et al. (2008) analyse census data on the Boston metropolitan population and find a strong, significant effect of shared block residence on the probability of working in the same census tract. Similarly, Hedstrom et al. (2003) document peer effects in the duration of unemployment among Stockholm youth. Their proxy of peer group is also given by small geographical units where a median number of 66 young people reside. Finally, Topa (2001) shows that unemployment pattern in Chicago's neighbourhoods are consistent with a peer effect model, albeit he focuses the analysis on geographical areas far larger than blocks.

[25]Overall, the empirical data consists of 15, 588 block-level dyads among 518 individuals in 19 blocks. In 1,804 cases i knows j and in 377 cases i declares to have given to or received from j job information or a referral.

[26]I include in the list individuals who reside in the block but were not available for interview (in most cases, they were out of town) at the time of fieldwork

In figures 2.8 and 2.9 I briefly summarise the data about job-contact networks. Figure 2.8 shows the structure of the network in one of the blocks in the sample. Blue squares represent individuals. Lines represent the job contact links.[27] The isolated squares on the left of the picture represent the individuals with no block-level job-contact links. I define the number of links an individual has as his or her degree. In figure 2.9 I plot the degree distribution of the survey respondents for two different networks: the knowledge network and the job-contact network.[28] In the figure on the left I define links using either the respondent's report or the reports of the other individuals in his block. In the panel on the right, I show the distribution of links which both respondent i and respondent j have reported, and links which at least one respondent has reported. In table 2.7 I present the summary statistics of all dyadic variables for individuals residing in the same block. Two descriptive findings emerge from this data. First, inter-personal economic interaction at residential block level is substantial. On average, a block dweller knows 11 percent of his neighbours, has received job-search assistance by 19 percent of the neighbours he knows and a loan by 14 percent of them. Second, even using the widest definition of a job-contact link, I find that almost 30 percent of individuals have *no* connections to the job-contact network of their residential block.

The survey includes standard socio-demographic variables such as household characteristics, age, gender, ethnicity, education, and migration status. A detailed module on labour market experience is administered, capturing employment status, job characteristics, search strategies (while in unemployment and on-the-job search) and referrals. A final module investigated expectations regarding employment, wage and unemployment exit rates.

I report here some summary statistics for the characteristics of the experimental

[27] In this figures, I include an undirected job-contact link if respondent i has reported any exchange of information or referrals from respondent i to respondent j, or from respondent j to respondent i.

[28] In the job contact network I report a link if respondent i has exchanged information or referrals with respondent j or has asked for his assistance in finding a job.

sample.[29] I am not able to match 17 IDs recorded in the experimental forms with a questionnaire.[30] I hence have 447 observations for experimental variables, and 430 observations for individual variables. Furthermore, 1 individual decline to respond the question about religion and 9 individuals decline to respond the question about migration status.

Table 2.2: Summary statistics: Binary variables

Variable	Proportion	Std. Dev.	N
Male	0.488	0.5	430
Muslim	0.536	0.499	429
Migrant	0.223	0.417	421
Employed	0.388	0.488	430
Inactive	0.374	0.485	430

Table 2.3: Summary statistics: Continuous variables

Variable	Mean	Std. Dev.	N
Age[31]	27.716	8.473	416
Education	9.606	4.394	430
Earnings	433.356	801.825	430
Real network degree	6.047	4.088	430
Amount kept in DG	13.938	4.158	430

I check for covariates balance for both levels of randomisation. Tables 2.12 and 2.13 in the appendix show the result of balance tests employing OLS regressions. Each column corresponds to a regression of a different dependent variable on dummies for individual and session treatments. The column heading specifies the dependent variable that is being tested. Table 2.12 confirms that the observable characteristics of participants assigned to different levels of degree centrality are not statistical different.

[29] I categorise workers as employed, unemployed, and inactive. A worker who currently has a job is classified as employed. A worker who does not have a job, has been searching for a job in the past seven days, and is currently available for work is defined as unemployed. A worker who does not have a job and either is not available for work or has not been searching for work in the past seven days (or both) is considered inactive. Also I report a variable called 'real network degree'. This is defined as the self-reported number of social ties with residents of the same block in the age group 20 to 40.

[30] While I have no particular reason to suspect that these individuals infringed the rules of the game in any way, I still exclude them from the analysis presented below. Inclusion of these 17 observations does not change any of the results.

Table 2.13 shows that there are some weak differences across individuals in the T1 and T2 treatments: individuals in the T1 treatments are more likely to be male and less central in their real-life network. These differences are not large and only significant at the 10 percent level. Finally, tables 2.14 and 2.15 show that there are no statistically significant differences in observables between individuals in T2a and in T2n and that individuals in T1n are more likely to be muslim and to and have a higher degree in their real networks, compared to the anonymous T1a condition. These differences are significant at the 10 percent level. In regression analysis, I include controls for the unbalanced characteristics. As religion is one of these, and I have one missing observation, my final sample comprises 429 individuals. All analysis from this point onwards refers to this sample.

I report the distribution of sessions and individuals across treatments in table 2.9 in the appendix. The last column of the table shows the number of dyads which I obtain for the individuals who participate.

I test understanding with five questions. For these I have complete responses for 426 of the 429 subjects in the final sample.[32] The first three questions deal with understanding of the network map. Very few people answer these questions wrongly. The last two questions test understanding of the relevant incentives.[33] There is somewhat more variation here and I hence create a binary variable for whether the participant answered both of these questions correctly. Reassuringly, about 80 percent of participants choose the right answer in both questions.

Table 2.16 in the appendix shows the result of a linear probability model where the understanding variable is regressed over a number of session treatment dum-

[32]Overall, I have complete responses for 444 subjects out of 447.

[33]In the T1 treatments, this is the probability of receiving a referral from player with different degree centrality. Participants are asked whether they would be more likely to receive a referral from a degree X or a degree Y player. In the T2 treatments I focus on the probability that a player with a given level of degree centrality would be referred for a job when unemployed. I ask participants whether a player of degree X is more likely to be referred for a job than a player of degree Y.

mies.[34] The first column includes dummies for four treatments: T1a is the residual category. The other treatments do not have a significantly different proportion of high-understanding subjects. The third column shows that in the T1a2 treatment, where no explanation of the relevant incentives is given, understanding of the incentives resulting from the structure of the network is not significantly lower than in other treatments. 77 percent of participants in the T1a2 treatment answer both questions correctly.

If participants do not want to express a preference for the two links, they have the option of writing the letter R. In this case a link is be picked at random for them.[35] A high percentage of random decisions could be interpreted as a signal of poor understanding. The data dispels such concern. The random link option is used in only 12 percent of decisions and only 8 percent of participants choose a random link twice. Furthermore, the likelihood of choosing a random link is not statistically different across treatment.

Taken together, these results reassure me that the experiment was well understood.

2.5 Results

Table 2.10 in the appendix describes the decisions of players in the link-formation experiment. I organise the discussion around four central results.

Result 2.1. *In the T1 treatments, links with degree-one players are significantly more likely than links with more connected players.*

Table 2.4 below shows results from estimation of the dyadic regression model (2.7). A connection to a degree-two partner is 17 percentage points less likely, and a connec-

[34] Standard errors are clustered at the session level, as in all other regressions in this chapter.

[35] Participants can write R in place of either of the links. Leaving the box blank is interpreted in the same way as the letter R.

tion to a degree-three partner is 20 percentage points less likely, than a connection to a degree-one partner. These effects are highly statistically significant and represent a broad confirmation of prediction 2.1. The understanding tests show that participants are aware of the incentives that arise from competition for referrals in a network where links are distributed unequally. Participants' decisions are thus consistent with a desire to maximise the chance of being referred for a job when unemployed.

Table 2.4: Linear probability model: T1 treatments

	(1)	(2)	(3)
$degree_j = 2$	-.170 (.002)***	-.170 (.002)***	-.184 (.018)**
$degree_j = 3$	-.198 (.002)***	-.199 (.002)***	-.305 (.006)***
T1n (identity disclosed)			-.019 (.699)
T1n * $degree_j = 2$			.021 (.787)
T1n * $degree_j = 3$			.131 (.086)*
T1a2 (no explanations)			-.057 (.551)
T1a2 * $degree_j = 2$			.021 (.879)
T1a2 * $degree_j = 3$			.199 (.272)
Const.	.399 (.000)***	.399 (.000)***	.423 (.000)***
Obs.	1534	1534	1534
Cluster N	30	30	30
Controls		✔	✔

Dyadic OLS regression. The dependent variable is a dummy which takes a value of one if player i wants to form a link with player j. Columns 2 and 3 include controls for gender, religion, and degree in the residential-block network. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors are corrected for clustering at session level. P-values obtained with wild bootstrap-t procedure are reported in parentheses.

In column 3 I test whether play differs when personal identity is disclosed or when incentives are not explained. The latter has no statistically significant effect on the likelihood of choosing a degree-one individual. Understanding of the incentives occurs without explicit explanation and such explanation does not seem to induce any significant change in behaviour. Disclosing personal identity increases somewhat the number of degree-three individuals who are chosen. This is possibly due to additional motives behind the decisions taken in non-anonymous treatments, for example, choosing real friends.[36]

The main effect is unchanged when I analyse the behaviour of high-understanding players (column 1 of table 2.17 in the appendix).[37] High-understanding players are less likely to choose degree-two and degree-three partners than low-understanding players (column 2). Participants who send a positive amount in the dictator game do not respond differently to the degree of potential partners (column 3).

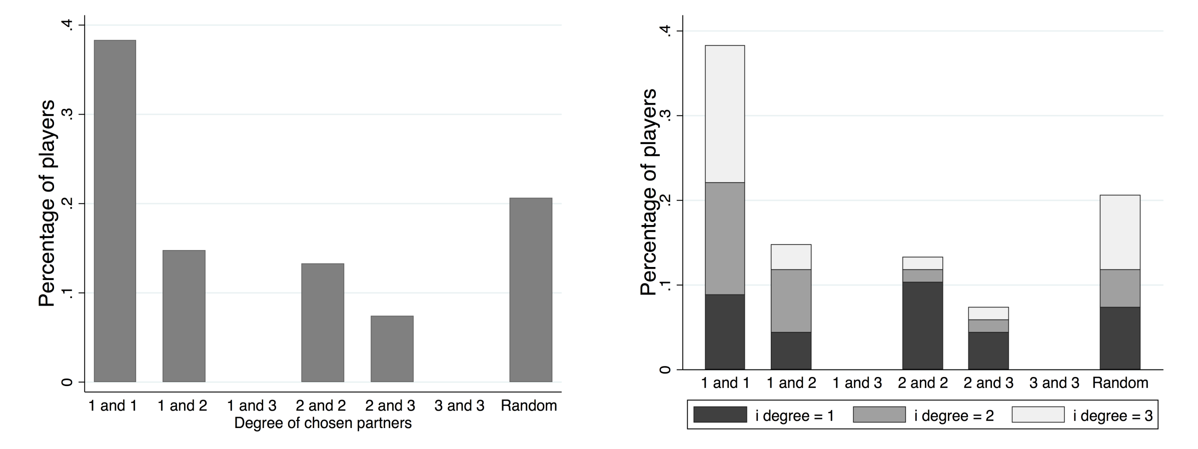
Not all participants choose to link with degree-one players in the T1 treatments. These players may be simply making mistakes, or they may be concerned with objectives other than the maximisation of expected payoff. To explore this second possibility further, I will study the strategies of high-understanding subjects, who are arguably less likely to commit mistakes. Furthermore, a strategy specifies two separate links. By analysing two decisions at a time, I can look for consistent decision making patterns. Figure 2.2 shows the relative frequencies of the strategies played by high-understanding players.

I find that the strategy that assigns the two links towards degree-one players is the most frequently chosen. The strategy of choosing a degree-one and a degree-two player is played less often, and the strategy of choosing a degree-one and a degree-three player is never chosen. There is however a significant group of players who specifies no link to a degree-one peer. Most of these players are themselves degree-one. In the prediction section for T2, I discussed the possibility that inequality-averse degree-one players may not want to link to degree-one peers to avoid an increase in the number of players who are better off than themselves. While in the T1 condition links

[36]I further explore this effect in the analysis below.

[37]In the specification reported in column 1 I interact the dummies for $\text{degree}_j = 2$ and $\text{degree}_j = 3$ with a dummy for whether player i has correctly answered both questions about understanding of the incentives. I include a separate dummy for whether player i has high understanding, but do not include separate dummies for the degree of player j. In this specification, the coefficient on the interaction captures whether an high-understanding player is less likely to choose a degree-two (or degree-three) player than to choose a degree-one partner.

Figure 2.2: Individual strategies for high-understanding subjects in T1a



(a) Aggregate (b) By i's position in the network

The category 'Random' includes both strategies that specify two random links, and strategies that specify only one random link.

do not influence the expected outcomes of the partner they are directed to, players may still care about the prestige that comes with having a high number of connections. If so, individuals may dislike increasing the number of better connected players even in T1.[38]

Overall, 47 percent of all links in the T1 treatments are directed towards a degree-one player. This has important welfare consequences. To fully tease this out I have to do some calculations. Recall that only the links specified by a single, randomly drawn unemployed player will be activated. I hence calculate the expected number of new links for a player i who has been chosen by b peers using the following formula:[39]

[38] I am unable to find evidence that dictator game giving correlates significantly with this type of behaviour in T1.

[39] I count all links to player i chosen by the other players, but do not consider the two links specified by player i himself, as the chance of obtaining these two links is not affected by play in the game. N is the set of players and n the cardinality of this set (n=8). B is the set of players who have chosen a link to player i and b the cardinality of this set. The first and third term in (2.9) are intuitive. The second term captures the probability that, given a total of z players who are unemployed, exactly w of the players who have chosen player i are unemployed. This probability is given by a fraction. In the denominator I have the $\binom{N}{z}$ possible combinations of z unemployed peers to choose from. In the numerator I have the product of two terms. First, the $\binom{B}{w}$ possible ways of choosing w players from the set B. Second, the $\binom{N-B}{z-b}$ ways of choosing the remaining $z-b$ unemployed individuals from the complement of B (the set of player who have not chosen a link to player i).

$$\sum_{z=0}^{n}\sum_{w=0}^{\min(z,b)} \underbrace{\left(\binom{N}{z}p^{n-z}(1-p)^{z}\right)}_{\text{Pr(z peers are unemployed)}} * \underbrace{\binom{B}{w} * \binom{N-B}{z-w} * \binom{N}{z}^{-1}}_{\text{Pr(w peers who have chosen i are unemployed)}} * \underbrace{\frac{w}{z}}_{\text{Pr(one of the w peers is drawn)}} \tag{2.9}$$

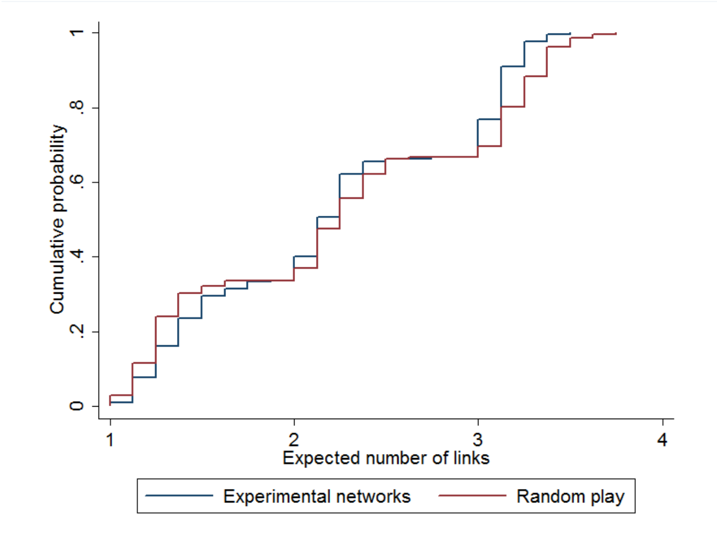
I sum the expected number of new links (2.9) to the original number of links and obtain the *total expected number of links*. I compare this number in the T1 treatments and in a simulation where every individual chooses new partners randomly, irrespective of their position in the network. I find that the number of expected links of degree-one players in T1 is 7.5 percent higher than the expected number of links under random link formation. This is a large effect, which corresponds to 63 percent of a standard deviation.

The distribution of expected links in the experimental networks is more equal (less dispersed) than in the simulated networks with random play.[40] The standard deviation drops by 9 percent and, as shown in figure 2.3, the experimental distribution of links (almost) weakly second order stochastically dominates the distribution of links under random play. A relationship of second order stochastic dominance implies a more equal distribution.

Let me now turn to the results of the T2 treatments. I start by studying behaviour in the dictator game, which I use to measure the prevalence of other-regarding concerns in the population. Figure 2.4 summarises the data: giving is substantial. About 82 percent of players send a positive amount to their partner. The mean amount sent is 6 Ethiopian Birr, which is about a third of the endowment, and is in line with the experimental evidence in other countries (Camerer, 2003). I interpret this result as evidence of substantial and widespread other-regarding preferences in the population. Furthermore, I note that the modal amount sent is half of the endowment. This potentially

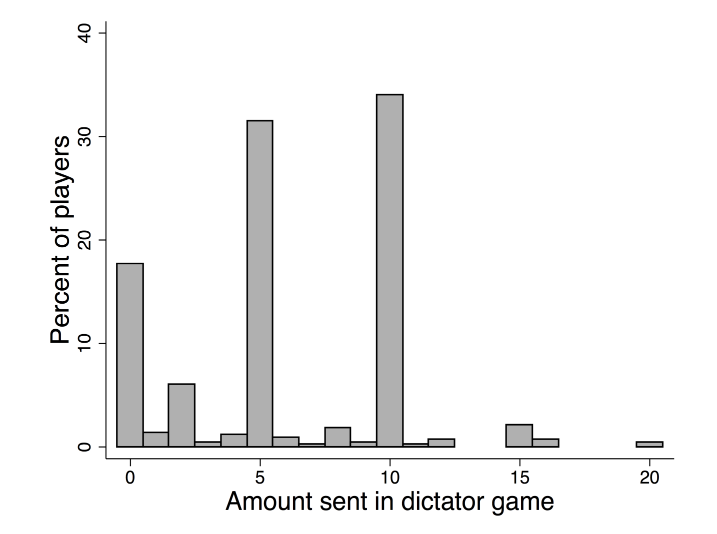
[40] As explained above, I am referring to the *total expected number of links*.

Figure 2.3: Cumulative distribution of total expected links in real and simulated networks



reflects a strong preference for equality.

Figure 2.4: Giving in the dictator game



Result 2.2. *82 percent of players send a positive amount in the dictator game. The equal split of the endowment is the modal choice.*

I now estimate model (2.7) with data from the T2a and T2n conditions and report the results in table 2.5. I find that the coefficients on the dummies for player j's degree are small and insignificant (columns 1 and 2). Subjects in non-anonymous treatments are more likely to link with degree-two and degree-three players than subjects in anonymous treatments, but this difference is not significant (column 3). In T2a, player i is 13.8 percentage point less likely to choose a degree-two player than a degree-one player and 5.5 percentage point less likely to choose a degree-three player. Again, these effects have p-values above conventional levels of significance (column 3). In short, I am unable to find statistically significant evidence to support prediction 2.2.

Result 2.3. *In the T2 treatments, links with degree-one players are* not *significantly more likely than links with better-connected players.*

Table 2.5: Linear probability model: T2 treatments

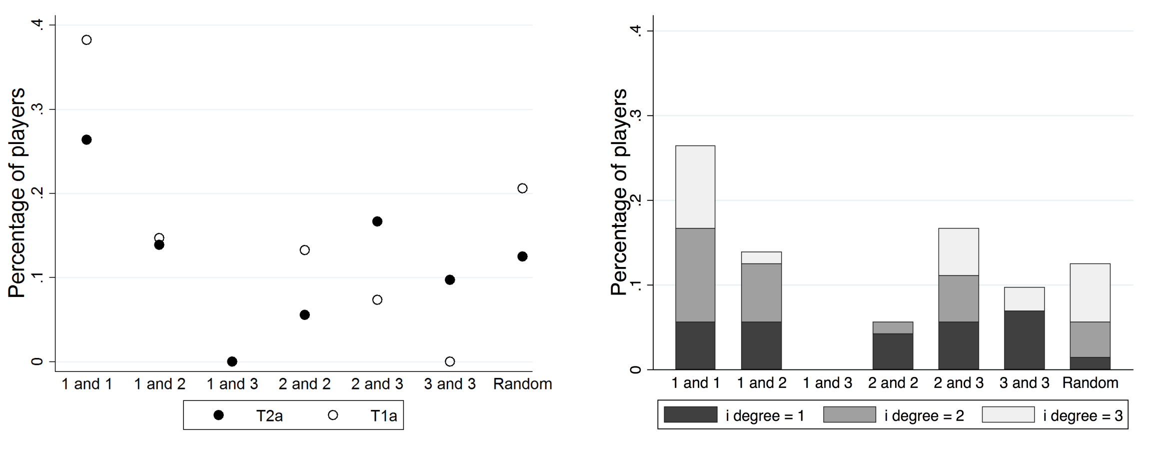
	(1)	(2)	(3)
$\text{degree}_j = 2$	-.068 (.226)	-.068 (.226)	-.138 (.162)
$\text{degree}_j = 3$	.013 (.847)	.012 (.859)	-.055 (.643)
T2n (identity disclosed)			-.058 (.557)
T2n * $\text{degree}_j = 2$			.138 (.248)
T2n * $\text{degree}_j = 3$			.135 (.396)
Const.	.303 (.000)***	.274 (.000)***	.332 (.000)***
Obs.	1022	1022	1022
Cluster N	20	20	20
Controls		✔	✔

Dyadic OLS regression. The dependent variable is a dummy which takes a value of one if player i wants to form a link with player j. Columns 2 and 3 include controls for gender, religion, and degree in the residential-block network. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors are corrected for clustering at session level. P-values obtained with wild bootstrap-t procedure are reported in parentheses.

In table 2.18 in the appendix I test whether I can find limited support for prediction 2.2 for individuals who understood the game well. This turns out not to be the case.[41]

[41] Again, in column 1 I am omitting the dummies $\text{degree}_j = 2$ and $\text{degree}_j = 3$. So the coefficients on the interactions tell me whether high-understanding players are significantly less likely to choose $\text{degree}_j = 2$ and $\text{degree}_j = 3$ partners.

Figure 2.5: Individual strategies for high-understanding subjects in T2a



(a) Aggregate (b) By i's position in the network

The category "Random" includes both strategies that specify two random links, and strategies that specify only one random link.

Surprisingly, individuals who send a positive amount in the dictator game are more likely to choose links with degree-two or degree-three partners. Neither of these effects is significant, however. In any case, this falsifies prediction 2.3, which conjectured a correlation in the opposite direction.

36 percent of all links in the T2 treatments are directed towards the degree-one players. This is an 11 percentage point drop compared to the T1 treatments. The fraction of high-understanding players who choose two links with degree-one players in the anonymous treatment also falls by 9 percentage points. Furthermore, 10 percent of players in the T2a treatment choose two links with degree 3 peers. No high-understanding player chose this strategy in T1a.

The fraction of links to degree-one players decreases irrespective of the network position of player i and of whether personal identities are disclosed. I show this by estimating the following regression models:

$$\text{link to degree-one player}_{zi} = \alpha + \beta_1 T2a + \beta_2 T2n + \beta_3 NA + e_{zi} \tag{2.10}$$

$$\text{link to degree-one player}_{zi} = \alpha + \sum_{n=1}^{3} \beta_n (T2 * dn_i) + \beta_4 d2_i + \beta_5 d3_i + e_{zi} \tag{2.11}$$

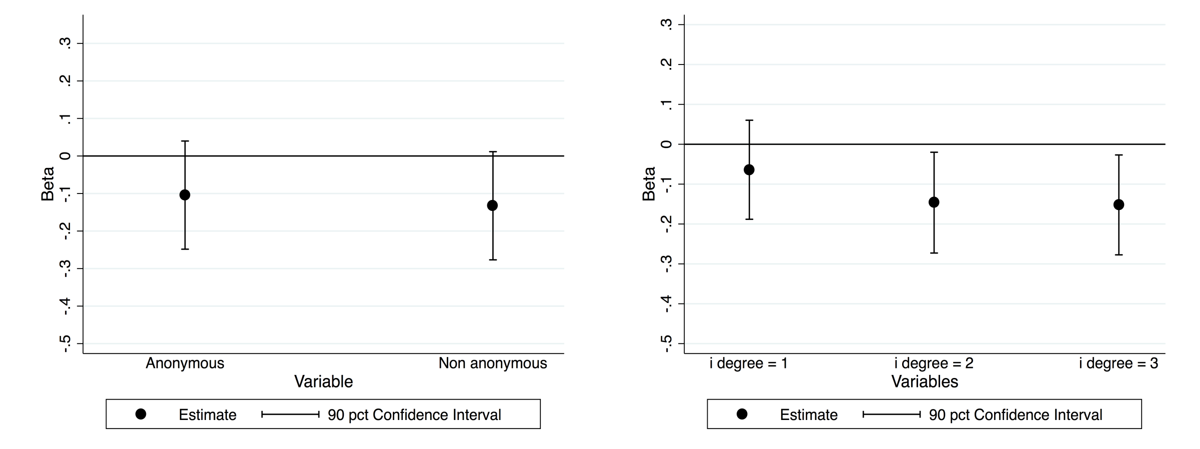
z refers to a link, i to a player, and s to a session. On the left hand side is an indicator variable which takes a value of one if link z by individual i is a link to a degree-one partner. NA_s is a dummy for non-anonymous sessions and $d1_{is}$, $d2_{is}$, $d3_{is}$ are dummies for player i's degree in the network. In model (2.10), coefficient β_1 captures the difference in the likelihood of a link to a degree-one player between the T2a and the T1a conditions, while coefficient β_2 captures the same difference between T2n and T1n. In model (2.11), coefficients β_1, β_2 and β_3 estimate the difference in the likelihood of a link to a degree-one player between the T2 and T1 conditions for decision makers of degree one, two and three, respectively. I estimate both models and report point estimates and confidence intervals in figure 2.6 below.[42]

Figure 2.6a shows that the probability of a link with a degree-one player drops by a similar amount in both the anonymous and non-anonymous conditions. Further, figure 2.6b shows that the probability of choosing a link with a degree-one player decreases for individuals in all initial network positions. Point estimates are larger and significant at the 10 percent level for degree 2 and degree 3 players. Thus, the drop in the proportion of links with degree-one partners in T2 is not driven by inequality-averse degree-one individuals who do not want to become envious of other degree-one players.

I further investigate the fall in links to degree-one players by separately analysing individuals who send a positive amount in the initial dictator game and individuals

[42]Full regressions tables are reported in tables 2.19 and 2.20 in the appendix

Figure 2.6: Regression coefficients of linear probability model: probability of choosing a degree-one player



(a) Model 2.10 (b) Model 2.11

All regressions include controls for gender, age and degree in the residential-block network. Standard errors are clustered at the session level.

who keep the whole endowment. For this purpose, I estimate the following models:

$$\text{link to degree-one player}_{zi} = \alpha + \beta_1 \left(T2 * sent > 0_i\right) + \beta_2 \left(T2 * sent = 0_i\right) + \beta_3 sent > 0_i + e_{zi} \tag{2.12}$$

$$\begin{aligned} \text{link to degree-one player}_{zi} = & \alpha + \beta_1 \left(T2a * sent > 0_i\right) + \beta_2 \left(T2a * sent = 0_i\right) \\ & + \beta_3 \left(T2n * sent > 0_i\right) + \beta_4 \left(T2n * sent = 0_i\right) \\ & + \beta_5 sent > 0_i + \beta_6 NA + e_{zi} \end{aligned} \tag{2.13}$$

With model (2.12), I check whether the T2/T1 fall in the probability of linking with a degree-one player separately occurs for both players who send a positive amount and players who send nothing. With model (2.13), I repeat this analysis keeping anonymous and non-anonymous sessions separate. Results are reported in figure 2.7 below.[43]

I find that individuals who send a positive amount in the dictator game are 12 percentage points less likely to choose degree-one players in T2 compared to T1. Surprisingly, individuals who have not sent anything are, on the other hand, almost equally

[43] Full regressions tables are reported in tables 2.21 and 2.22 in the appendix

likely to choose a degree-one player than individuals in T1.[44] When I further split the data between the anonymous and non-anonymous conditions, I find that this pattern is only observed for the anonymous conditions. However, I cannot reject the equality of the β_1 and β_2 coefficients. The β_3 and β_4 coefficients, referring to the non-anonymous condition, are similar in magnitude (and statistically indistinguishable).

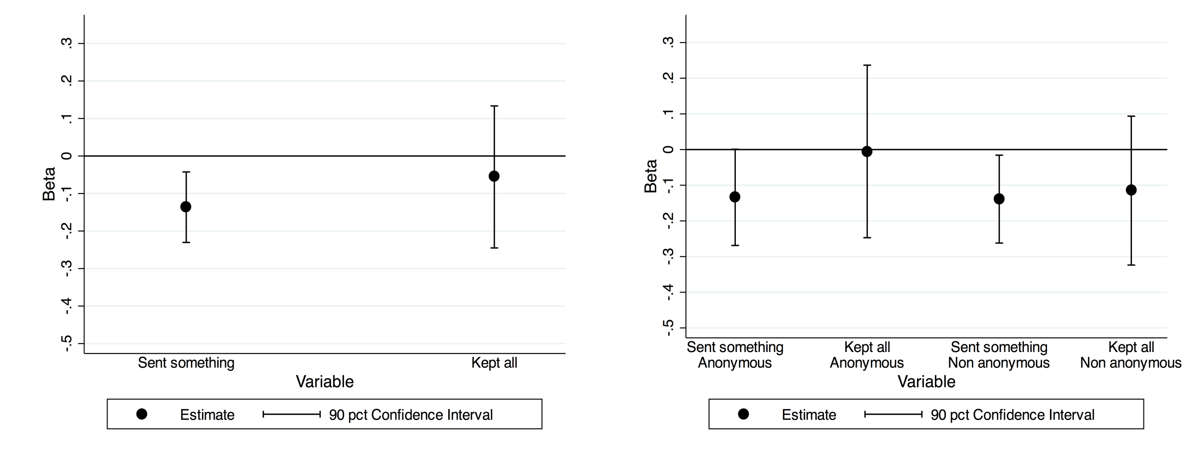
This result is striking. Subjects who transfer resources to others at a personal cost in the dictator game are less likely to choose degree-one partners when links have no effect on personal payoff. Subjects who do not make any transfer in the dictator game are, on the other hand, equally likely to pick degree-one partners in the T1 and in the T2 conditions.

I put forward three possible explanations for this result. First giving in the dictator game reflects conformity to social norms rather than preferences (Bicchieri, 2006). While the equal split is very often perceived as the fair allocation in distribution problems, norms that rule link-formation behaviour may actually encourage preferential attachment to high-degree players. Preferential attachment of this kind is commonly observed in empirical networks. It is thus not unconceivable that field norms prescribe preferential attachment. Second, individuals are guilt averse as in Charness and Dufwenberg (2006) and they believe that (i) dictator-game receivers expect to be sent a positive amount, and (ii) high-degree players expect to be proposed at least some links. Third, giving in the dictator game satiates the other-regarding inclinations of givers.

The first and second explanations rest on the existence of norms and expectations which I do not observe directly. Furthermore I cannot offer an independent test of the satiation hypothesis. As a result, I cannot rigorously assess the relevance of these potential explanations. However, I flag one piece of evidence which appears consistent

[44] A Wald test however cannot reject the null hypothesis of the equality of the two coefficients at the 15 percent confidence level.

Figure 2.7: Regression coefficients of linear probability model: probability of choosing a degree-one player



(a) Model 2.12 (b) Model 2.13

All regressions include controls for gender, age and degree in the residential-block network. Standard errors are clustered at the session level.

only with the mechanisms based on norms or expectations: individuals who send a positive amount in the dictator game are 11 percentage points more likely to choose a degree-three partner in T2, while they are virtually equally likely to choose a link with a degree-two partner or a random link.

Finally, I study whether in non-anonymous treatments individuals choose players whom they know in real life, and whether they take into account the characteristics of the other players.[45] My first prediction is that knowing individual j will attract the links of player i in the T2n treatment, but not in T1n. Estimates are reported in table 2.6.

Subjects in T2n are more likely to link with individuals they know in real life. In a specification without controls or interactions, knowing player j is associated with a 25 percentage point increase in the likelihood of choosing a link him. The effect is significant at the 10 percent and is of a large magnitude: it amounts to almost a doubling of the likelihood that player i chooses a player j he does not know.

[45] As explained above, I lack survey data about 17 individuals. ij dyads where these individuals feature as player i have been excluded from the analysis so far. I will now also exclude dyads where these 17 individuals feature as j.

Contrary to what hypothesised, subjects in T1n are also more likely to link with the players they know. Personal knowledge of player j raises the probability of a link by about 13 percentage points. This is surprising as in this game knowing one's new partners brings no benefit to either the proposer or the receiver of the link, and the identity of the proposer remains private information. One possibility is that choosing real friends may be a heuristic which economises on cognitive costs. While the coefficients in the T1n and T2n regressions differ in magnitude, a Wald test cannot reject the null hypothesis that the two coefficients are equal.

As a robustness check, I repeat the exercise in table 2.6 for the three anonymous treatments. Here player identities are not disclosed and hence knowing a particular player j in real life should not affect player i's propensity to link to him. Table 2.23 in the appendix confirms that personal knowledge is a not a significant predictor of link formation in any of the anonymous treatments.

Result 2.4. *In T1n and T2n, subjects are more likely to link with they players they know in real life.*

Table 2.6: Linear probability model: non-anonymous treatments

	(1)	(2)
$\text{degree}_j = 2$	-.160 (.028)**	-.005 (.967)
$\text{degree}_j = 3$	-.162 (.002)***	.078 (.410)
i knows j	.137 (.034)**	.251 (.078)*
Const.	.390 (.000)***	.262 (.000)***
Obs.	596	498
R^2	.037	.024
Treatment	T1n	T2n
Cluster N	11	10

Dyadic OLS regression. The dependent variable is a dummy which takes a value of one if player i wants to form a link with player j. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors are corrected for clustering at session level. P-values obtained with wild bootstrap-t procedure are reported in parentheses.

Once identities are disclosed, subjects may also take into account individuals' real characteristics. For example, players with a special concern for the welfare of the least well off may choose to target their links to individual whom they know to be socially isolated or unemployed in real life. In table 2.24 I explore this possibility by including on the right hand side variables that measure the number of links in social and job-contact networks and a dummy for whether the individual is employed. For each characteristic x, I estimate a model of this form:

$$link_{ij} = \alpha + \beta_1 d2_j + \beta_2 d3_j + \beta_3 \text{i knows j} + \beta_4 \text{x}_j + \beta_5 \left(\text{i knows j* x}_j\right) + e_{ij} \quad (2.14)$$

For the employment dummy, I imprecisely estimate a negative coefficient β_5 in both T1n and T2n (columns 3 and 6). Player i is 3.4-5.5 percentage points less likely to link with a employed player he knows in real-life than with a non-employed known player. This result is consistent with the story that altruist also consider the characteristics of players outside of the lab.

For ties in the job-contact network, the estimated coefficient is not consistent with altruistic motives: among known potential partners player i prefers those with a higher number of connections in job-contact networks. The effect is large in magnitude, but again, is not significantly estimated.

I further test whether individuals who send a positive amount in the dictator game target their links to socially isolated or non-employed peers in T2n. I do this by interacting the dyadic variable of interest with dummies capturing whether player i sends a positive amount in the dictator game or sends nothing:

$$
\begin{aligned}
\text{link}_{ij} = &\alpha + \beta_1 c2_j + \beta_2 c3_j + \beta_3 \text{i knows j} + \beta_4 \text{x}_j \\
&+ \beta_5 \left(\text{i knows j} * x_j * sent > 0_i\right) \\
&+ \beta_6 \left(\text{i knows j} * x_j * sent = 0_i\right) \\
&+ \beta_7 sent > 0_i + e_{ij} \qquad (2.15)
\end{aligned}
$$

The results I report in table 2.25 in the appendix show no statistically significant effects.

2.6 Conclusion

Using a novel experiment from urban Ethiopia, I study whether individuals form new links with the least-connected players in a network. In my design, links with the least-connected individuals maximise the expected payoff of the player proposing the link. Further, least connected players have the smallest expected payoff in the network and experience the largest increase in expected payoff from a new link. The first feature should motivate selfish players, while the last two features should be of concern to other-regarding players. Manipulation of the rule determining when new links are added to the network allows me to separate self-regarding and other-regarding motives for link formation.

I find broad support for selfish network formation. In treatments where links primarily affect the welfare of the player proposing the tie, new links are mostly formed with the least-connected individuals. In treatments where links affect the welfare of the second player in the tie, links with the least-connected individuals are less common, and not significantly more likely than ties to better-connected players. Surprisingly, this effect is driven by players who send a positive amount in an initial dictator game. In sessions where I disclose personal identity, I further document that subjects prefer links with the players they know in real life. Among known peers, links with the unemployed and the well-connected in real job-contact networks are more likely.

The participants in this experiment do not seem intrinsically motivated to strengthen the position of the least well off individuals in the network. This carries policy implications as job-contact networks that are persistently asymmetric can create economic inequality.[46] Interventions can try to weaken reliance on informal channels that distribute information and referrals. For example, this can be done through the introduction of anti-nepotism policies or the strengthening of formal systems of job search and personnel selection. Alternatively, when the economic function of informal channels cannot be easily replaced by formal intermediation, interventions can try to promote the social integration of poorly connected individuals. Mentoring programs, for example, create links between individuals at different stages of their career. Moreover, employment programmes can allow participants to refer other individuals for the next round of jobs, in order to help participants reinforce their job-contact networks. Strengthened networks would in turn prolong the welfare gains associated with participation in the programme.

[46] The external-validity caveats that apply to all lab-in-the-field experiments should of course be considered as well.

Appendix

A.1 Derivations

A.1.1 From equation (2.3) to (2.4)

$$\sum_{z=0}^{n_k(g)-1} \binom{n_k(g)-1}{z} \frac{p^{n_k(g)-1-z}(1-p)^z}{z+1}$$

Define $m \equiv n_k(g) - 1$, so that:

$$\sum_{z=0}^{m} \binom{m}{z} \frac{p^{m-z}(1-p)^z}{z+1}$$

Multiply by (1-p)(m+1):

$$(1-p)(m+1)\sum_{z=0}^{m} \binom{m}{z} \frac{p^{m-z}(1-p)^z}{z+1}$$

$$(1-p)(m+1)\sum_{z=0}^{m} \frac{m!}{z!(m-z)!} \frac{p^{m-z}(1-p)^z}{z+1}$$

$$(1-p)(m+1)\sum_{z=0}^{m} \frac{m!}{(z+1)!(m-z)!} p^{m-z}(1-p)^z$$

$$\sum_{z=0}^{m} \frac{m+1!}{(z+1)!(m-z)!} p^{m-z}(1-p)^{z+1}$$

$$\sum_{z=0}^{m} \frac{m+1!}{(z+1)!(m+1-(z+1))!} p^{m+1-(z+1)}(1-p)^{z+1}$$

Define $s \equiv z+1$:

$$\sum_{s=1}^{m+1} \frac{m+1!}{s!(m+1-s)!} p^{m+1-s}(1-p)^s$$

The following expression is the sum of the probability mass function of the binomial distribution over its support, which has to be equal to one:

$$\sum_{s=0}^{m+1} \frac{m+1!}{s!(m+1-s)!} p^{m+1-s}(1-p)^s = 1$$

Thus:

$$\begin{aligned}\sum_{s=1}^{m+1} \frac{m+1!}{s!(m+1-s)!} p^{m+1-s}(1-p)^s &= 1 - \frac{m+1!}{0!(m+1-0)!} p^{m+1-0}(1-p)^0 \\ &= 1 - p^{n_k(g)}\end{aligned}$$

This has to be divided by $(1-p)(m+1)$, as I initially multiplied the expression by that number. Also: $m+1 = n_k(g)$. Thus:

$$\sum_{z=0}^{n_k(g)-1} \binom{n_k(g)-1}{z} \frac{p^{n_k(g)-1-z}(1-p)^z}{z+1} = \frac{1-p^{n_k(g)}}{(1-p)n_k(g)}$$

□

A.1.2 Proposition 2.1

To simplify the exposition $n_k(g)$ is abbreviated with n_k.

$$q(n_k) = \frac{1-p^{n_k}}{(1-p)n_k}$$

$$\begin{aligned}\frac{\partial q(n_k)}{\partial n_k} =& \frac{-p^{n_k}\ln(p)(1-p)n_k - (1-p^{n_k})(1-p)}{(1-p)^2 n_k^2} \\ =& \frac{-p^{n_k}\ln(p)n_k - (1-p^{n_k})}{(1-p)n_k^2} \\ =& \frac{p^{n_k} - p^{n_k}\ln(p)n_k - 1}{(1-p)n_k^2} \\ =& \frac{p^{n_k}(1-\ln(p)n_k) - 1}{(1-p)n_k^2}\end{aligned} \tag{2.16}$$

In the game I study $p \in (0,1)$ and $n_k \geq 1$. Thus the denominator of the fraction on the right hand side of equation (2.16) is positive. The whole fraction is negative when:

$$
\begin{aligned}
p^{n_k}(1-\ln(p)n_k)-1<0 \\
\ln(p)n_k > 1-\frac{1}{p^{n_k}}
\end{aligned}
\tag{2.17}
$$

When $n_k = 1$, this simplifies to:

$$\ln(p) > 1 - \frac{1}{p}$$

which holds for $p \in (0,1)$. As n_k increases beyond one, both sides of the inequality decrease. The left hand side decreases at a fixed rate given by $\ln(p)$. The right hand side decreases at a rate given by $\frac{1}{p^{n_k}}\ln(p)$. Note that $\frac{1}{p^{n_k}}\ln(p) > \ln(p)$ as $\frac{1}{p^{n_k}} > 1$: the right hand side decreases faster than the left hand side. Together with the fact that when $n_k = 1$ the right hand side is smaller than the left hand side, this shows that condition (2.17) holds. □

A.1.3 Proposition 2.2

In what follows, $n_i(g)$ is abbreviated with n_i. Furthermore, I assume for simplicity that in the initial network all players k have the same number of connections, so that $q(n_j(g)) = \bar{q}\ \forall\ j \in N_i(g)$. I know that $\bar{q} \in (0,1]$.[47] Thus:

$$
\begin{aligned}
Q_i(g) &= 1 - \prod_{j \in N_i(g)} (1 - pq(n_j(g)) \\
Q_i(g) &= 1 - (1 - p\bar{q})^{n_i}
\end{aligned}
$$

$$\frac{\partial Q_i(g)}{\partial n_i} = -(1-p\bar{q})^{n_i}\ln(1-p\bar{q}) \tag{2.18}$$

[47]When all individuals have only one link, $\bar{q} = 1$. As the number of links increases, q drops, but never reaches zero.

Notice that $0 < 1 - p\bar{q} < 1$ as $\bar{q} \in (0, 1]$ and $p \in (0, 1)$. This implies that $\ln(1 - p\bar{q}) < 0$ and that $(1 - p\bar{q})^{n_i} > 0$. The whole expression is multiplied by -1 and so $\frac{\partial Q_i(g)}{\partial n_i} > 0$, which proves the first part of proposition 2.

$$\frac{\partial^2 Q_i(g)}{\partial n_i^2} = -(1 - p\bar{q})^{n_i}(\ln(1 - p\bar{q}))^2 \tag{2.19}$$

Now $\left(\ln(1 - p\bar{q})\right)^2 > 0$ and $(1 - p\bar{q})^{n_i} > 0$. Hence $\frac{\partial^2 Q_i(g)}{\partial n_i^2} < 0$. □

A.2 Figures

Figure 2.8: The job-contact network of a residential block in urban Ethiopia

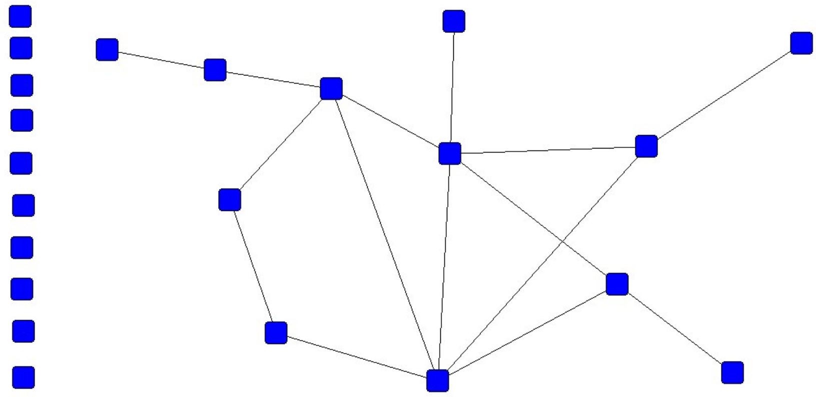
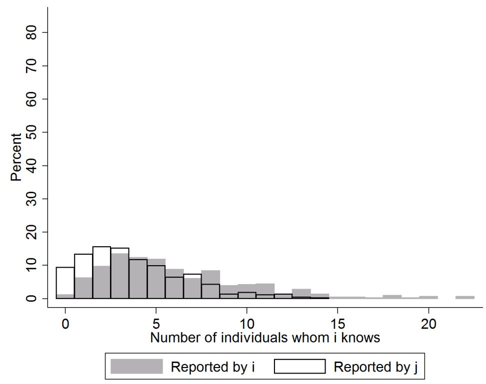
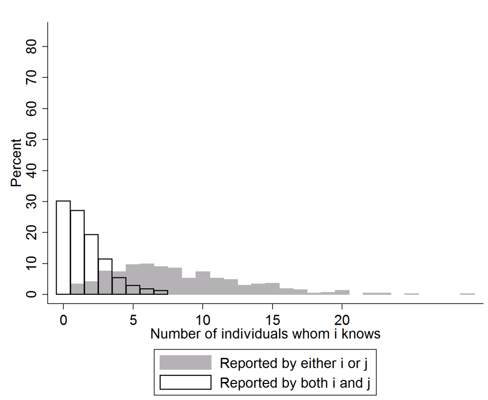


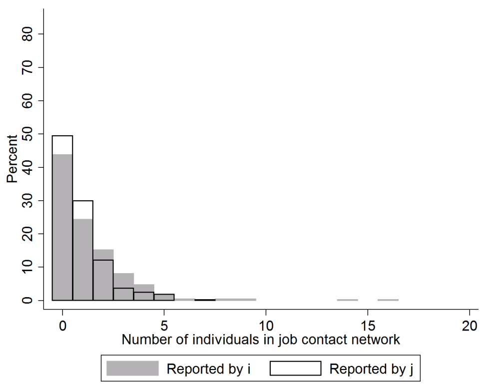
Figure 2.9: Distribution of links in real networks



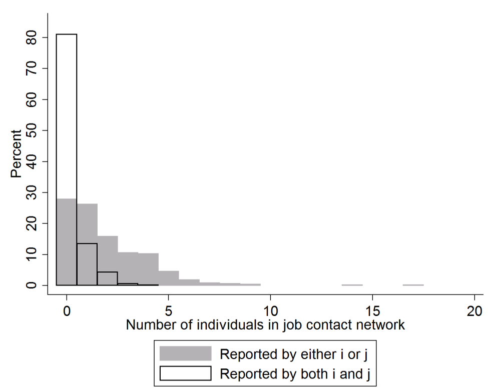
(a) Social network



(b) Social network



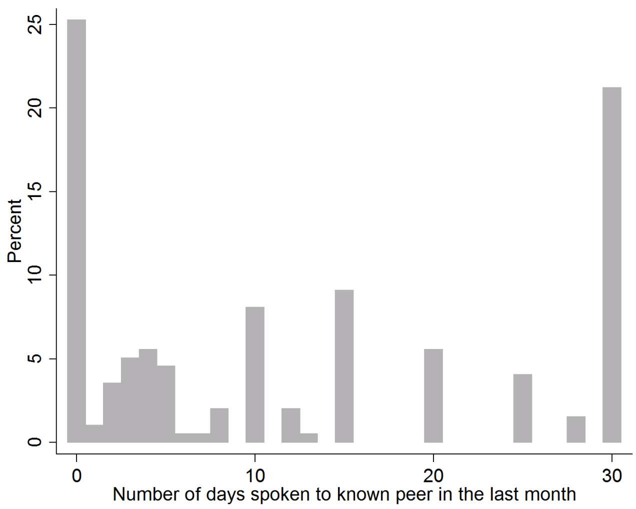
(c) Job-contact network



(d) Job-contact network

The sample includes all individuals who take part in the experiment. Links to neighbours who have not participated in the experiment are included. Links in figures (a) and (b) indicate personal knowledge. Links in figures (c) and (d) indicate exchange of job information or request of assistance for finding a job.

Figure 2.10: Number of days in which respondent i has talked to respondent j in the last 30 days



The sample includes all pairs of individuals who take part in the experiment and know each other.

A.3 Tables

Table 2.7: Summary statistics for within-block dyadic variables

Variable	Mean	Std. Dev.	Min.	Max.	N
i knows j	0.116	0.32	0	1	15588
j is a family member	0.124	0.33	0	1	1804
j is a close friend	0.135	0.342	0	1	1804
j is a worship place acquaintance	0.654	0.476	0	1	1804
i has asked job assistance to j	0.192	0.394	0	1	1804
i has given assistance to j	0.166	0.372	0	1	1804
i has received assistance to j	0.19	0.393	0	1	1804
i has lent money to j	0.146	0.354	0	1	1804
i has borrowed money from j	0.142	0.349	0	1	1804
i has given a gift to j	0.115	0.319	0	1	1804
i has received a gift from j	0.113	0.317	0	1	1803

Table 2.8: Summary statistics for within-session dyadic variables

Variable	Mean	Std. Dev.	Min.	Max.	N
i knows j	0.074	0.262	0	1	2666
Days spoken	12.298	11.677	0	30	198
Family member	0.152	0.359	0	1	198
Close friend	0.121	0.327	0	1	198
Worship place acquaintance	0.626	0.485	0	1	198
i has asked job assistance to j	0.237	0.427	0	1	198
i has given assistance to j	0.192	0.395	0	1	198
i has received assistance to j	0.232	0.423	0	1	198
i has lent money to j	0.152	0.359	0	1	198
i has borrowed money from j	0.146	0.354	0	1	198
i has given a gift to j	0.121	0.327	0	1	198
i has received a gift from j	0.111	0.315	0	1	198

Table 2.9: Summary of experimental observations by treatment

Treatment	Sessions	Players	Dyads	W Sample	Dyads of W Sample
T1a	10	88	520	85	500
T1n	12	108	648	103	620
T1a2	8	72	432	69	414
all T1	30	268	1600	257	1534
T2a	10	90	540	85	510
T2n	10	89	526	87	512
all T2	20	179	1066	172	1022
All sessions	50	447	2666	429	2556

'Players' refers to all individuals who participated in the experimental sessions. 'Dyads' refers to the pairs that are obtained from the sample of players: 7 pairs for each degree-one player, 6 pairs for each degree-two player, 5 pairs for each degree-three player. 'W sample' refers to all individuals in the 'working sample': individuals who participate in the experimental sessions and have been surveyed. 'Dyads of W sample' refers to all pairs where individual i and individual j are in the working sample.

Table 2.10: Summary of experimental decisions by treatment

	Treatment					
	T1a	T1n	T1a2	T2a	T2n	Total
Degree 1	84	98	60	68	55	365
Degree 2	42	54	31	34	51	212
Degree 3	19	39	30	43	51	182
Random	25	15	17	25	17	99
Total	170	206	138	170	174	858

Each player form two links and hence takes two decisions. This table summarises the decisions of the 430 surveyed players.

Table 2.11: OLS regression: selection into the experiment

	Male	Age	Education	Muslim	Migrant	Employed	Inactive	Earnings	Degree	Degree2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Selected	-.065	-1.394	.876	-.226	-.041	-.041	.014	-146.037	1.426	.527
	(.062)	(.885)	(.605)	(.050)***	(.058)	(.060)	(.051)	(154.368)	(.400)***	(.509)
Obs.	496	496	501	495	488	501	501	501	501	503

OLS regression. Column headings indicate the dependent variable. 'Degree' measures the self-reported number of friends in the block. Degree2 captures to the number of residents of the blocks who report i as their friend. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at residential block level are reported in parenthesis.

Table 2.12: OLS regression: covariates balance across network centrality

	Male	Age	Education	Muslim	Migrant	Employed	Inactive	Earnings	Degree
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Centrality j=2	.056	1.313	-.524	-.072	.017	.047	-.009	103.861	-.405
	(.045)	(1.371)	(.600)	(.047)	(.051)	(.051)	(.067)	(93.997)	(.338)
Centrality j=3	-.006	.751	-.379	-.054	.022	.067	-.025	-32.703	.335
	(.050)	(.973)	(.548)	(.058)	(.051)	(.043)	(.056)	(85.064)	(.255)
Obs.	430	430	430	429	421	430	430	430	430

OLS regression. Column headings indicate the dependent variable. 'Degree' measures the self-reported number of friends in the block. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parenthesis.

Table 2.13: OLS regression: covariates balance across T1 and T2 treatments

	Male	Age	Education	Muslim	Migrant	Employed	Inactive	Earnings	Degree
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Centrality j=2	.056	2.132	-.524	-.072	.017	.047	-.009	103.861	-.405
	(.045)	(1.485)	(.600)	(.047)	(.051)	(.051)	(.067)	(93.997)	(.338)
Centrality j=3	-.010	.712	-.407	-.054	.024	.070	-.029	-30.071	.349
	(.051)	(1.150)	(.533)	(.058)	(.052)	(.044)	(.057)	(86.062)	(.250)
Obs.	429	415	429	429	420	429	429	429	429

OLS regression. Column headings indicate the dependent variable. 'Degree' measureso the self-reported number of friends in the block. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parenthesis.

Table 2.14: OLS regression: covariates balance in T1

	Male	Age	Edu	Muslim	Migrant	Employed	Inactive	Earnings	SocNet	JobNet
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
T1n	-.083	-1.319	-.448	.125	-.080	-.073	-.092	-194.403	.983	.352
	(.063)	(1.133)	(.451)	(.066)*	(.055)	(.064)	(.060)	(128.144)	(.530)*	(.311)
Obs.	257	252	257	257	252	257	257	257	257	257

OLS regression. Column headings indicate the dependent variable. 'Degree' measures the self-reported number of friends in the block. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parenthesis.

Table 2.15: OLS regression: covariates balance in T2

	Male	Age	Edu	Muslim	Migrant	Employed	Inactive	Earnings	SocNet	JobNet
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
T2n	-.070	-.490	-.128	.010	-.003	-.061	.030	-92.096	-.558	-.097
	(.082)	(1.317)	(.638)	(.076)	(.057)	(.055)	(.069)	(106.410)	(.558)	(.239)
Obs.	172	163	172	172	168	172	172	172	172	172

OLS regression. Column headings indicate the dependent variable. 'Degree' measures the self-reported number of friends in the block. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parenthesis.

Table 2.16: Linear probability model: understanding

	(1)	(2)	(3)	(4)
T1n	.092	.089		
	(.078)	(.078)		
T2a	.017	.010		
	(.087)	(.086)		
T2n	-.172	-.171		
	(.134)	(.132)		
T1a2	-.050	-.043	-.043	-.033
	(.100)	(.102)	(.080)	(.084)
Const.	.800	.803	.793	.782
	(.070)***	(.083)***	(.037)***	(.074)***
Obs.	426	426	426	426
Cluster N	50	50	50	50
Controls		✔		✔

OLS regression. The dependent variable is a dummy which takes a value of one if the respondent answered correctly both questions about the incentives of the game. Columns 2 and 4 include controls for whether the individual is male, muslim and the degree in the residential-block network. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parenthesis.

Table 2.17: Linear probability model: T1 treatments

	(1)	(2)	(3)
$\text{degree}_j = 2$		-.014	-.156
		(.973)	(.208)
$\text{degree}_j = 3$		.004	-.266
		(.977)	(.038)**
High understanding_i	.132	.129	
	(.000)***	(.008)***	
High understanding_i * $\text{degree}_j = 2$	-.200	-.186	
	(.002)***	(.110)	
High understanding_i * $\text{degree}_j = 3$	-.240	-.244	
	(.002)***	(.012)**	
sent$> 0_i$			-.032
			(.725)
sent$> 0_i$ * $\text{degree}_j = 2$			-.015
			(.865)
sent$> 0_i$ * $\text{degree}_j = 3$			.079
			(.471)
Const.	.290	.293	.428
	(.000)***	(.000)***	(.000)***
Obs.	1523	1523	1534
Cluster N	30	30	30
Controls	✔	✔	✔

Dyadic OLS regression. The dependent variable is a dummy which takes a value of one if player i wants to form a link with player j. All columns include controls for gender, religion, and degree in the residential-block network. 'High understanding' is a dummy for players who answered correctly all questions about understanding of the incentives of the game. 'sent$> 0_i$' is a dummy for players who sent a positive amount in the dictator game. Sample size in columns (1) and (2) excludes 11 dyads for two individuals for whom I do not have a complete understanding questionnaire. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors are corrected for clustering at session level. P-values obtained with wild bootstrap-t procedure are reported in parentheses.

Table 2.18: Linear probability model: T2 treatments

	(1)	(2)	(3)
$degree_j = 2$		-.090 (.410)	-.152 (.198
$degree_j = 3$		-.029 (.869)	-.153 (.224)
High $understanding_i$	.020 (.725)	-.017 (.829)	
High $understanding_i$ * $degree_j = 2$	-.060 (.344)	.030 (.797)	
High $understanding_i$ * $degree_j = 3$	.032 (.737)	.061 (.693)	
$sent > 0_i$			-.103 (.158)
$sent > 0_i$ * $degree_j = 2$			.107 (.368)
$sent > 0_i$ * $degree_j = 3$			.210 (.112)
Const.	.249 (.000)***	.287 (.000)***	.352 (.000)***
Obs.	1015	1015	1022
Cluster N	20	20	20
Controls	✔	✔	✔

Dyadic OLS regression. The dependent variable is a dummy which takes a value of one if player i wants to form a link with player j. All columns include controls for gender, religion, and degree in the residential-block network. 'High understanding' is a dummy for players who answered correctly all questions about understanding of the incentives of the game. '$sent > 0_i$' is a dummy for players who sent a positive amount in the dictator game. Sample size in columns (1) and (2) excludes 7 dyads for one individuals for whom I do not have a complete understanding questionnaire. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors are corrected for clustering at session level. P-values obtained with wild bootstrap-t procedure are reported in parentheses.

Table 2.19: Linear probability model (2.10): all treatments

$degree_i = 2$	.240 (.056)***
$degree_i = 3$	.110 (.059)*
$degree_i = 1$ * T2	-.064 (.075)
$degree_i = 2$ * T2	-.146 (.077)*
$degree_i = 3$ * T2	-.152 (.076)**
Const.	.302 (.067)***
Obs.	858
Cluster N	50
Controls	✔

OLS regression. Dependent variable is a dummy which takes a value of one if player i chose to establish a link with a degree-one partner. Includes controls for gender, religion, and degree in the residential-block network. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parentheses.

Table 2.20: Linear probability model (2.11): all treatments

T2a	-.104
	(.088)
T2n	-.133
	(.063)**
Non anonymous	.003
	(.064)
Const.	.436
	(.064)***
Obs.	858
Cluster N	50
Controls	✔

OLS regression. Dependent variable is a dummy which takes a value of one if i chose to establish a link with a degree 1 j. Includes controls for gender, religion, and degree in the residential-block network. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parentheses.

Table 2.21: Linear probability model (2.12): all treatments

Panel a	
T2 * sent>0_i	-.136
	(.057)**
T2 * sent=0_i	-.056
	(.115)
sent>0_i	-.028
	(.088)
Const.	.460
	(.097)***
Panel b	
$\beta_1 - \beta_2$=0	.051
	(.481)
Obs.	858
Cluster N	50
Controls	✔

OLS regression. Dependent variable is a dummy which takes a value of one if i chose to establish a link with a degree 1 j. Includes controls for gender, religion, and degree in the residential-block network. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parentheses. Panel b reports the F statistics (and p value in parenthesis) for a Wald test of the equality of coefficients β_1 and β_2 in model 2.12.

Table 2.22: Linear probability model (2.13): all treatments

Panel a	
T2a * sent>0_i	-.134
	(.082)
T2a * sent=0_i	-.005
	(.147)
T2n * sent>0_i	-.139
	(.075)*
T2n * sent=0_i	-.115
	(.127)
sent>0_i	-.028
	(.087)
Non anonymous	.002
	(.065)
Const.	.461
	(.107)***
Panel b	
$\beta_1 - \beta_2$=0	1.07
	(.307)
$\beta_3 - \beta_4$=0	.03
	(.871)
Obs.	858
Cluster N	50
Controls	✔

OLS regression. Dependent variable is a dummy which takes a value of one if i chose to establish a link with a degree 1 j. Includes controls for gender, religion, and degree in the residential-block network. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parentheses. Panel b reports the F statistics (and p value in parenthesis) for a Wald test of the equality of coefficients β_1 and β_2, and β_3 and β_4 in model 2.13.

Table 2.23: Linear probability model: anonymous treatments

	(1)	(2)	(3)	(4)
$degree_j = 2$	-.182	-.180	-.128	-.163
	(.006)***	(.154)	(.196)	(.008)***
$degree_j = 3$	-.303	-.110	-.054	-.158
	(.010)**	(.444)	(.845)	(.034)**
i knows j	-.013	-.018	.125	.044
	(.903)	(.693)	(.200)	(.400)
Const.	.421	.384	.324	.375
	(.000)***	(.002)***	(.000)***	(.000)***
Obs.	482	397	485	1364
Cluster N	10	9	10	29
Sample	T1a	T1a2	T2a	T1a, T2a2, T2a

Dyadic OLS regression. The dependent variable is a dummy which takes a value of one if player i wants to form a link with player j. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors are corrected for clustering at session level. P-values obtained with wild bootstrap-t procedure are reported in parentheses.

Table 2.24: Linear probability model: non-anonymous treatments

	(1)	(2)	(3)	(4)	(5)	(6)
Panel a						
$\text{degree}_j = 2$	-.162	-.156	-.159	-.004	-.005	-.005
	(.026)**	(.036)**	(.034)**	(.919)	(.961)	(.887)
$\text{degree}_j = 3$	-.161	-.161	-.162	.075	.078	.079
	(.006)***	(.004)***	(.004)***	(.491)	(.426)	(.465)
i knows j	.282	.024	.153	.249	.237	.262
	(.118)	(.935)	(.038)**	(.280)	(.090)*	(.250)
Social network degree_j	.002			.002		
	(.799)			(.679)		
Social network degree_j* i knows j	-.020			-.0001		
	(.401)			(.553)		
Job network degree_j		-.004			.008	
		(.591)			(.679)	
Job network degree_j* i knows j		.058			.011	
		(.164)			(.553)	
Employed_j			.028			-.005
			(.400)			(.843)
Employed_j* i knows j			-.055			-.034
			(.513)			(.895)
Const.	.377	.395	.381	.249	.254	.264
	(.000)***	(.000)***	(.000)***	(.000)***	(.000)***	(.004)***
Panel b						
$\beta_4 + \beta_5$=0 (social network degree)	.87			0.01		
	(.372)			(.938)		
$\beta_4 + \beta_5$=0 (job network degree)		2.95			0.19	
		(.115)			(.675)	
$\beta_4 + \beta_5$=0 (employed)			0.09			0.02
			(.776)			(.879)
Obs.	596	596	596	498	498	498
Cluster N	11	11	11	10	10	10
Sample	T1n	T1n	T1n	T2n	T2n	T2n

Dyadic OLS regression. The dependent variable is a dummy which takes a value of one if player i wants to form a link with player j. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors are corrected for clustering at session level. P-values obtained with wild bootstrap-t procedure are reported in parentheses. Panel b reports the F statistics (and p value in parenthesis) for a Wald test on the sum of coefficients β_4 and β_5, for different variables x in model (2.14).

Table 2.25: Linear probability model: T2n

	(1)	(2)	(3)	(4)	(5)
Panel a					
$\text{degree}_j = 2$	-.005 (.967)	-.005 (.977)	-.004 (.923)	-.003 (.989)	-.005 (.889)
$\text{degree}_j = 3$	.078 (.406)	.079 (.398)	.075 (.495)	.076 (.438)	.082 (.457)
i knows j	.251 (.076)*		.251 (.271)	.273 (.064)*	.261 (.254)
i knows j * sent $>0_i$		.270 (.060)*			
i knows j * sent $=0_i$		.193 (.707)			
social network degree_j			.002 (.709)		
social network degree_j * i knows j * sent $>0_i$			.0009 (.931)		
social network degree_j * i knows j * sent $=0_i$			-.007 (.909)		
job network degree_j				.007 (.675)	
job network degree_j * i knows j * sent $>0_i$				.031 (.453)	
job network degree_j * i knows j * sent $=0_i$				-.187 (.212)	
employed_j					-.007 (.841)
employed_j * i knows j * sent $>0_i$					.195 (.523)
employed_j * i knows j * sent $=0_i$					-.572 (.106)
sent $>0_i$	-.017 (.378)	-.022 (.444)	-.021 (.593)	-.035 (.150)	-.037 (.220)
Const.	.276 (.000)***	.262 (.000)***	.266 (.000)***	.281 (.000)***	.292 (.000)***
Panel b					
$\beta_5 - \beta_6$=0 (i knows j)	0.14 (.717)				
$\beta_5 - \beta_6$=0 (social network degree)	.09 (.768)				
$\beta_5 - \beta_6$=0 (job network degree)	15.6 (.003)***				
$\beta_5 - \beta_6$=0 (employed)	22.16 (.001)***				
Obs.	498	498	498	498	498
Cluster N	10	10	10	10	10
Sample	T2n	T2n	T2n	T2n	T2n

Dyadic OLS regression. The dependent variable is a dummy which takes a value of one if player i wants to form a link with player j. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors are corrected for clustering at session level. P-values obtained with wild bootstrap-t procedure are reported in parentheses. Panel b reports the F statistics (and p value in parenthesis) for a Wald test of the equality of coefficients β_5 and β_6, for different variables x in model (2.15).

Instructions

Introduction

Good evening! Thank you for joining us tonight. You will take part in an experiment that will last no more than 1 hour.

Please switch off your mobile phones. Also, during the experiment you are not allowed to talk. You will have a chance to ask questions after I finish each part of this explanation. But while the experiment is on, we ask you please not to talk. You are given a paper copy of the instructions I am about to read. You can refer to that at any point in the game.

You are of course allowed to leave the game whenever you want. However, payments will be made only at the end of the whole game. Hence if you choose to leave, you will not receive any payment.

In front of you, you will find a game sheet, and a copy of these instructions.

The game sheet is where all decisions and outcomes in this game will be recorded.

Please answer the first questions on the game sheet. Please add your name, surname, and mobile phone number. Leave the other boxes empty.

The ID-Letter

You will be identified with a letter of the alphabet. I will now let each of you pick a letter from this urn. Please write the letter in the dedicated space of your game sheet.

You are not allowed to inform anyone in the room of which letter you have picked. If you are found doing so, you will not be allowed to continue the game and will not collect any winnings.

Tonight your activities will be divided in two phases. We will explain each of these in turn.

The first phase

In this phase, you and everyone else starts with 20 Birr.

You will have to decide how to split these 20 Birr between yourself and a partner in the room. Your identity and that of your partner will remain anonymous. Also, no one will be informed of your decision.

You should write down on the game sheet how much you want to keep for yourself. You can write any amount from 0 to 20 Birr. What you do not keep for yourself, will be sent to your receiving partner.

Please now ask any questions you may have.

Please write down on your game sheet the amount you wish to keep out of the 20 Birr.

The job

Each of you in the room has a chance to be given a small job. This job consists of filling 5 bags with exactly 20 beans. The job will pay you 30 Birr.

To determine whether you got this small job or not, you will pick a ball from this urn. The urn has 5 blue balls and 5 green balls. If you pick the blue ball, you got the job. If you pick the green ball, you did not get the job.

Whoever got the job will be able to refer another person to do the same job. The referred person will have to come back tomorrow to this room, at the same time. If the referred person comes back tomorrow, he/she will do the envelope job and will be paid 30 Birr for it.

Tonight, each of you is assigned a number of friends. Those who get the job tonight will give their referral to one of these friends. Only friends who did not get the job today are eligible for a referral. If a player who gets the job has more than one unemployed friend, he/she will pick one of them at random. This is done by picking one card from an urn which contains a card for each of his unemployed friends. The game assistant will carry out this operation.

Let us give you an example. If Selam gets the job, and has only one unemployed friend, things are easy: her friend is referred to carry out the job tomorrow. If Selam gets the job, and has two unemployed friends, for example, then the game assistant will pick a letter from an urn containing a small card for each one of her unemployed friends. Whoever is drawn gets the referral.

Look for your ID-letter in the following picture to find out about how many friends you have and who they are. For example, if your ID-letter is G you have one friend: B. Another example: if your ID-letter is E, your friends are F and D. The number next to the letter reminds you of the number of friends a player has. D, for example, has 3 friends. Hence the number 3 next to the letter.

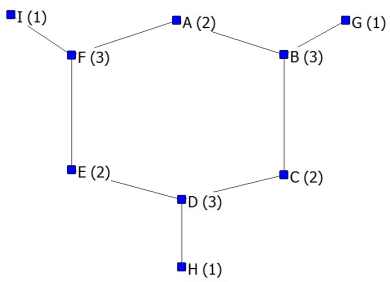


Figure 1 The lab friends

Before jobs are assigned, you have a chance to make two new friends. Write down on your game sheet the ID-letters of two more people with whom you would like to be friend tonight. After the job lottery, and before the referral phase, we will randomly draw one participant among those who DID NOT GET THE JOB and we will add his lab friends to the list. This person will have more friends tonight and has hence a chance to get a referral either by his original friends or by his new friends.

Notice that the new friendships that you specify on the game sheet may be established only if you do not get the job. In other words, you may RECEIVE a referral from the friends you are adding, but you will not be able to GIVE a referral to the friends you are adding.

Also notice that the chance of getting a referral from another player depends on how many friends this player already has. The chance to be referred by a player with many friends is smaller than that to be referred by a player with few friends.

Let us illustrate this with an example. Say Selam becomes friend with Tesfaye, who has only one friend: Beletu. If Tesfaye gets the job and both Selam and Beletu did not get the job, he can refer either Selam or Beletu. Say, instead, Selam becomes friend with Tamrat, who is friend with Ibrahim, Tarik and Haile. If Tamrat gets the job and no friend of his gets the job, he can refer either Selam, or Ibrahim, or Tarik, or Haile. So it is more likely that Selam gets a referral from Tesfaye than from Tamrat.

Let's be precise. If you do not get the job and become friend with a player who initially had only one friend, the probability of getting a referral from this player is 75%. This corresponds to the chance of drawing a blue ball from an urn with 75 blue balls and 25 green balls. If the player you befriend had two initial friends, the probability of getting a referral from him is 58%. This corresponds to the probability of drawing a blue ball from an urn with 58 blue balls and 42 green ones. Finally, if the player had three friends, the probability of getting a referral is 47%. These numbers are summarized in the table below.

Initial number of friends of player you choose	Probability of being referred by this player when you are unemployed
1 friend	75%
2 friend	58%
3 friend	47%

You can choose to become friend with any two players whom you are not currently friend with. If you do not want to specify either or both of these friends, you can write R and we will randomly pick for you.

Your decision about which additional friends to make will remain private. No one will be informed of who you have chosen. So to summarize, you will now:

- Choose which additional friends you would like to have. Please write their ID-letters down in your game sheet
- Give your game sheet to the game assistant
- Pick from the urn to determine whether you got the job
- Answer a short list of questions on a separate sheet
- If you do not get the job, you will be informed whether you got a referral and then collect your winnings from the game assistant and leave
- If you got the job, you will first carry out the job. After completing the job you will collect your winnings from the game assistant and leave

Please now ask any questions you may have. Please answer the questions on the additional question sheet, to test understanding of the rules of the game.

Please write down on your game sheet the ID-letters of the additional friends you would like to make.

Thank you very much for participating in the experiment!

3 Risk attitudes and information networks among cocoa farmers in Ghana

3.1 Introduction

What characteristics do individuals value in their advisors? Exchange of advice and information plays an important role in the diffusion of technology, opinions and products (Conley and Udry, 2010; Jackson, 2014). A rich theoretical literature has developed from the observation that individuals consider the costs and benefits of the social connections they form (Jackson and Wolinsky, 1996; Bala and Goyal, 2000; Galeotti and Goyal, 2010). This literature makes predictions on the structure of networks, including information and advice networks, that can be tested against a growing body of field evidence (Fafchamps, 2011; Santos and Barrett, 2010). Researchers typically observe existing connections, but have imperfect knowledge of the underlying process of link formation. When connections can be refused or it is unclear whether both individuals value the relationship, the links that are observed may differ from the links that individuals would have liked to form (Comola and Fafchamps, 2013).

Direct evidence on the links that individuals would like to form and on the characteristics they value in their partners is scarce. This limits our understanding of the process through which networks are created and hence our capacity to predict how networks will respond to interventions and shocks. For example, suppose I observe farmers with a similar level of experience seeking each other's advice. This could be

This chapter is a revised version of the thesis I submitted for the M.Phil. in Economics at the University of Oxford. I am grateful to the respondents, enumerators, the Ghana Cocoa Board and Mark Fiorello for their patience and professionalism in putting together the data without which this thesis would have not been written and to Francis Teal and Andrew Zeitlin for granting use the data. I thank Meg Meyer, Marcel Fafchamps, Andrew Zeitlin for their valuable advice. I also thank Abigail Barr, Lea Cassar, Mauro Caselli, Paolo Falco, Edoardo Gallo, Julien Labonne, Juana Lamote De Grignon Perez, William Lee, Esteban Ortiz Ospina and seminar participants at the University of Sassari for useful comments.

because farmers value such advice the most, or because more experienced farmers refuse to exchange information with less experienced peers. In the latter case, interventions that increase the assets of inexperienced farmers may enable them to trade some of these assets off in exchange for agricultural advice. To predict this effect, I need an understanding of the characteristics that make advisors valuable in the eyes of farmers. My objective in this chapter is to contribute evidence to build such understanding.

I analyse the responses of a sample of Ghanaian cocoa farmers to an hypothetical question on their willingness to seek the agricultural advice of other farmers in their village. Using a matched lottery experiment for the elicitation of risk attitudes, I am able to document a relationship between interest in another farmer's advice and the difference in risk attitudes between the two farmers. The more risk averse the potential advisor compared to the respondent, the less likely is the respondent to desire his advice. The magnitude of the effect is meaningful. One standard deviation increase in the difference in risk aversion is associated with a 7.4 percent decrease in the likelihood that the respondent would like the advice of the other farmer.

This effect is robust to the inclusion of a number of controls motivated by the literature on risk aversion and on information networks (Dohmen et al., 2011; Santos and Barrett, 2010), and to the use of village fixed effects. Difference in risk aversion does not influence other dimensions of interaction, such as speaking, exchanging loans, or walking past each other farms. This rules out unobserved heterogeneity that is correlated with generic social interaction. I also present evidence in support of my choice of functional form.

I propose a model of strategic experimentation between two farmers that offers an explanation for my finding: in equilibrium, the more risk-averse[1] farmer free rides on the information produced by the experiments of the other farmer. I present evi-

[1] I derive a specific condition on the difference in risk aversion, which is explained in result 3.

dence suggesting that information exchange is often bilateral, and predict that farmers are unwilling to initiate communication with other farmers if they expect the latter will free ride on the information they produce. This interpretation is supported the study of Bandiera and Rasul (2006), which documents that farmers who observe the experiments of many peers delay the adoption of a new technology.

I contribute to two distinct strands of the literature. First, I contribute to the empirical literature on the formation of social networks (Fafchamps and Gubert, 2007; Krishnan and Sciubba, 2009; Karlan et al., 2009; Santos and Barrett, 2010). My results suggest that in information networks individuals may be interested in the relative willingness to take risks of their advisors. This highlights how risk aversion has a role to play in dimensions of social interactions other than risk sharing and extends our understanding of the cases where preferences for connections do not display homophily (McPherson et al., 2001).

Second, the results I present contribute to the literature that studies how individuals consider strategically the preferences of others. In the experiment reported in Attanasio et al. (2012), for example, socially proximate players create risk-sharing groups on the basis of the similarity of individual risk preferences. Finan and Schechter (2012) document that political middlemen are more like to attempt to buy the votes of individuals who show higher level of reciprocity in an experimental game. In both cases, targeting individuals with certain preferences is in the interest of the decision maker. These findings suggest that the preferences that determine individuals' economic decisions can influence their broader pattern of social interaction. The results of this chapter provide further support for this hypothesis. I explore some of the implications for policy in the conclusion.

The rest of this chapter is organised as follows. Section 3.2 proposes a model of strategic experimentation and discusses or predictions and testing strategy. The data set and the lottery experiment are described in section 3.3. I present the results and the robustness checks in section 3.4. Section 3.5 concludes.

3.2 Model

I develop a simple model of strategic experimentation. This model studies interaction between two farmers *after* an undirected link has been formed.[2] As in Conley and Udry (2010), farmers are Bayesian learners who face uncertainty about a universally available new technology. In this model, *the returns to adopting the new technology are unknown*. Farmers acquire more precise beliefs about the true returns of the technology by carrying out experiments, or by observing the experiments of the other farmers. The information produced from these trials is non-rival and non-excludable. Farmers can hence *free ride* on the information produced by their neighbours.

I explore the possibility that the two farmers have different attitudes to risk: one farmer finds it less costly to bear the risk of experimentation. Farmers are identical in all other characteristics: the returns of the technology are common, they have equal prior beliefs about these returns, and they are equally good at experimenting, that is, they receive signals of the same precision.

The two farmers are called i and j. They have to choose an action $a^t \in \{L, H\}$ at time $t \in \{0, 1\}$. I can think of H as a farmer's decision to use inorganic fertilizer, and L as the decision not to use it. Alternatively, L and H can be two different quantities of fertilizer, the first of which has known returns.

When a farmer chooses L, he makes a certain profit of y_L. Technology H, on the other hand, gives profits y_H^*. Farmers do not know the value of y_H^*, but, at the beginning of the game, they share the same belief about y_H^*, which is given by:

$$y_H = \mu_0 + \sigma_0 u_0$$

[2]In an undirected information link both farmers observe each other.

where $u_0 \sim N(0,1)$. I assume $\mu_0, ph, y_l, pl \in R_+$ and $\sigma_0 \in R_{++}$ and all parameters are finite.

In this model, experimentation resolves all uncertainty. In other words, when a farmer experiments with H in the first period, he observes the true y_H^* and his belief about technology h in period two is accordingly updated to $y_H = y_H^*$.[3]

I assume that the information created by an experiment is non-excludable: at the end of period 1 both the actions and the profits of farmer i are observed by farmer j and viceversa. This is a fairly natural assumption to make if the two farmers are are connected in an information network. As I assume that a single experiment is sufficient to eliminate all uncertainty, there is no additional value in the information provided by a farmer's period 1 experiment with technology H when the other farmer experiments in period 1. Farmers' experiments are hence perfect substitutes.

In order to introduce risk aversion explicitly, I specify the following exponential utility function:[4]

$$u_i(y_{a^1}, y_{a^2}) = 1 - \exp\{-r_i(y_{a^1} + y_{a^2})\} \text{ for } r_i > 0 \tag{3.1}$$

For simplicity, I assume no discounting takes place. The parameter r captures the

[3]This is a special case of Bayesian updating in which the variance of the signal received by the agent is equal to zero. To see this, suppose instead that choosing technology H in period 1 generates profits $\pi_H = y_H^* + \sigma_e v_1$ where $v_1 \sim N(0,1)$. That is, the farmer receives a noisy signal of y_H^*. The higher σ_e^2, the noisier the signal. A Bayesian farmer uses π_H to update his beliefs about y_H^*. Using standard Bayesian updating formulae from DeGroot (1970), after updating y_H will be normally distributed with variance $\frac{\sigma_0^2\sigma_e^2}{\sigma_0^2+\sigma_e^2}$ and mean $\frac{\sigma_0^2}{\sigma_0^2+\sigma_e^2} y_h^1 + \frac{\sigma_e^2}{\sigma_0^2+\sigma_e^2}\mu_0$. It is easy to see that when σ_e^2 is equal to zero, y_H has variance zero and mean y_H^*. This assumption allows me to simplify the analysis of period 2 behaviour, while not departing from the Bayesian framework.

[4]I use this functional form because of its tractability. The drawback is that this function implies constant absolute risk aversion, whereas the lottery games I use in the empirical part estimate coefficients of constant relative risk aversion (CRRA). A CRRA utility function implies decreasing absolute risk aversion and is thus not consistent with the exponential utility function adopted here. Reformulating the model with a CRRA utility function is an area I am planning to work on in the future.

intensity of risk aversion and is allowed to vary across farmers. I assume that an agent's utility function and coefficient of risk aversion are common knowledge.

The timing of the game is as follows:

- $t = 1$: farmer i chooses a^1 and, at the end of the period, observes the profits of both farmers. On the basis of that, he updates his beliefs;
- $t = 2$: farmer i chooses a^2 and, at the end of the period, collects the output from both periods.

I first look at period 2 behaviour and then analyse strategic interaction in period 1.

Period 2

In the second period each farmer chooses a^2 to maximise expected utility given his current belief about the profitability of the technology. There are two cases to analyse: (i) nobody has experimented with H in period 1, and (ii) at least one farmer has experimented with H in period 1.

Suppose *neither farmer has experimented with technology H in period 1*. Initial beliefs have not been updated. Farmer i will compare the certain return y_L with the expected utility he gets from the random variable y_H. For exponential utility, I can easily calculate the certainty equivalent of the gamble with technology H. This is given by:

$$\begin{aligned} \mathrm{CE}(y_H) &= \mathrm{E}(y_H) - \frac{1}{2} r_i \mathrm{Var}(y_H) \\ &= \mu_0 - \frac{1}{2} r_i \sigma_0^2 \end{aligned} \tag{3.2}$$

Technology H will be chosen in period 2 if $\mu_0 - \frac{1}{2} r_i \sigma_0^2 \geq y_L$, and technology L will be chosen otherwise. In other words, in period 2, farmer i will choose technology H

Table 3.1: The game

		Player j H	L
Player i	H	$V_i(HH)$ $V_j(HH)$	$V_i(HL)$ $V_j(HL)$
	L	$V_i(LH)$ $V_j(LH)$	$V_i(LL)$ $V_j(LL)$

if, given his level of risk aversion, he is sufficiently convinced about the profitability of this technology.

Suppose, on the other hand, that *at least one farmer has tried technology H in period 1*. Both farmers now know y_H^*. Farmer i will use technology H whenever $y_H^* \geq y_L$. Period 2 profits when at least one farmer has experimented in period 1 are given by:

$$z^* = \max\left(y_H^*, y_L\right) \tag{3.3}$$

Hence, when the value of technology H is known, period 2 profits weakly increase from either y_L or y_H^*, to $\max\left(y_H^*, y_L\right)$. In period 1, farmers take this into account when deciding whether to experiment with the risky technology.

Period 1

In period 1, farmers choose what technology to use weighting two factors. First, the profits produced by the technology in period 1. Second, the value of knowing about the true profitability of technology H in period 2.

As the knowledge created by an experiment with technology H is available to both farmers in period 2, there is a strategic element to period 1's decision. Table 3.1 represents the one-shot, non-cooperative game that the two farmers play in period 1.

I use the notation $V_i(XY)$ to indicate the expectation in period 1 of player i's utility when the initial technology choices of farmer i and farmer j are, respectively, X and Y. This is the expected utility that is generated by the sum of period 1 and period 2 profits. Technology decisions in the first period have an effect on the profits expected in period 2. In particular, given the analysis so far, I know the following about period 1 and period 2 profits.

Period 1 profits. If in period 1 farmer i uses technology L, period 1 profits are given by the known constant y_L. If he uses technology H, farmer i is uncertain about period 1 profits. His beliefs are captured by the normally distributed random variable y_H.

Period 2 profits. (i) If nobody has experimented with H in period 1 and farmer i is pessimistic about the technology, farmer i gets profits of y_L. If nobody has experimented and farmer i is optimistic, farmer i chooses technology H and is uncertain about period 2 profits. His beliefs are captured by y_H. (ii) If at least one farmer has experimented with H, then, from period 1's perspective, period 2 profits will be given by the censored normal variable $z = \max\left(y_H, y_L\right)$.[5]

These observations allow me to calculate the values of $V_i(HH)$, $V_i(LH)$, $V_i(HL)$, and $V_i(LL)$. I report the formulae in the appendix.[6] To calculate the equilibrium of the one-shot game in period 1, I need to study farmer i's optimal decision when farmer j in period 1 uses technology H, and when he uses technology L. I rely on numerical simulations methods and study how (i) $V_i(HH)-V_i(LH)$ and (ii) $V_i(HL)-V_i(LL)$ vary with the risk aversion of farmer i. The simulations are reported in sections A.1.2 and A.1.3 of the appendix. I obtain two main results.

[5]In period 2, once y_H^* is revealed, profits are given by $z^* = \max\left(y_H^*, y_L\right)$. In period 1, however, farmer i does not know y_H^*. He maximises utility on the basis of the information he has about y_H^*, which is contained in y_H. His belief about period 2 profits is hence given by $z = \max\left(y_H, y_L\right)$. As y_H is normally distributed, z follows a normal distribution censored from below at y_L.

[6]Note that $V_i(HH) = V_i(HL)$: if farmer i uses technology H in the first period, farmer j's period 1 experiment with H brings no additional information and hence has no effect on his expected utility.

Result 3.1. *In the interval* $r \in (0,1]$ *there is at most one value of r- call it* r^**- such that:*

$$V_i(LH) > V_i(HH) \qquad \text{if } r_i > r^*$$
$$V_i(LH) = V_i(HH) \qquad \text{if } r_i = r^*$$
$$V_i(LH) < V_i(HH) \qquad \text{if } r_i < r^*$$

Result 1 shows that there is a cut-off point in risk aversion that determines whether farmer i uses the risky technology when farmer j is experimenting with it. If farmer i is more risk averse than r^*, he prefers to use the safe technology. When $r_i > r^*$, I say that farmer i *satisfies* the cut-off point r^*.

Result 3.2. *In the interval* $r \in (0,1]$ *there is at most one value of r- call it* r^{**}*- such that:*

$$V_i(LL) > V_i(HL) \qquad \text{if } r_i > r^{**}$$
$$V_i(LL) = V_i(HL) \qquad \text{if } r_i = r^{**}$$
$$V_i(LL) < V_i(HL) \qquad \text{if } r_i < r^{**}$$

$$\text{and } r^{**} \geq r^*$$

Result 2 shows that there is a second cut-off point that determines farmer i's technology choice when farmer j does not experiment with the risky technology. Again, farmers who are more risk averse than the cut-off point r^{**} prefer the safe technology.

Importantly, this second cut-off point is to the right (ie. it is larger) than the first cut-off. Choosing technology H when farmer j has chosen L allows farmer i to make a better decision in period 2. On the other hand, choosing technology H when farmer j is already experimenting with H brings no additional benefits in period 2, and it generates the same uncertainty in period 1. A farmer who is willing to bear the uncertainty in both cases has to be more tolerant of risk than a farmer who is only prepared to choose the risky technology in the first case.

Table 3.2: Nash equilibria in pure strategies

	$r_j \leq r^* \leq r^{**}$	$r^* \leq r_j \leq r^{**}$	$r^* \leq r^{**} \leq r_j$
$r_i \leq r^* \leq r^{**}$	H,H	H,L	H,L
$r^* \leq r_i \leq r^{**}$	L,H	H,L and L,H	H,L
$r^* \leq r^{**} \leq r_i$	L,H	L,H	L,L

The Nash equilibrium of the game is determined by the particular combination of risk attitudes of the two farmers. The analysis is presented in detail in section A.1.4 in the appendix. Table 3.2 illustrates. I summarise this analysis in the following result:

Result 3.3. *If $r_j > r_i$ and r_j satisfies at least one more cut-off point than r_i, then the unique equilibrium of the game is one where, in period 1, farmer i experiments with the risky technology and farmer j chooses the safe technology.*

Predictions

The model above studies equilibrium technology choices after a link has been formed. How will this affect the desire of farmer i to form a link with farmer j?

I hypothesise that farmer i is less willing to exchange advice with farmer j if he does not expect farmer j to experiment with the new technology. Links may impose costs on both parties, for example the opportunity cost of the time spent communicating information. Farmer i may not be prepared to bear these costs if he receives no benefits in return.[7] A sense of negative reciprocity may compound the cost of maintaining the link, if farmer i perceives free riding as an unkind act.

[7]Farmer j may actually possess valuable information because he is able to observe the results of the experiments of other farmers. I return to this possibility below.

In making this prediction, I am assuming that information links are undirected: once farmer i approaches farmer j for advice, information will flow both ways. In the data section below I present some descriptive evidence consistent with this assumption.

Farmer j does not experiment in equilibrium if (i) $r_j > r_i$, and (ii) r_j satisfies at least one more cut-off than r_i. In my data, I observe r_i and r_j, but do not observe the cutoffs r^* and r^{**} directly. For the empirical analysis, I hence proxy conditions (i) and (ii) with the difference between r_j and r_i. Intuitively, the higher this difference, the more likely it is that farmer j will satisfy at least one more cut-off point than farmer j. Furthermore, if the games measure risk aversion imperfectly, conditions (i) and (ii) above are more likely to hold for larger differences between the noisy measures of r_i and r_j.[8] I will hence use the following empirical specification:

$$\text{wants advice}_{ijs} = \beta_0 + \beta_1(\text{risk aversion}_{js} - \text{risk aversion}_{is}) + W_{ijs}\gamma_1 + \nu_s + e_{ijs} \quad (3.4)$$

'wants advice' is a dummy which takes the value of 1 if farmer i in village s answers yes to the question 'Would you like to turn to farmer j for advice on farming matters?'. W_{ijs} is a matrix of 'dyadic' covariates: the simple difference in the case of a continuous variable, and a dummy for having the same value of a binary variable. ν_s is a village fixed effect.

I test the prediction that $\beta_1 < 0$. A negative estimate of β_1 would document that farmer i becomes less likely to desire farmer j's advice as the difference between the risk aversion of farmer j and farmer i increases. I standardise difference in risk aversion, so that the estimate of β_1 can be interpreted as the effect of a one-standard-deviation increase in difference in risk aversion on the probability that farmer i would like farmer

[8] A similar argument can be made if farmer i only observes a noisy, unbiased signal of r_j. Farmer i will be more confident that conditions (i) and (ii) are satisfied the larger the difference between the signals he receives about r_j and r_i.

j's advice.

I will estimate model (3.4) using OLS. To guarantee the consistency of the OLS estimator I require:

$$E(e_{ijs}|X) = 0 \tag{3.5}$$

Assumption (3.5) states that the error in model (3.4) is mean independent of the covariates X which I introduce on the right hand side of the model. In the next section I discuss the set of controls I rely on. Identification of the parameter β fails if, after the inclusion of all controls, there is still residual correlation between the error e_{ijs} and the difference in risk aversion between the two farmers.

I include village dummies in the regression equation to capture village-specific features that may be correlated both with a higher willingness to seek advice from one's neighbours and with a higher heterogeneity in risk aversion. Such features are hence controlled for and cannot cause failure of assumption (3.5).

I cluster standard errors at the level of the village. This allows for arbitrary correlation between the errors of pairs of individuals sampled from the same village. Such correlation can arise, for example, if farmer i imitates the decisions of farmer k. This would create correlation between errors involving farmer i and errors involving farmer k. Alternatively, farmers may derive diminishing utility from additional links, so that a positive e_{ij} error is negatively correlated with the other errors involving farmer i. The clustering correction allows for these types of influence across pairs of farmers in the same village. To perform inference, however, I still require the following assumption about the influence between pairs of farmers in different villages:

$$E(e_{ijs'} e_{kms''}|X) = 0 \tag{3.6}$$

$\forall$ farmers i and j in village s' and farmers k and m in village s'', $s' \neq s''$

Assumption (3.6) rules out correlation between the error terms of pairs of farmers drawn from different villages. My data covers 22 different villages, randomly sampled from a list of all villages in the cocoa growing regions of Ghana. The probability that a farmer sampled in one of these villages interacts with a farmer sampled in a different village is likely to be extremely small. This makes assumption (3.6) plausible. Cameron et al. (2008) show that the clustering correction to standard error over-rejects the null when it is applied to datasets with less than 30 independent clusters. I hence rely on the wild-bootstrap correction to p-values proposed by these authors. In regression tables, I report the corrected p-values below the coefficient estimates.

I consider three factors that could affect the estimates. First, *farmers may be uncertain about the risk preferences of their peers*. Without information on r_j, they may not be able to discriminate between farmers who will free ride on their experiments and farmers who will not. Farmers may however have a better sense of the preferences of those peers with whom they interact frequently. I calculate a dummy variable capturing whether both j and i report speaking to each other in the previous 7 days. Using specification (3.7) below, I study whether, among pairs of farmers who have recently interacted, farmer i's willingness to seek advice from farmer j decreases in the difference in risk aversion between farmer j and i.

$$\begin{aligned} \text{wants advice}_{ijs} =& \beta_0 + \beta_1 \text{speak}_{jis} + \beta_2 \text{speak}_{jis} * (\text{ra}_{js} - \text{ra}_{is}) \\ & + W_{ijs}\gamma_1 + \nu_s + e_{ijs} \end{aligned} \tag{3.7}$$

Second, *farmers may anticipate rejection* and they may only mention farmers with whom they would actually be able to communicate.[9] In the data, I observe farmer j's

[9]The problem of interpretation of hypothetical questions is common to other areas of research in empirical economics, such as the study of the marriage market. Hitsch et al. (2006), for example, study the determinants of the decision to contact browsed profiles in an internet dating website. They worry that true

willingness to link to farmer i. I can use this variable to proxy for the possibility of rejection of the ij link by farmer j. Coefficient β_2 in specification (3.8) is estimated using variation in difference in risk aversion among pairs of farmers where farmer j is willing to link with farmer i. Misreporting due to anticipated rejection should not affect this coefficient. The dummy 'wants advice'$_{jis}$ controls for the effect of j's willingness to link on i's willingness to link when farmers i and j have the same level of risk aversion.

$$\begin{aligned} \text{wants advice}_{ijs} =& \beta_0 + \beta_1 \text{wants advice}_{jis} + \beta_2 \text{wants advice}_{jis} * (\text{ra}_{js} - \text{ra}_{is}) \\ &+ W_{ijs}\gamma_1 + \nu_s + e_{ijs} \end{aligned} \tag{3.8}$$

Farmer j's desire for the advice of farmer i may also proxy for the fact that this particular pair of farmers use similar technologies or face similar soil conditions and hence have much advice to share with each other. I control for similarity in technology used (in particular, for whether both farmers use fertiliser), but do not have information on soil quality. I will hence not be able to disentangle between these two possible interpretations.

Third, *some risk averse farmers* may actually have a substantial amount of information because they *maintain many links*. Communication with these farmers would be highly valuable. Failing to control for peer-to-peer learning could depress the estimates of the effect of the difference in risk aversion if farmer i values both farmer j's experiments and the knowledge farmer j has accumulated from other farmers. As I explain below, for a part of the sample I have data on actual requests for advice. Using this data, I can explore whether the effect of the difference in risk aversion is stronger once I condition on the number of sources of information that farmer j has tried to access. I only observe a limited subset of farmers in every village, so I do not attempt to calculate measures of centrality that are more sophisticated than the count of links.

preferences may not be revealed if unattractive partners refrain to contact the most attractive partners for fear of rejection

$$\text{wants advice}_{ijs} = \beta_0 + \beta_1(\text{ra}_{js} - \text{ra}_{is}) + \beta_2\text{No. farmers approached for advice}_{js} + W_{ijs}\gamma_1 + \nu_s + e_{ijs} \quad (3.9)$$

3.3 Data

The GCFS data set

The data I use comes from the 2008 wave of the Ghana Cocoa Farmers' Survey (GCFS), a panel of farmers from Ghana's main cocoa growing regions. The panel has been collected by the Centre for the Study of African Economies at the University of Oxford since 2002. Sampling for the first wave of the panel was implemented using the 1998/1999 Ghana Living Standards Survey (GLSS). Details can be found in Zeitlin and Vigneri (2005). To retain representativeness, a number of new farmers, sampled from a more recent GLSS, were included in the 2008 wave.

I focus on the 2008 data because it includes both a network module and a matched lab-in-the-field experiment. Collecting data on social interactions is time-consuming. To limit this problem, a sub-sample of farmers in every village was randomly selected from the list of all farmers to be interviewed. I refer to the set of farmers that are interviewed in a village as the 'village sample' and to the randomly chosen subset as the 'network sample'. Each farmer in the 'village sample' is asked a small set of questions about each farmer in the 'network sample'. I restrict the analysis to farmers in the 'network sample' and the links between them. This corresponds to the 'induced-subgraph' data collection scheme of Chandrasekhar and Lewis (2011).

For each farmer in the 'network sample', respondents were first asked whether they 'know of' that particular farmer[10] and then a number of questions about membership

[10] For dyads where a farmer declares not to 'know of' the other person, no data on other dimensions of interaction is collected. I set all of these variables to zero. This means that, for example, I assume that farmer i has not spoken in the last 7 days with farmer j, if farmer i does not know of the existence of farmer j.

in the same kinship, monetary loans, labour exchange, and whether, in the previous seven days, the respondent spoke to the other farmer or walked past the farm of the other farmer. Furthermore, respondents were asked the following two questions about each farmer in the 'network sample' : 'Would you turn to farmer j for advice on farming matters?' and 'Have you ever turned to farmer j for advice on farming matters?'. I interpret answers to the first question as capturing farmers' *willingness to establish an information link* with farmer j. Answers to the second questions capture *an actual attempt to establish an information link*.[11] Unfortunately, the second question was only introduced in the second half of the survey. This means that it is available for a smaller number of villages.

Chandrasekhar and Lewis (2011) emphasise the importance of the sampling technique employed to collect network data. A problem with the method I use is that it under-represents the connections of individuals with many links. Chandrasekhar and Lewis (2011) propose a correction for this problem which would require information on the total number of cocoa farmer in every village. I do not have this data and hence cannot apply their correction. This however should not pose problems for the analysis unless individuals who want the advice of may farmers respond differently to relative risk aversion.

I have data for 22 villages and a total of 443 individuals. The network sample includes 229 farmers. I have a full set of covariates, including data on decisions in the lottery game, for 225 of them. These 225 farmers form the sample I use for the analysis. I thus have data for an average of 10.2 farmers per village. The exact number varies by village, with a standard deviation of about 2 farmers. This sample of farmers generates 2,166 dyads.

I have data on actual communication attempts for 115 of the 225 farmers in the

[11] As I cannot ascertain whether farmer j has indeed satisfied farmer i's request for advice, I will conservatively interpret this data as recording actual communication attempts.

network sample, in 10 of the villages. These generate 1,208 dyads.

I present descriptives on the variables from the network module in table 3.3. In 75 percent of pairs farmer i 'knows of' farmer j. In about half of these dyads, farmer i would like to seek advice from farmer j. For the restricted subsample where I have data on communication attempts, I observe that these occur in about half of the where communication is desired. Lending money or exchanging labour is infrequent. About 18 percent of farmer pairs have spoken to each other in the week prior to the interview.

In 21 percent of cases where at least one party reports approaching the other for advice, the request has been made both ways. This number gives a lower bound to the rate of information exchange that occurs in both directions: misreporting is likely to inflate the number of cases where I observe a communication attempt in only one direction. Misreporting is common in network data (Comola and Fafchamps, 2013). To investigate the potential scale of the problem, I can look at a variable that is undirected by definition: membership in the same kinship. Farmer i reports belonging to the same kinship of farmer j in 15.8 percent of the pairs. This is confirmed by farmer j only in 7.3 percent of pairs. Suppose that communication *always takes place in both directions* and that misreporting of communication attempts occurs at a rate similar to kinship. I would then observe bilateral communication in about 46 percent of the pairs of farmers where at least one party reports asking information to the other party. In the data I observe bilateral communication in 21 percent of such pairs. In the light of the benchmark that takes misreporting into account, this suggests that information is often exchanged in both directions.

Risk aversion

I play an incentivised experiment with each farmer in the 'village sample'. In the experiment, a farmer is presented with six pairs of lotteries. For each pair, the farmer has to choose between a risky lottery and a safe lottery. In exchange for a higher spread

Table 3.3: Summary statistics of variables from the network module

Variable	Percent	N
i would like advice from j	37.8%	2166
Both would like advice from each other	16.1%	2166
i has asked advice from j	18.9%	1208
Both have asked advised from the other farmer	4.1%	1208
i 'knows of' j	75.3%	2166
i and j have spoken in the last 7 days	18.5%	2166
i has lent money to j	1.6%	2166
i has exchanged labour with j	9.3%	1316
i has walked past j's farm in the last 7 days	11.2%	2166

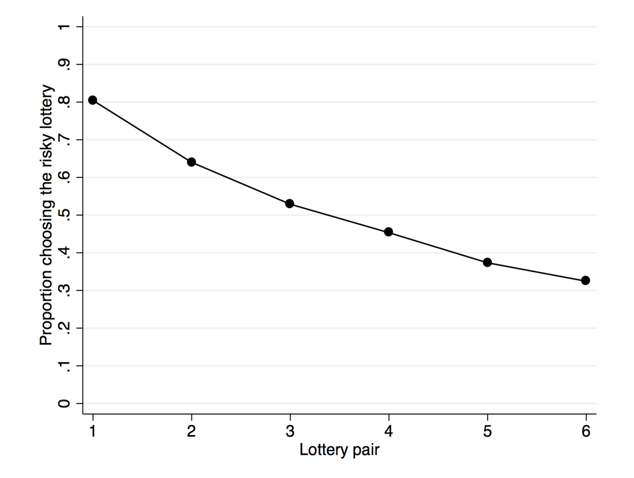
of outcomes, the risky lottery offers higher expected returns. The sequence of lotteries is calibrated so that a player with CRRA utility will choose the risky lotteries in the initial pairs and will eventually switch to the safe lottery. The point at which the player switches to the safe lottery can be used to derive bounds on the level of risk aversion of the player. The structure of this game follows the design of Holt and Laury (2002) and was adapted to the context of Ghana for CSAE by Dr Abigail Barr. In tables 3.17 and 3.18 in the appendix I show the six pairs of lotteries and the CRRA coefficients that correspond to specific patterns of choices.

After a farmer declares his six decisions, one of the lottery pairs is randomly selected by the enumerator. The farmer then plays the lottery within the pair which he has previously chosen. Monetary payments are disbursed immediately after play. The expected value of the lottery varies in range from 8.25 USD PPP, for the risky lottery with the highest value, to 2.25 USD PPP, for the safe lottery with the lowest value.[12] The difference in expected payoff varies from 5.15 to 0.85 USD PPP.

I start by plotting the relative frequency of the decision to play the risky lottery, for each of the six pairs. As expected, this frequency decreases as the risky lottery becomes less appealing compared to the safe lottery.

[12] I convert Ghana Cedis to US Dollars using the 2008 PPP conversion factor reported by the Millennium Development Goals Indicators Website at this address: http://mdgs.un.org/unsd/mdg/SeriesDetail.aspx?srid=699.

Figure 3.1: Proportion of farmers choosing the risky lottery

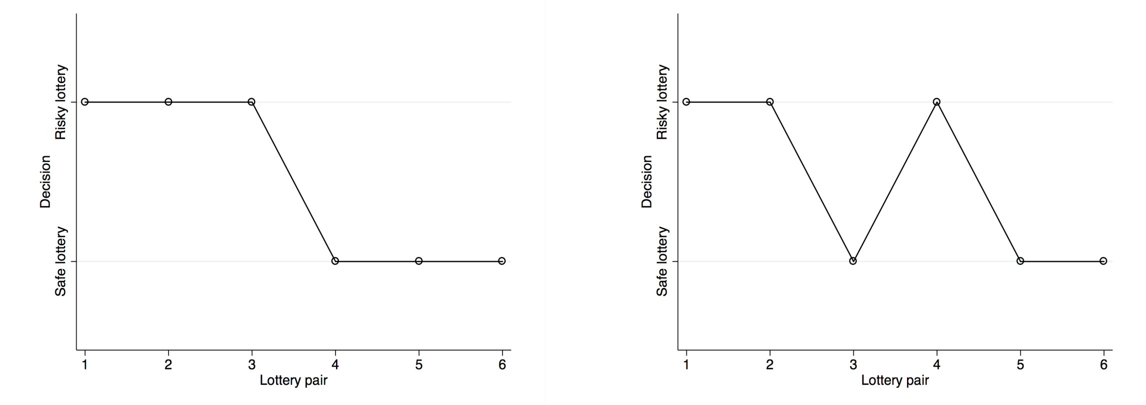


I study individual decisions in two ways. First, I compute the number of times a player chooses the safe lottery. This strategy is followed, for example, by Holt and Laury (2002) and, for a game measuring loss aversion, by Abeler et al. (2011). It has the important advantage that it does not require the assumption that farmers' utility function is correctly captured by the CRRA formulation. This is appealing since, as I explain below, only about half of the subjects conform strictly with the prediction of monotonic switching of the CRRA model. Thus, the 'number of safe choices' is my preferred measure of risk aversion.

I also calculate an explicit CRRA coefficient. Famers close to risk neutrality should always choose the risky lottery. Particularly risk averse farmers, on the other hand, should choose the safe lottery in all six decisions. Farmers between these two extreme cases should choose the risky lottery for one or more of the initial pairs and then, when the higher expected value of the risky lottery is not sufficient to offset the cost of risk, they should switch to the safe lottery for the remaining decisions. I say that players who follow one of these three patterns conform with 'monotonic switching'.

I find that 130 farmers (57.5 percent) conform with 'monotonic switching'. A second group of 33 farmers (14.5 percent) switches to the safe choice twice. Examples of these two patterns are represented in figure 3.2.

Figure 3.2: Patterns of decisions



(a) Monotonic switching (b) Switches twice

I can derive a CRRA coefficient for a monotonic switcher by taking the mid-point of the bounds on risk aversion implied by his pattern of choices. Furthermore I assume that farmers who always choose the risky lottery have a CRRA coefficient of 1 and farmers who always choose the safe lottery are risk neutral. For farmers who switch twice, I calculate CRRA by taking the mid-point between (i) the value of CRRA consistent with the first switch and (ii) the value of CRRA consistent with the second switch.[13] This amounts to treating the fact that the farmer switches twice as a form of mistake. Following this logic, the two switching points give bounds on where the actual desired switching point lies. This is a relatively strong assumption to make. I thus always study the robustness of regression results to the exclusion of individuals who switch twice. To avoid confusion, I refer to two CRRA variables: CRRA1, which is set to missing for farmers who switch twice, and CRRA2, which includes values both for monotonic switchers and for farmers who switch twice.

I can calculate CRRA2 for 72 percent of the farmers who play the game. It is not clear whether the subjects who switch multiple times do so because they have a limited

[13]These are the values of CRRA which I would attribute to a monotonic switcher who switches at that lottery.

understanding of the game, they are truly indifferent, or whether their decisions follow a logic that is not captured by the CRRA expected utility framework. A number of studies in the literature enforce monotonic switching: once a player chooses the safe lottery, he is not allowed to choose the risky lottery in subsequent tasks (Tanaka et al., 2010; Liu, 2013). Holt and Laury (2002) allow subjects to switch between the two lotteries as many times as they like. They find that a number of subjects do not conform with monotonic switching. For example, in the first set of decisions of their experiment, 13 percent of farmers switch more than once. In urban Ghana, Falco (2014) is unable to calculate a coefficient of risk aversion for 10 percent of his sample. Dohmen et al. (2011) and de Brauw and Eozenou (2014) also allow players to switch as many times as they like, but find that almost all individuals conform with monotonic switching.

To summarise, these are the three measures of risk aversion that I will use in the analysis:

1. **Number of safe choices**;
2. **CRRA1**: CRRA coefficient, calculated for the sample of monotonic switchers;
3. **CRRA2**: CRRA coefficient, calculated for the sample of farmers who are monotonic switchers or who switch twice.

In figure 3.11 in the appendix I present the distribution of risk aversion under the three measures. The distribution of the number of safe choices is roughly symmetric around a mode of three safe choices. For both CRRA1 and CRRA2, on the other hand, the mode is at 0. The mean of CRRA1 is 0.50 (st. dev. 0.34) and the mean of CRRA2 is 0.52 (st. dev. 0.31). These values are similar, for example, to Tanaka et al. (2010), who document that among vietnamese farmers the curvature of the utility function is 0.59 for farmers in northern villages and 0.63 for farmers in southern villages. The game of Tanaka et al. (2010) differs from that Holt and Laury (2002) and allows the authors to further estimate a loss aversion parameter and a probability weighting factor, in line with prospect theory.[14] Liu (2013) replicates the game of Tanaka et al. (2010) among chinese farmers and finds an average coefficient of risk aversion of 0.48 (st. dev. 0.33). Falco (2014), for workers in urban Ghana and using a game similar to the one I am using, finds an average CRRA of 0.43.[15]

Controls

I control for a number of factors that could be correlated both with difference in risk aversion and with willingness to seek advice. If left in the error term, these factors would cause assumption (3.5) to fail. Past studies have documented correlation between risk aversion and age, education, gender, income, wealth, parental background, height (Binswager, 1980; Cardenas and Carpenter, 2008; Dohmen et al., 2011). Santos and Barrett (2010) find that advice exchange is correlated with similar amount of experience with a crop, while Conley and Udry (2004) report correlations with similar soil quality and, to a lesser extent, differences in wealth.

I include controls for all the main correlates of risk aversion apart from parental background and height. Because risk-aversion enters the model as the difference between the attitudes of farmer j and farmer i, I control for the difference in age, land owned and number of animals owned. Land and animal ownership are two forms of wealth that are directly related to income generation. I also include dummies for having the same gender, the same education level, and for belonging to the same kinship.[16] I do not have data on soil quality, nor on total years of experience with growing cocoa. To capture similarity in technology, I introduce a dummy for pairs where both farmers use the same technology.[17]

In table 3.4 I report summary statistics of the variable of interest for the network

[14] The curvature parameter I report is comparable to the CRRA parameter I estimate.

[15] Cardenas and Carpenter (2008) summarise a number of studies on risk aversion in developing countries. The estimates they report are often in a similar range to the studies mentioned above.

[16] I treat discordant answers as mistakes and consider that farmer i and farmer j belong to the same kinship only when both report so.

[17] This dummy takes the value of one if both farmers use fertiliser or if neither uses fertiliser.

sample. Average age is 55 years, about 77 percent of farmers are male, 61 percent have completed a level of education beyond primary school. Only 28 percent of farmers use inorganic fertiliser.

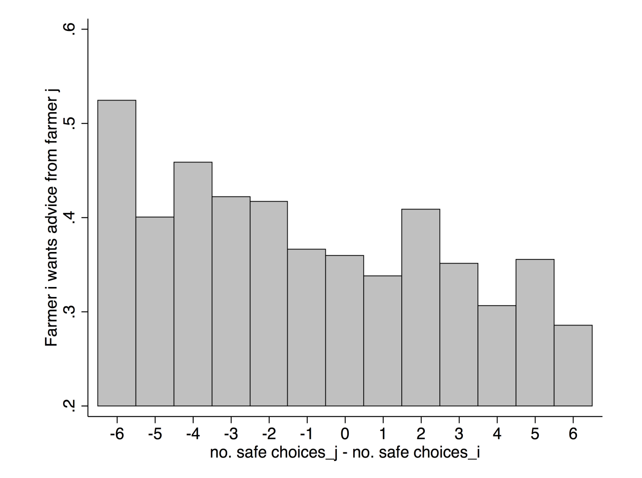
Table 3.4: Summary statistics

Variable	Mean/Percentage	Std. Dev.	N
Male	76.7%		223
Post-primary education	61%		223
Age	55	15.7	223
Number of animals	18.5	27.9	223
Land owned	6.7	8.4	216
Uses fertiliser	28.3%		223
Number of safe choices in risk game	2.9	1.8	225
CRRA2	0.52	0.31	163
CRRA1	0.50	0.34	130

3.4 Results

I start by plotting the relationship between difference in risk aversion and willingness to seek advice. Figure 3.3 shows a clear negative association between the two variables: at higher values of difference in risk aversion, farmer i is less likely to be willing to seek advice from farmer j.

Figure 3.3: Relationship between difference in risk aversion and willingness to seek advice



I present the main results in table 3.5 below. I am able to detect a negative effect of the difference between farmer j and farmer i's risk aversion on farmer i's willingness to seek advice from farmer j. The effect is in the hypothesised direction. One standard deviation increase in the difference in risk aversion is associated with a 2.8 percentage point decrease in the likelihood that farmer i would like advice from farmer j. This corresponds to a 7.4 percent reduction compared to the probability that farmer i would like advice from a farmer with the same level of risk aversion. If I compare a pair of farmers with the same level of risk aversion to a pair with the maximum difference in risk aversion, the reduction in the likelihood of wanting advice is 7 percentage points (18.5 percent). These are meaningful magnitudes when compared to other effects. For example, having the same education level is associated with an increase in the probability that farmer i would like to seek farmer j's advice of 4.8 percentage points, while an increase in the age difference of one standard deviation raises this probability by 6.2 percentage points.

When I do not include controls, the estimate has a p-value of .104, so it is just above the conventional level of statistical significance. When I include controls, the magnitude of the effect goes up by 1 percentage point, and significance is at the 5 percent level.

I report on the goodness-of-fit of the models at the bottom of table 3.5. The R^2 statistics performs poorly for the linear probability model, so I omit it. Instead, I use the model to predict whether farmer i would like to learn from farmer j and compare the predictions with the data. This is the strategy recommended by Wooldridge (2012).[18] When I include controls, the model correctly predicts almost 63 percent of the observations. I correctly predict 'successes' (i wants advice from j = 1) and 'failures' (i wants advice from j = 0) at a similar rate.

I explore three factors that could be affecting my estimates. First: is the effect of the

[18]After estimation of the linear probability model, the OLS fitted value gives a predicted 'probability of success' for each dyad, that is, a prediction of the probability that farmer i would like to seek advice from farmer j. I set a threshold t, such that when the predicted probability of success is greater than t, I predict a success, and I predict a failure otherwise. Following Wooldridge (2012) I set t to the proportion of successes in the sample used for the regression.

Table 3.5: Dyadic linear probability model

	(1)	(2)
risk aversion$_j$ - risk aversion$_i$	-.028 (.104)	-.038 (.046)**
Const.	.378 (.000)***	.589 (.070)*
Obs.	2166	2010
% correctly predicted	49.68	62.84
% successes correctly predicted	60.02	65.21
% failures correctly predicted	43.03	61.40
Cluster N	22	22
Risk aversion measure	No. safe choices	No. safe choices
Controls		✔

Dyadic OLS regression. The dependent variable is a dummy which takes a value of one if farmer i would like to receive advice from farmer j. The regression in column 2 includes controls for difference in age, land owned, and number of animals owned between farmer j and farmer i. Further controls include dummies for having the same education level, same gender, having the same fertiliser adoption pattern, and village fixed effects. Standard errors are corrected for clustering at the village level. P-values obtained with wild bootstrap-t procedure reported in parentheses. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%.

difference in risk aversion more pronounced when farmer i interacts frequently with farmer j? In a risk pooling experiment in rural Colombia, for example, Attanasio et al. (2012) find that only close friends match assortatively on risk attitudes. This effect could be driven by the fact that individuals only have information about the aversion to risk of the people with whom they communicate frequently. I explore the possibility that the effect is stronger among socially proximate farmers in model (3.7), where I interact difference in risk aversion with a dummy for whether farmer i has spoken with farmer j in the last 7 days. This dummy proxies for frequent interaction between the two farmers. Results are reported in table 3.6 below. In my first specification, farmer i is almost 50 percent more likely to desire advice from farmer j (an increase in probability of 24.2 percentage points) when the two farmers have spoken in the previous seven days. The effect is significant at the 1 percent level. Among pairs of farmers who have spoken in the previous week, one standard deviation increase in the difference in risk aversion is associated with a 4.6 percentage point increase in the likelihood that farmer i wants advice from farmer j. This effect however is statistically imprecise. In column 2 of table 3.6 I independently include difference in risk aversion. Now the coefficient on the interaction term captures whether the effect of difference in risk aversion is stronger among pairs of farmers who have spoken in the last 7 days. The negative coefficients I estimate suggest that the effect among socially proximate farmers is 1 percentage point stronger. However, this effect is not significant.

Table 3.6: Dyadic linear probability model: interactions

	(1)	(2)	(3)	(4)
(risk aversion$_j$ - risk aversion$_i$) * ij speak	-.046 (.232)	-.010 (.833)		
ij speak	.242 (.000)***	.242 (.000)***		
(risk aversion$_j$ - risk aversion$_i$)* j wants advice from i			-.051 (.022)**	-.021 (.080)*
j wants advice from i			.006 (.887)	.009 (.837)
risk aversion$_j$ - risk aversion$_i$		-.036 (.086)*		-.030 (.110)
Const.	.541 (.080)*	.541 (.082)*	.596 (.144)	.588 (.128)
Obs.	2010	2010	2010	2010
Cluster N	22	22	22	22
Risk aversion measure	No. safe choices	No. safe choices	No. safe choices	No. safe choices
Controls	✔	✔	✔	✔

Dyadic OLS regression. The dependent variable is a dummy which takes a value of one if farmer i would like to receive advice from farmer j. All regressions include controls for difference in age, land owned, and number of animals owned between farmer j and farmer i. Further controls include dummies for having the same education level, same gender, having the same fertiliser adoption pattern, and village fixed effects. Standard errors are corrected for clustering at the village level. P-values obtained with wild bootstrap-t procedure reported in parentheses. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%.

Second: do farmers only report interest in the advice of those peers with whom they will be able to communicate? And does this depress the effect of difference in risk aversion which I estimate? I hypothesise that farmer i's request for advice is unlikely to be refused when farmer j also reports to be interested in farmer i's advice. In model (3.8), I control for farmer j's desire for farmer i's advice, and I further include the interaction between farmer j's desire and the difference in risk aversion between farmer j and farmer i. Columns 3 and 4 of table 3.6 report the results of this analysis. Before I include controls, I estimate that farmer i is almost 8 percentage points more likely to be willing to link with farmers that reciprocate his desire. However, this effect disappears once I include controls. Among pairs where farmer j would like to receive farmer i's advice, one standard deviation increase in the difference in risk aversion is associated with a 5.1 percentage point decrease in farmer i's willingness to seek advice from farmer j. The regression in column 4 reports that the difference in the effect between pairs where farmer j desires farmer i's advice and pairs where he does not is a significant 2 percentage points.

After estimation of model (3.8) I cannot reject that $\beta_1 = 0$. In other words, when the two farmers have the same level of risk aversion, farmer j's desire for farmer i's advice is not a significant predictor of farmer i's willingness to seek j's advice. This is inconsistent with both interpretation I have initially proposed: (the lack of) anticipated rejection or technological affinity. If farmer j's desire for farmer i's advice proxies for either of these, it should also increase the probability that i wants to learn from j when both farmers have the same level of risk aversion. Another possibility is that when farmer j is interested in farmer i's advice, perhaps because farmer i knows about topics of particular interest to j, the risk that farmer j free rides on the information produced by farmer i is higher. Then farmer i (i) is not concerned about farmer j's interest in his advice when they both have the same level of risk aversion, and (ii) when farmer j is more risk averse than farmer i, farmer i is particularly concerned to avoid a link with farmer j.

Third: is the effect I study confounded by the fact that some risk averse farmers are valuable advisors because of the information they obtain from others? To answer this question I include among the controls the number of farmers whom farmer j approached for advice in the past. This variable should proxy for the amount of information farmer j has accumulated by learning from others. To perform this exercise, I have to restrict attention to the sample of farmers for whom I have data on actual communication attempts. In column 1 of table 3.7 I thus first report the estimates of my baseline model for this restricted sample. In columns 2 and 3 I then control for the number of farmers whom farmer j approached for advice. In column 3, I add a full set of controls. The point estimates I obtain in these three regressions are virtually identical. So I have no evidence that failing to control for actual links is affecting the estimates. The direct effect I estimate for the number of (potential) information sources of j is positive, but small and insignificant.

To sum up, I have documented a negative effect of difference in risk aversion on the likelihood that farmer i desires advice from farmer j. This effect is of a meaningful economic magnitude and becomes stronger when farmer j is interested in farmer i's

Table 3.7: Dyadic linear probability model: farmer j's requests for other farmers' information

	(1)	(2)	(3)
risk aversion$_j$ - risk aversion$_i$	-.042 (.052)*	-.042 (.062)*	-.043 (.020)**
No. farmers approached for advice$_j$		.0005 (.933)	.020 (.557)
Const.	.420 (.000)***	.413 (.104)	.419 (.555)
Obs.	1208	1208	1154
Cluster N	10	10	10
Risk aversion measure	No. safe choices	No. safe choices	No. safe choices
Controls			✔

Dyadic OLS regression. The dependent variable is a dummy which takes a value of one if farmer i would like to receive advice from farmer j. The regression in column 3 includes controls for difference in age, land owned, and number of animals owned between farmer j and farmer i. Further controls include dummies for having the same education level, same gender, having the same fertiliser adoption pattern, and village fixed effects. Standard errors are corrected for clustering at the village level. P-values obtained with wild bootstrap-t procedure reported in parentheses. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%.

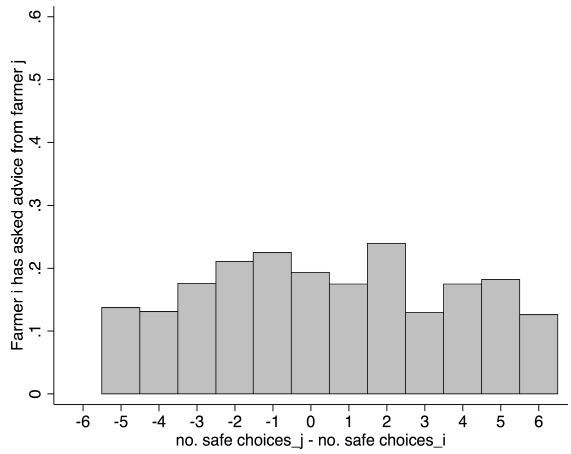
advice, and among pairs of farmers who have recently spoken (albeit insignificantly so in this last case). Is difference in risk aversion also associated with actual requests for information exchange? I provide some evidence to answer this question by estimating model (3.4) using the data on advice requests that is available for the subsample of 10 villages. Table 3.8 shows that estimated coefficients are insignificant and close to 0. The plot in figure 3.4 confirms these results graphically: there is no clear relationship between the two variables. This could be due to a number of reasons. For example, risk-averse farmers may transfer resources in exchange for the information provided by their more risk-tolerant advisors. The data I have does not allow me to propose a convincing test of this hypothesis.

Table 3.8: Dyadic linear probability model: request for advice

	(1)	(2)
risk aversion$_j$ - risk aversion$_i$	.001 (.895)	-.001 (.925)
Const.	.189 (.000)***	.173 (.076)*
Obs.	1208	1154
Cluster N	10	10
Risk aversion measure	No. safe choices	No. safe choices
Controls		✔

Dyadic OLS regression. The dependent variable is a dummy which takes a value of one if farmer i has asked for advice from farmer j. The regression in column w includes controls for difference in age, land owned, and number of animals owned between farmer j and farmer i. Further controls include dummies for having the same education level, same gender, having the same fertiliser adoption pattern, and village fixed effects. Standard errors are corrected for clustering at the village level. P-values obtained with wild bootstrap-t procedure reported in parentheses. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%.

Figure 3.4: Relationship between difference in risk aversion and request for advice



Robustness

I offer four types of robustness tests of the main result. First, I look for evidence of mis-specification of functional form. I have studied the role of the difference in the risk attitudes between farmer i and farmer j. This variable is mechanically correlated with the risk aversion of farmer j. In table 3.10 I study the effect of the risk aversion of j on willingness to seek advice. The effect I estimate is smaller than that of the difference in risk aversion (one standard deviation increase reduces willingness to seek advice by 1.2-2 percentage points) and never significant. Without a formal test, I cannot interpret these differences as statistically significant evidence in favour of my specification, but I am reassured to see that the qualitative comparison favours my model. In columns 3 and 4 I also report estimates of the effect of the absolute difference in risk aversion on willingness to seek advice. This variable has no correlation with difference in risk aversion. I estimate a coefficient that is very small and insignificant.

Second, in table 3.11 I further explore the robustness of my specification. In columns 1 and 2 I include a quadratic term, which has a small and insignificant coefficient. In columns 3 and 4 I study the symmetry of positive and negative effects. I replace the continuous difference in risk aversion with dummies for whether farmer j has greater or smaller risk aversion than farmer i. Pairs with the same level of risk aversion are the omitted category. The coefficients I estimate are in the hypothesised direction: when farmer j is more risk averse than i, farmer i is *less likely* to desire his advice; when farmer j is less risk averse, farmer i is *more likely* to be interested in his advice.

Third, I study whether difference in risk aversion is a significant predictor of other types of interaction between farmers. Table 3.12 shows the estimation results. Difference in risk aversion is not significantly associated with knowledge of farmer j, with speaking to farmer j, lending him money, or walking past his farm. The coefficients on all these interactions are generally very small: for example, an increase in the difference in risk aversion of one standard deviation increases the probability of lending money to farmer j by 0.2-0.3 percentage points. The effect of difference in risk aversion that I have uncovered seems unique to willingness to seek to advice. This rules out unobserved heterogeneity that is correlated both with difference in risk aversion and with generic social interaction.

Fourth, I repeat the analysis of tables 3.5, 3.6, 3.7 using the measures of risk aversion CRRA1 and CRRA2. In table 3.13 I report estimation of the basic model for the restricted samples for which I can calculate a coefficient of risk aversion. I present next to each other estimates using the three measures. The coefficients I estimate for the restricted sample are all in the hypothesised direction. However, when I do not include controls, the coefficients are small. When I include controls and using CRRA1 (CRRA2), I estimate that a standard deviation increase in the difference in risk aversion is associated with a 2.5 (2) percentage point decrease in the likelihood that farmer i is interested in farmer j's advice. The magnitude of this effect is similar to that estimated in table 3.5, however, in the smaller sample, it is not significant. In tables 3.14, 3.15, 3.16 I study models (3.7), (3.8), (3.9) using CRRA1 and CRRA2. The results are qualitatively similar to those reported in tables 3.6 and 3.7, but, again the effects are statistically imprecise in the smaller sample.[19]

[19]The sign of coefficient on the difference in risk aversion turns positive when I use CRRA2 and further re-

3.5 Conclusion

This chapter studies the characteristics that individuals value in the people they approach for advice. Using cross-sectional data on 225 cocoa farmers in 22 Ghanaian villages and a matched lottery experiment, I find an association between the difference in the aversion to risk of two farmers and the probability that one farmer is interested in the other farmer's advice. This probability decreases by 7.4 percent when the difference in risk attitudes between the potential advisor and the farmer seeking advice increases by one standard deviation. The effect is robust to the inclusion of a set of controls that may be correlated with difference in risk aversion and willingness to seek advice, and does not extend to other dimensions of interaction in networks. The effect of difference in risk aversion is stronger when the desire for advice is reciprocated.

I propose an explanation of this results using a simple of model of strategic experimentation between two farmers with different risk preferences. The model shows that in equilibrium the farmer who is more risk averse (in a specific sense) free rides on the information produced by his partner. I argue that, in anticipation of this, farmers would like to avoid advice exchange with more risk averse peers.

I provide evidence that individuals strategically consider the preferences of others. This carries implications for policy as interventions are often targeted to specific groups defined by preferences. For example, the provision of insurance decreases the cost of trying out new technologies for risk-averse individuals (Karlan et al., 2012) and thus reduces the rationale to free ride on the experiments of others. This is likely to change the way information networks are formed. It may for instance allow risk-averse individuals to establish a larger number of connections, or to reduce transfers to risk-tolerant peers given in exchange for their advice. Interventions that allow time-inconsistent individuals to commit (Ashraf et al., 2006), or individuals concerned with their social image to signal pro-sociality (Ashraf et al., 2012), may also influence the

strict the analysis to the villages for which I have information about advice requests. However, the magnitude is very close to 0. I do not think this is suggestive of particular problems.

incentives to associate and exchange information and favours. A full assessment of the impacts of these interventions has to take into account their effects on the structure of networks.

Appendix

A.1 The model

A.1.1 Expected utility given period 1 strategies

When both farmers choose technology L in period 1, player i does not update his beliefs. Farmer i is thus going to earn profit y_L for period 1. For period 2, he is going to earn profits y_H if $\mu_0 - \frac{1}{2}r_i\sigma_0^2 \geq y_L$ and he applies technology H, and profits y_L otherwise. Expected utility is then given by:

$$V_i(LL) = \begin{cases} E_1(1-\exp\{-r_i(y_L+y_H)\}) & \text{if } \mu_0 - \frac{1}{2}r_i\sigma_0^2 \geq y_L \\ E_1\left(1-\exp\{-r_i(2y_L)\}\right) & \text{otherwise} \end{cases}$$

In the second case, the payoff is certain, so I can drop the expectation sign. In the first case, I know the distribution of y_H and I can calculate the value of the expectation. I can then express $V_i(LL)$ as:

$$V_i(LL) = \begin{cases} 1-\int_{-\infty}^{\infty}\exp\{-r_i(y_L+y_H)\}\frac{1}{\sigma_0}\phi\left(\frac{y_H-\mu_0}{\sigma_0}\right)\mathrm{d}y_H & \text{if } \mu_0 - \frac{1}{2}r_i\sigma_0^2 \geq y_L \\ 1-\exp\{-r_i(2y_L)\} & \text{otherwise} \end{cases} \tag{3.10}$$

I use $V_i(LL,L)$ to refer to expected utility when nobody experiments in period 1 and farmer i chooses technology L in period 2. Similarly $V_i(LL,H)$ refers to expected utility when the farmer chooses the risky technology in period 2 and nobody has experimented in period 1.

I will now look at the case where, in period 1, farmer i uses technology L and farmer j uses technology H. Farmer i 's profits in period 1 are given by y_L. On the basis of the information that will be produced by farmer j, in period 1 farmer i expects period 2 profits to be defined by the random variable $z = \max(y_H, y_L)$. In other words, in period 2 he will at least make profits of y_L and, with a positive probability, he will make profits above y_L.

$$\begin{aligned} V_i(LH) &= 1 - E_1\left(\exp\{-r_i\left(y_L + \max\left(y_H, y_L\right)\right)\}\right) \\ &= 1 - E_1\left(\exp\{-r_i \max[(y_H + y_L), 2y_L]\}\right) \end{aligned}$$

Let $W_1 = \max[(y_H + y_L), 2y_L]$. W_1 has the following censored distribution:

$$W_1 = \begin{cases} y_H + y_L & \text{if } y_H \geq y_L \\ 2y_L & \text{otherwise} \end{cases}$$

I can thus now calculate expected utility $V_i(LH)$:

$$\begin{aligned} V_i(LH) = & 1 - \int_{-\infty}^{y_L} \exp\{-r_i 2y_L\} \frac{1}{\sigma_0}\phi\left(\frac{y_H - \mu_0}{\sigma_0}\right) \mathrm{d}y_H - \\ & \int_{y_L}^{\infty} \exp\{-r_i(y_H + y_L)\} \frac{1}{\sigma_0}\phi\left(\frac{y_H - \mu_0}{\sigma_0}\right) \mathrm{d}y_H \end{aligned} \tag{3.11}$$

Now, let me consider the case where farmer i experiments with technology H in period 1, while farmer j does not experiment. Here I have:

$$\begin{aligned} V_i(HL) &= 1 - E_1\left(\exp\{-r_i(y_H + \max(y_H, y_L))\}\right) \\ &= 1 - E_1\left(\exp\{-r_i \max(2y_H, y_H + y_L)\}\right) \end{aligned}$$

Let $W_2 = \max[(y_H + y_L), 2y_H]$. W_2 has the following censored distribution:

$$W_2 = \begin{cases} 2y_H & \text{if } y_H \geq y_L \\ y_H + y_L & \text{otherwise} \end{cases}$$

I can thus now calculate expected utility $V_i(HL)$:

$$\begin{aligned} V_i(HL) =& 1 - \int_{-\infty}^{y_L} \exp\{-r_i(y_H + y_L)\} \frac{1}{\sigma_0} \phi\left(\frac{y_H - \mu_0}{\sigma_0}\right) \mathrm{d}y_H - \\ & \int_{y_L}^{\infty} \exp\{-r_i 2y_H\} \frac{1}{\sigma_0} \phi\left(\frac{y_L - \mu_0}{\sigma_0}\right) \mathrm{d}y_H \end{aligned} \tag{3.12}$$

Finally consider the case where both farmers experiment in period 1. Expected utility is given by:

$$V_i(HH) = 1 - E_1\left(\exp\{-r_i(y_H + \max(y_H, y_L))\}\right)$$

Notice that this is identical to expected utility $V_i(HL)$, which I have just derived. When farmer i experiments with technology H in period 1, the expected utility of this farmer is not affected by whether farmer j experiments or not, as the additional experiment does not provide new information.

These expressions allow me to compare the expect utility from the alternative strategies available to farmer i. In the next section I will use numerical methods to make such comparisons. Here I can prove the following:

Proposition 3.1. $V_i(HL) > V_i(LL, H)$.

Proposition 1 says that, when farmer j is not experimenting, trying the risky technology H in period 1 is superior to trying it for the first time in period 2. In the latter case, farmer i uses each technology once. In the former case, farmer i uses technology H in period 1, and then the best technology in period 2. Hence he cannot do worse than in the first case. In expectation he is better off, because he can exploit the states of the world where $y_H^* > y_L$. I also offer a mathematical proof of this.

$V_i(HL) > V_i(LL, H)$ implies:

$$\int_{-\infty}^{\infty} \exp\{-r_i(y_H + y_L)\} \frac{1}{\sigma_0} \phi\left(\frac{y_H - \mu_0}{\sigma_0}\right) \mathrm{d}y_H > \\ \int_{-\infty}^{y_L} exp\{-r_i(y_H + yL)\} \frac{1}{\sigma_0} \phi\left(\frac{y_H - \mu_0}{\sigma_0}\right) \mathrm{d}y_H + \\ \int_{y_L}^{\infty} \exp\{-r_i 2y_H\} \frac{1}{\sigma_0} \phi\left(\frac{y_H - \mu_0}{\sigma_0}\right) \mathrm{d}y_H \tag{3.13}$$

I can rewrite the above as:

$$\int_{-\infty}^{y_L} \exp\{-r_i(y_H + y_L)\} \frac{1}{\sigma_0} \phi\left(\frac{y_H - \mu_0}{\sigma_0}\right) \mathrm{d}y_H + \\ \int_{y_L}^{\infty} \exp\{-r_i(y_H + y_L)\} \frac{1}{\sigma_0} \phi\left(\frac{y_H - \mu_0}{\sigma_0}\right) \mathrm{d}y_H > \\ \int_{-\infty}^{y_L} \exp\{-r_i(y_H + y_L)\} \frac{1}{\sigma_0} \phi\left(\frac{y_H - \mu_0}{\sigma_0}\right) \mathrm{d}y_H + \\ \int_{y_L}^{\infty} \exp\{-r_i 2y_H\} \frac{1}{\sigma_0} \phi\left(\frac{y_H - \mu_0}{\sigma_0}\right) \mathrm{d}y_H \tag{3.14}$$

I can cancel out the first term on each side of the inequality, as these two terms are identical. I am then left with the following condition:

$$\int_{y_L}^{\infty} \exp\{-r_i(y_H+y_L)\}\frac{1}{\sigma_0}\phi\left(\frac{y_H-\mu_0}{\sigma_0}\right)\mathrm{d}y_H > \int_{y_L}^{\infty} \exp\{-r_i 2y_H\}\frac{1}{\sigma_0}\phi\left(\frac{y_H-\mu_0}{\sigma_0}\right)\mathrm{d}y_H \tag{3.15}$$

Both integrals in (3.15) are defined over the same interval $[y_L,\infty)$. For all values of y_H in this interval, $y_H \geq y_L$ and $\exp\{-r_i(y_H+y_L)\} \geq \exp\{-r_i 2y_H\}$. For all values of y_H in (y_L,∞), $y_H > y_L$ and $\exp\{-r_i(y_H+y_L)\} > \exp\{-r_i 2y_H\}$. As y_H is normally distributed with infinite support, there is some positive density over values in the interval (y_L,∞). This guarantees that (3.15) holds as a strict inequality.

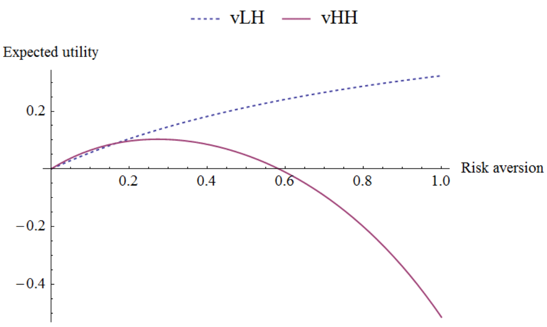
A.1.2 Result 1

In this subsection I compute the values of $V_i(LH)$ and $V_i(HH)$ for various levels of risk aversion in the interval (0,1]. I also compute $V_i(LH)-V_i(HH)$. When this difference is positive, farmer i prefers to choose the safe technology L when farmer j chooses the risky technology H in period 1. I show that there is a cut-off level of risk aversion that determines the sign of $V_i(LH)-V_i(HH)$. Farmers with risk aversion above the cut-off, prefer the safe technology. Farmers with risk aversion below the cut-off, prefer the risky technology. I call the cut-off level of risk aversion r_i^*.

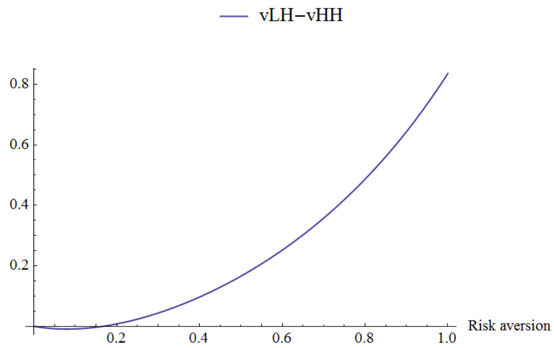
To find a numerical value for the expected utility, I need to posit values for y_L, μ_0 and σ_0. Throughout this section I fix $y_L = 0$. In the first simulation, I set $\mu_0 = 0.24$ and $\sigma_0 = 1.2$. The increase in expected profits in this simulation is 0.2 time a standard deviation of y_H. Results are presented in figure 3.5 below. The left panel shows separately the values of $V_i(LH)$ and $V_i(HH)$, for different levels of risk aversion. The right panel shows the difference $V_i(LH)-V_i(HH)$. For low levels of risk aversion, this difference is negative. A farmer close to risk neutrality would like to experiment in period 1 even when farmer j is experimenting. For levels of risk aversion above a CRRA

Figure 3.5: Simulation 1

$\mu_0 = 0.24$, $\sigma_0 = 1.2$



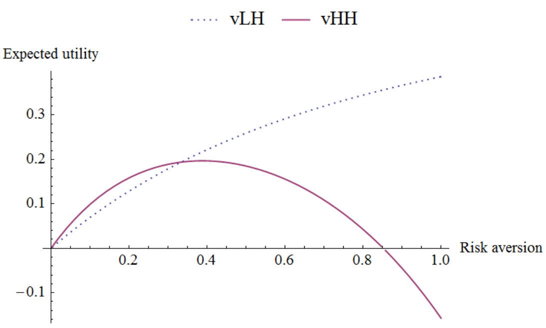
(a) $V_i(LH)$ and $V_i(HH)$



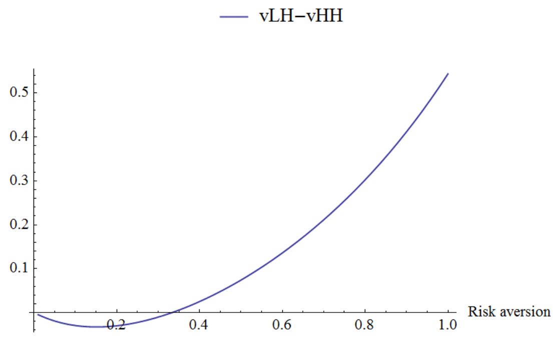
(b) $V_i(LH) - V_i(HH)$

Figure 3.6: Simulation 2

$\mu_0 = 0.48$, $\sigma_0 = 1.2$



(a) $V_i(LH)$ and $V_i(HH)$



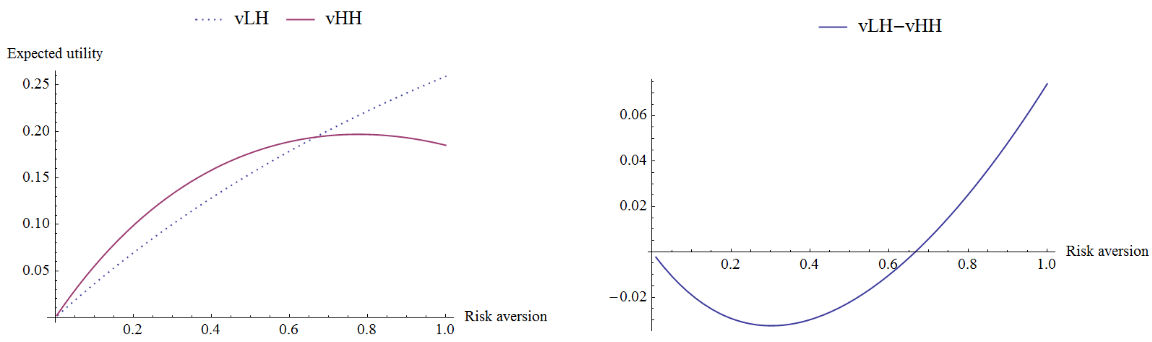
(b) $V_i(LH) - V_i(HH)$

coefficient of 0.15, the expected utility from the safe technology becomes higher.

In the second simulation, presented in figure 3.6, I double the value of μ_0. Now the increase in expected profits is 0.4 of a standard deviations of y_H. This makes the risky technology more appealing. It now takes more risk aversion to turn down the risky technology. The cut-off level of risk aversion at which the farmer is indifferent between the two technologies is now higher (close to 0.35).

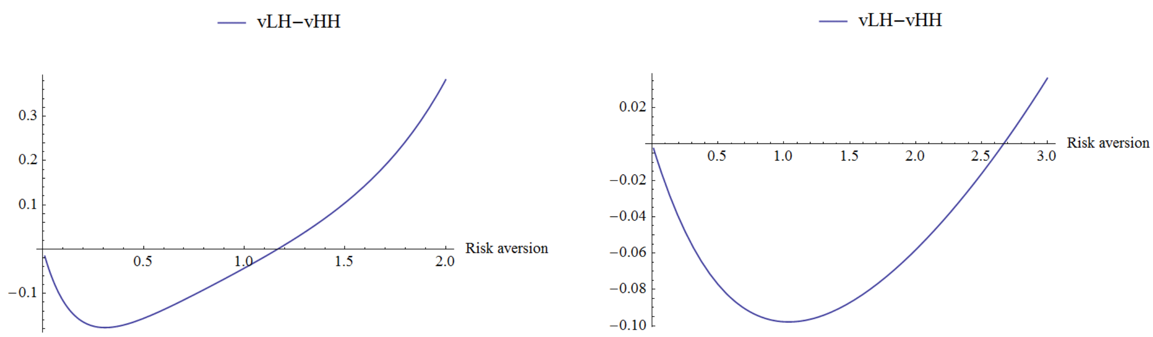
In the third simulation, I halve the value of σ_0. The brings the increase in expected profits to 0.4 times a standard deviation of y_H, as in simulation 2. Similarly, the effect is to increase the cut-off point. Lower dispersion makes the risky technology more

Figure 3.7: Simulation 3
$\mu_0 = 0.24$, $\sigma_0 = 0.6$



(a) $V_i(LH)$ and $V_i(HH)$ (b) $V_i(LH) - V_i(HH)$

Figure 3.8: Simulations 4 and 5



(a) $\mu_0 = 1.68$, $\sigma_0 = 1.2$ (b) $\mu_0 = 0.24$, $\sigma_0 = 0.3$

appealing, and, compared to simulation 1, it now takes a larger risk aversion to forgo the risky technology.

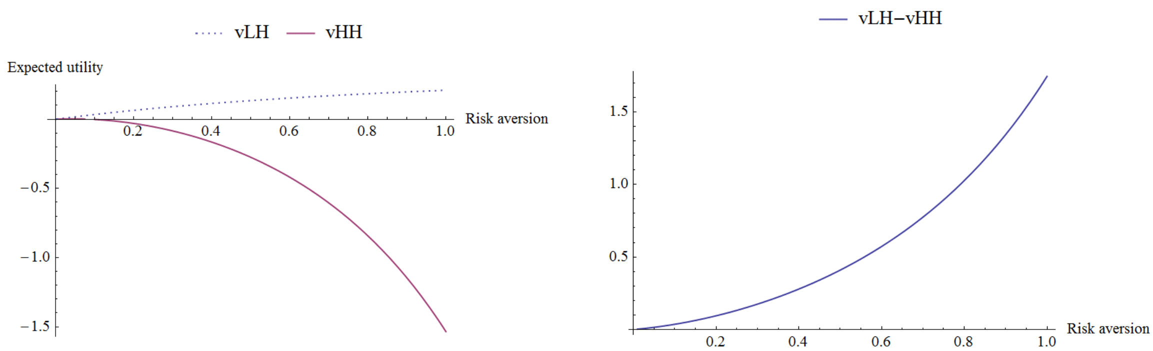
If I further increase μ_0 or decrease σ_0, the cut-off occurs for levels of risk aversion outside of the interval I have considered so far. Figure 3.8 illustrates. Finally, if the expected value of the risky technology is negative, farmer i always chooses technology L when farmer j chooses H. I present this result in figure 3.9.

A.1.3 Result 2

I now want to compare $V_i(HL)$ and $V_i(LL)$. Proposition 1 tells me that $V_i(HL) > V_i(LL,H)$. I thus focus on the comparison between $V_i(HL)$ and $V_i(LL,L)$, If I find a level of risk aversion such that $V_i(LL,L) > V_i(HL)$, I know that for that level of risk

Figure 3.9: Simulation 6

$\mu_0 = -0.24, \sigma_0 = 1.2$



(a) $V_i(LH)$ and $V_i(HH)$

(b) $V_i(LH) - V_i(HH)$

aversion $V_i(LL, L) > V_i(LL, H)$ and hence farmer i will indeed choose technology L in period 2. In the rest of this section, I use $V_i(LL)$ to refer to the longer $V_i(LL, L)$.

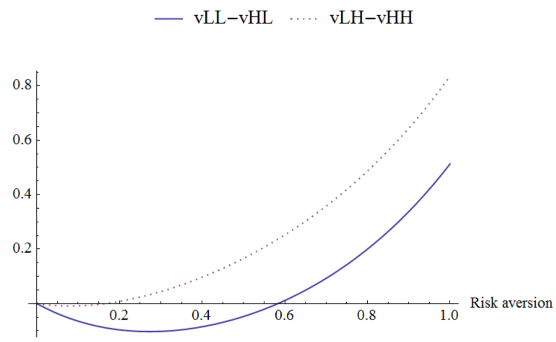
As for result 1, I fix the value of y_L to 0. This means that the value of $V_i(LL)$ is also fixed as 0. I thus do not report separate values for $V_i(HL)$ and $V_i(LL)$. To facilitate comparison with result 1 above, I report both the value of $V_i(LL) - V_i(HL)$[20] and of $V_i(LH) - V_i(HH)$. Whenever $V_i(LL) - V_i(HL)$ is positive, farmer i finds it optimal to use technology L when farmer j is not experimenting with the risky technology.

The cut-off property holds also in this set of simulations. I call this second cut-off r^{**}. Further, I find that $r^{**} \geq r^*$. In other words, it takes more risk aversion to prefer the safe technology when farmer j does not experiment with H. The intuition is simple: when farmer j does not experiment with H in period 1, farmer i can produce valuable information by choosing H. This increases the value of technology H compared to technology L.

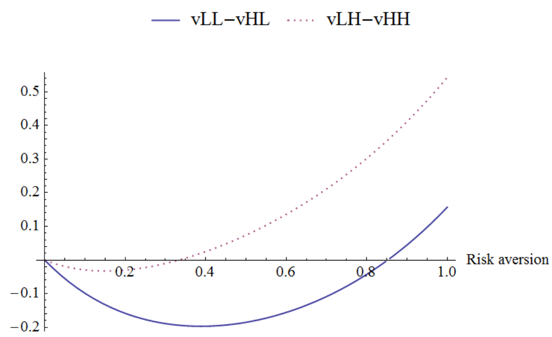
Figure 3.10 illustrates the simulations. The comparative statics of r^{**} as I change μ_0 and σ_0 are qualitatively similar to those of r^*.

[20] $V_i(LL) - V_i(HL) = -V_i(HL)$.

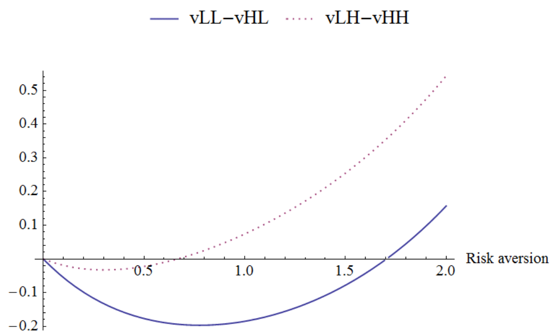
Figure 3.10: Simulations for result 2



(a) $\mu_0 = 0.24$, $\sigma_0 = 1.2$



(b) $\mu_0 = 0.48$, $\sigma_0 = 1.2$



(c) $\mu_0 = 0.24$, $\sigma_0 = 0.6$

A.1.4 Nash equilibria for the 9 games

In what follows I explore the nine possible combinations of risk aversion coefficients presented in table 3.9. For both r^* and r^{**} I assume that, when the farmer has exactly the cut-level of risk aversion, he will choose the risky technology. I restrict attention to equilibria in pure strategies. I underline a player best response to the other player's strategy. For sets of strategies that induce a Nash equilibrium, both payoffs in the corresponding cell are underlined.

Table 3.9: Possible combinations

		Player j		
		$r_j \leq r^* \leq r^{**}$	$r^* \leq r_j \leq r^{**}$	$r^* \leq r^{**} \leq r_j$
Player i	$r_i \leq r^* \leq r^{**}$	case 1	case 4	case 7
	$r^* \leq r_i \leq r^{**}$	case 2	case 5	case 8
	$r^* \leq r^{**} \leq r_i$	case 3	case 6	case 9

Case 1 (H,H) unique equilibrium

		Player j	
		H	L
Player i	H	$\underline{V_i(HH)}$ $\underline{V_j(HH)}$	$\underline{V_i(HL)}$ $V_j(HL)$
	L	$V_i(LH)$ $\underline{V_j(LH)}$	$V_i(LL)$ $V_j(LL)$

Case 2 (L,H) unique equilibrium

		Player j	
		H	L
Player i	H	$V_i(HH)$ $\underline{V_j(HH)}$	$\underline{V_i(HL)}$ $V_j(HL)$
	L	$\underline{V_i(LH)}$ $\underline{V_j(LH)}$	$V_i(LL)$ $V_j(LL)$

Case 3 (L,H) unique equilibrium

		Player j	
		H	L
Player i	H	$V_i(HH)$ $\underline{V_j(HH)}$	$V_i(HL)$ $V_j(HL)$
	L	$\underline{V_i(LH)}$ $\underline{V_j(LH)}$	$\underline{V_i(LL)}$ $V_j(LL)$

Case 4 (H,L) unique equilibrium

		Player j	
		H	L
Player i	H	$\underline{V_i(HH)}$ $V_j(HH)$	$\underline{V_i(HL)}$ $\underline{V_j(HL)}$
	L	$V_i(LH)$ $\underline{V_j(LH)}$	$V_i(LL)$ $V_j(LL)$

Case 5 (H,L) and (L,H) are both equilibria

		Player j	
		H	L
Player i	H	$V_i(HH)$ $V_j(HH)$	$\underline{V_i(HL)}$ $\underline{V_j(HL)}$
	L	$\underline{V_i(LH)}$ $\underline{V_j(LH)}$	$V_i(LL)$ $V_j(LL)$

Case 6 (L,H) unique equilibrium

		Player j	
		H	L
Player i	H	$V_i(HH)$ $V_j(HH)$	$V_i(HL)$ $\underline{V_j(HL)}$
	L	$\underline{V_i(LH)}$ $\underline{V_j(LH)}$	$\underline{V_i(LL)}$ $V_j(LL)$

Case 7 (H,L) unique equilibrium

		Player j	
		H	L
Player i	H	$\underline{V_i(HH)}$ $V_j(HH)$	$\underline{V_i(HL)}$ $\underline{V_j(HL)}$
	L	$V_i(LH)$ $V_j(LH)$	$V_i(LL)$ $\underline{V_j(LL)}$

Case 8 (H,L) unique equilibrium

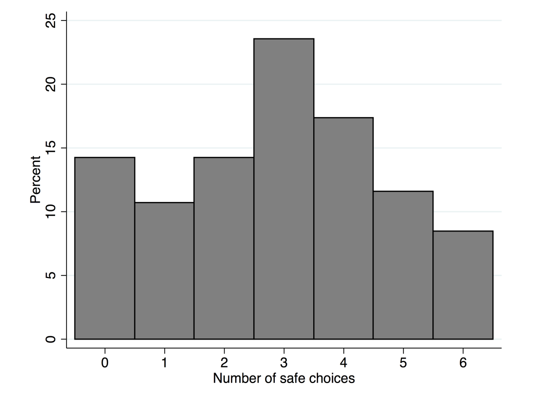
		Player j	
		H	L
Player i	H	$V_i(HH)$ $V_j(HH)$	$\underline{V_i(HL)}$ $\underline{V_j(HL)}$
	L	$\underline{V_i(LH)}$ $V_j(LH)$	$V_i(LL)$ $\underline{V_j(LL)}$

Case 9 (L,L) unique equilibrium

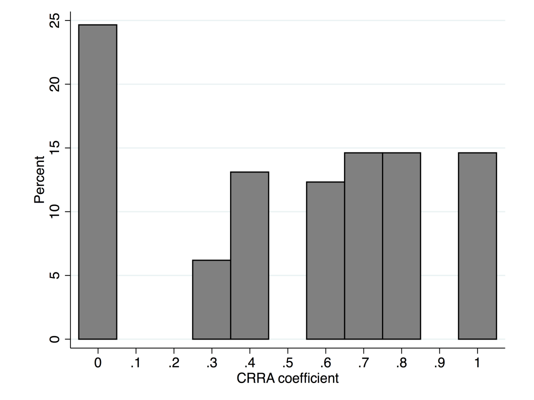
		Player j	
		H	L
Player i	H	$V_i(HH)$ $V_j(HH)$	$V_i(HL)$ $\underline{V_j(HL)}$
	L	$\underline{V_i(LH)}$ $V_j(LH)$	$\underline{V_i(LL)}$ $\underline{V_j(LL)}$

A.2 Figures

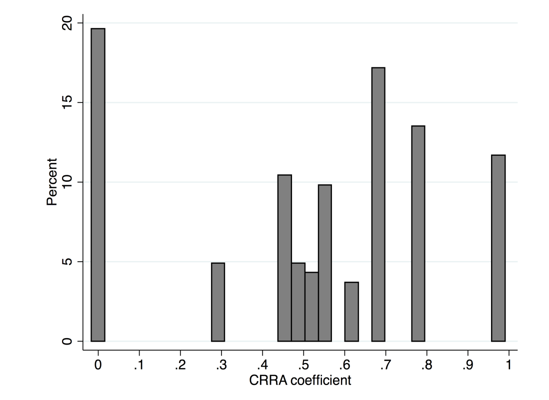
Figure 3.11: Distribution of the different measures of risk aversion



(a) Number of safe choices



(b) CRRA1



(c) CRRA2

A.3 Regression tables

Table 3.10: Dyadic linear probability model: functional form 1

	(1)	(2)	(3)	(4)
risk aversion$_j$	-.020 (.200)	-.013 (.354)		
$\lvert$risk aversion$_j$ - risk aversion$_i\rvert$			.020 (.270)	.001 (.945)
Const.	.411 (.000)***	.599 (.068)*	.361 (.000)***	.588 (.022)**
Obs.	2166	2010	2166	2010
Cluster N	22	22	22	22
Risk aversion	No. safe choices	No. safe choices	No. safe choices	No. safe choices
Controls		✔		✔

Dyadic OLS regression. The dependent variable is a dummy which takes a value of one if farmer i would like to receive advice from farmer j. The regressions in column 2 and 4 include controls for difference in age, land owned, and number of animals owned between farmer j and farmer i. Further controls include dummies for having the same education level, same gender, having the same fertiliser adoption pattern, and village fixed effects. Standard errors are corrected for clustering at the village level. P-values obtained with wild bootstrap-t procedure reported in parentheses. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%.

Table 3.11: Dyadic linear probability model: functional form 2

	(1)	(2)	(3)	(4)
risk aversion$_j$ - risk aversion$_i$	-.028 (.016)*	-.038 (.017)**		
(risk aversion$_j$ - risk aversion$_i$)^2	.007 (.008)	-.003 (.008)		
risk aversion$_j$ $<$ risk aversion$_i$			.047 (.166)	.061 (.072)*
risk aversion$_j$ $>$ risk aversion$_i$			-.002 (.943)	-.010 (.773)
Const.	.371 (.000)***	.591 (.064)*	.361 (.000)***	.588 (.012)**
Obs.	2166	2010	2166	2010
Cluster N	22	22	22	22
Risk aversion	No. safe choices	No. safe choices	No. safe choices	No. safe choices
Controls		✔		✔

Dyadic OLS regression. The dependent variable is a dummy which takes a value of one if farmer i would like to receive advice from farmer j. The regressions in column 2 and 4 include controls for difference in age, land owned, and number of animals owned between farmer j and farmer i. Further controls include dummies for having the same education level, same gender, having the same fertiliser adoption pattern, and village fixed effects. Standard errors are corrected for clustering at the village level. P-values obtained with wild bootstrap-t procedure reported in parentheses. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%.

Table 3.12: Dyadic linear probability model: other types of interaction

	i knows of j		i and j spoke		i lent money to j		i walked past j's farm	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ra$_j$ - ra$_i$	-.008 (.348)	-.013 (.164)	8.20e-19 (.305)	4.07e-19 (.388)	-.002 (.244)	-.003 (.116)	-.002 (.851)	-.003 (.813)
Const.	.753 (.000)***	.834 (.010)***	.186 (.000)***	.199 (.360)	.016 (.000)***	.021 (.340)	.112 (.000)***	.082 (.426)
Obs.	2166	2010	2166	2010	2166	2010	2166	2010
Cluster N	22	22	22	22	22	22	22	22
Risk aversion	No. safe choices		No. safe choices		No. safe choices		No. safe choices	
Controls		✔		✔		✔		✔

Dyadic OLS regression. The dependent variable is indicated in the column heading. The regressions in column 2, 4, 6 and 8 include controls for difference in age, land owned, and number of animals owned between farmer j and farmer i. Further controls include dummies for having the same education level, same gender, having the same fertiliser adoption pattern, and village fixed effects. Standard errors are corrected for clustering at the village level. P-values obtained with wild bootstrap-t procedure reported in parentheses. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%.

Table 3.13: Dyadic linear probability model: different measures of risk aversion

	(1)	(2)	(3)	(4)	(5)	(6)
risk aversion$_j$ - risk aversion$_i$	-.004	-.008	-.020	-.009	-.008	-.025
	(.380)	(.747)	(.452)	(.380)	(.781)	(.388)
Const.	.329	.329	.435	.341	.341	.454
	(.016)***	(.000)***	(.093)*	(.016)**	(.000)***	(.016)**
Obs.	1124	1124	1056	724	724	680
Cluster N	22	22	22	22	22	22
Risk aversion measure	No. safe choices	CRRA2	CRRA2	No. safe choices	CRRA1	CRRA1
Controls			✔			✔

Dyadic OLS regression. The dependent variable is a dummy which takes a value of one if farmer i would like to receive advice from farmer j. The regressions in columns 3 and 6 include controls for difference in age, land owned, and number of animals owned between farmer j and farmer i. Further controls include dummies for having the same education level, same gender, having the same fertiliser adoption pattern, and village fixed effects. Standard errors are corrected for clustering at the village level. P-values obtained with wild bootstrap-t procedure reported in parentheses. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%.

Table 3.14: Dyadic linear probability model: different measures of risk aversion

	(1)	(2)	(3)	(4)
(risk aversion$_j$ - risk aversion$_i$) * ij speak	-.040	-.023		
	(.316)	(.567)		
ij speak	.256	.256		
	(.000)***	(.000)***		
(risk aversion$_j$ - risk aversion$_i$)* j wants advice from i			-.033	-.019
			(.260)	(.192)
j wants advice from i			-.008	-.007
			(.949)	(.961)
risk aversion$_j$ - risk aversion$_i$		-.016		-.014
		(.531)		(.551)
Const.	.368	.368	.450	.445
	(.124)	(.146)	(.426)	(.489)
Obs.	1056	1056	1056	1056
Cluster N	22	22	22	22
Risk aversion measure	CRRA2	CRRA2	CRRA2	CRRA2
Controls	✔	✔	✔	✔

Dyadic OLS regression. The dependent variable is a dummy which takes a value of one if farmer i would like to receive advice from farmer j. All regressions include controls for difference in age, land owned, and number of animals owned between farmer j and farmer i. Further controls include dummies for having the same education level, same gender, having the same fertiliser adoption pattern, and village fixed effects. Standard errors are corrected for clustering at the village level. P-values obtained with wild bootstrap-t procedure reported in parentheses. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%.

Table 3.15: Dyadic linear probability model: different measures of risk aversion

	(1)	(2)	(3)	(4)
(risk aversion$_j$ - risk aversion$_i$) * ij speak	-.026 (.537)	-.001 (.813)		
ij speak	.268 (.006)***	.268 (.006)***		
(risk aversion$_j$ - risk aversion$_i$)* j wants advice from i			-.050 (.134)	-.039 (.014)**
j wants advice from i			.012 (.885)	.013 (.867)
risk aversion$_j$ - risk aversion$_i$		-.025 (.360)		-.011 (.667)
Const.	.385 (.064)*	.385 (.042)**	.467 (.238)	.463 (.236)
Obs.	680	680	680	680
Cluster N	22	22	22	22
Risk aversion measure	CRRA1	CRRA1	CRRA1	CRRA1
Controls	✔	✔	✔	✔

Dyadic OLS regression. The dependent variable is a dummy which takes a value of one if farmer i would like to receive advice from farmer j. All regressions include controls for difference in age, land owned, and number of animals owned between farmer j and farmer i. Further controls include dummies for having the same education level, same gender, having the same fertiliser adoption pattern, and village fixed effects. Standard errors are corrected for clustering at the village level. P-values obtained with wild bootstrap-t procedure reported in parentheses. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%.

Table 3.16: Dyadic linear probability model: different measures of risk aversion

	(1)	(2)	(3)	(4)	(5)	(6)
risk aversion$_j$ - risk aversion$_i$	.009 (.683)	.010 (.657)	-.002 (.961)	-.011 (.747)	-.009 (.779)	-.023 (.527)
No. farmers approached for advice$_j$		.011 (.531)	.009 (.569)		.026 (.519)	.002 (.863)
Const.	.401 (.006)***	.273 (.262)	.351 (.521)	.400 (.000)***	.105 (.703)	.419 (.471)
Obs.	466	466	452	320	320	310
Cluster N	10	10	10	10	10	10
Risk aversion measure	CRRA2	CRRA2	CRRA2	CRRA1	CRRA1	CRRA1
Controls			✔			✔

Dyadic OLS regression. The dependent variable is a dummy which takes a value of one if farmer i would like to receive advice from farmer j. The regressions in columns 3 and 6 include controls for difference in age, land owned, and number of animals owned between farmer j and farmer i. Further controls include dummies for having the same education level, same gender, having the same fertiliser adoption pattern, and village fixed effects. Standard errors are corrected for clustering at the village level. P-values obtained with wild bootstrap-t procedure reported in parentheses. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%.

Table 3.17: The six lottery pairs

	Risky lottery		Safe lottery	
Lottery Pair	Pr(prize=6 Cedi)	Pr(prize=0 Cedi)	Pr(prize=2 Cedi)	Pr(prize=1 Cedi)
1	0.8	0.2	0.8	0.2
2	0.7	0.3	0.7	0.3
3	0.6	0.4	0.6	0.4
4	0.5	0.5	0.5	0.5
5	0.4	0.6	0.4	0.6
6	0.3	0.7	0.3	0.7

Table 3.18: CRRA estimation

A subject who chooses ...	... has CRRA of
The risky lottery in all six pairs	$CRRA < 0.22$
The risky lottery in pairs 1 to 5 and the safe lottery in pair 6	$0.22 \leq CRRA < 0.38$
The risky lottery in pairs 1 to 4 and the safe lottery in pairs 5 and 6	$0.38 \leq CRRA < 0.51$
The risky lottery in pairs 1 to 3 and the safe lottery in pairs 4 to 6	$0.51 \leq CRRA < 0.62$
The risky lottery in pairs 1 to 2 and the safe lottery in pairs 3 to 6	$0.62 \leq CRRA < 0.72$
The risky lottery in pairs 1 and the safe lottery in pairss 2 to 6	$0.672 \leq CRRA < 0.82$
The safe lottery in all six pairs	$CRRA \geq 0.82$

4 Cooperation and expectations in networks: Evidence from a network public good experiment in rural India

4.1 Introduction

Well-connected individuals play an important role in social networks. They can effectively diffuse information, influence opinions, and arbiter transactions that require trust (DeMarzo et al., 2003; Golub and Jackson, 2010; Breza et al., 2014). When information is the product of costly experiments and its quality deteriorates with each relay in the network, socially central innovators have a large influence on group welfare (Bramoulle and Kranton, 2007). Leaders, who can motivate groups to increase public goods provision, are often highly connected (Bonacich, 1987; Grossman and Baldassarri, 2012; Jack and Recalde, 2013). Policy makers that want to promote human cooperation should pay attention to the behaviour of central individuals in networks.

Do well-connected individuals supply the public goods they are well-placed to generate? Consider a game where a star network determines who can benefit from the public good contributions of a particular individual (Bramoulle and Kranton, 2007). A large literature in experimental economics suggests that players are often prepared to contribute just as much as the other players contribute (Gächter, 2006; Chaudhuri,

This chapter is joint work with Marcel Fafchamps (Stanford University). We are first of all indebted to Gaurav Puntambekar and Kabira Namit, whose hard work made this project possible. Palash Puntambekar also contributed significantly to all field work activities. We would like to thank Johannes Abeler, Francesco Cecchi, Elwyn Davies, Edoardo Gallo, Kabira Namit, Esteban Ortiz Ospina, Cecilia Piantanida and seminar participants in Oxford for useful comments and suggestions. Our gratitude also goes to the the enumerators and supervisors whose contributions made this study possible. We finally would like to thank the study participants for their goodwill and patience. All mistakes remain naturally ours. This research was funded by the UK Department for International Development (DFID) as part of the iiG, a research programme to study how to improve institutions for pro-poor growth in Africa and South-Asia. The views expressed are not necessarily those of DFID.

2011). In this game, however, consideration of both relative efficiency and equality would motivate the player at the centre of the star to invest more than the average of the other players. This is because his contributions are most effective at increasing overall welfare, and because, as we explain later, payoff equality requires higher contributions from the centre player. Further, players may 'expect' the centre of the star to supply more than the rest. A recent literature hypothesises that individuals experience guilt when they determine a payoff for other players that is lower than what these players expect (Battigalli and Dufwenberg, 2007). Sufficiently guilt-averse players contribute to the public good as much as the other players expect them to contribute.

In this chapter we use an experiment with farmers in rural India to study the public good contributions of the central player in a star network. We answer three related questions. First, does the player at the centre of the star contribute more than the average of the other players? Second, is he influenced by the expectations that the other players hold about his decisions? Third, do players expect the centre of the star to contribute more than the rest?

In our experiment, as in standard public good games, individuals are given an initial endowment that must be divided between a private account and investment in a public good. The game is one-shot. Contributions to the public good increase aggregate payoff, while decreasing the payoff of the contributor. Following the model of Bramoulle and Kranton (2007), the star network determines who can benefit from the public good contributions of a particular player. The contributions of the centre of the star thus reach more individuals and have a larger impact on the total welfare of the group than the contributions of the spoke players. Furthermore, the centre of the star benefits from the contributions of all the spokes, while each spoke is only affected by the contribution of the centre. When every player contributes the same positive amount, the centre player earns a higher payoff than the spokes. Higher contributions from the centre are required to equalise payoffs.[1]

Positions in the network are randomly assigned. This allows us to use the strategy method in two ways. First, to elicit contribution decisions for both the case where the player is assigned to the centre position and the case where the player is assigned to the spoke position. Second, to allow the centre of the star to contribute different amounts depending on the average contribution of the spokes. We also collect players' expectations about how much other players will contribute when they are deciding as centre of the star.[2]

We analyse the decisions that players would like to take when placed at the centre of the star in the experiment, regardless of their position in real networks. This could affect the external validity our results if individuals who are well-connected in real networks have other-regarding preferences that are systematically different from those of other individuals.[3] Reassuringly, using data on the real connections between participants, we do not find evidence suggesting that individuals who are more central in real networks behave differently in the experiment.

In selected sessions we disclose the average value of the expectations of the players in the network.[4] This captures what farmers in the experimental session, on average,

[1]This type of heterogeneity has some similarities with heterogeneity in the rate of return to contributions, which is explored, for example, in the experiments of Reuben and Riedl (2009) and Nikiforakis et al. (2012). In particular, both the centre of the star and a player with a high rate of return benefit more than others when everybody is contributing the same positive amount. Further, both types of players find increasing aggregate payoff less costly than the other players in their sessions. The crucial difference is that in our design these effects are obtained by changing the number of players with which an individual interacts, while in the design of Reuben and Riedl (2009) each player interacts with everybody else, but the benefits and costs of this interaction vary. For example, in our set-up the centre of the star is more efficient than the spokes at increasing aggregate welfare because *a higher number of players benefit from his contributions*. Players with a higher rate of return in Reuben and Riedl (2009), on the other hand, are more efficient at increasing aggregate welfare because their sacrifice of personal payoff has *a greater effect on each of the other players*.

[2]The centre of the star has to specify a contribution level for each of four possible (rounded) average contribution levels of the spokes. We thus elicit expectations about the average contribution in each of the four decisions.

[3]Burt et al. (1998), for example, document a correlation between personality traits and network position.

[4]Throughout the chapter, we will refer to these as 'group expectations'. We disclose the average value of each of the four forecasts that players make. See footnote 2.

expect the centre of the star to contribute. Sufficiently guilt-averse individuals will contribute as much as the group expects them to contribute. They will do so more frequently as the monetary cost of contribution is decreased. We thus randomly vary the monetary cost of contribution across sessions.

We play the experiment with adult male farmers, randomly selected from villages in the Indian state of Maharashtra. Information and innovation networks are salient for farmers, making this an interesting population for our research question. Further, the policy implications of our findings are particularly relevant in this context: agricultural interventions, for example extension services, often try to mobilise central farmers in networks (Kondylis et al., 2014).

Our first finding is that, despite the efficiency and equality considerations presented above, the centre of the star contributes an amount that approximates the average contribution of the spokes. This is what the literature calls 'conditional cooperation', a strategy that is often observed in public good games played by homogeneous groups (Fischbacher et al., 2001). Our adaptation of the strategy method to a network setting enables us to observe this pattern. Both regression analysis and the relative frequency of the strategies chosen by players support the finding. Conditional cooperation has implications for efficiency: farmers in the baseline treatment are able to capture only about 50 percent of the potential gains from cooperation. If the centre of the star contributed the full amount regardless of the average of the spokes, they would capture 82 percent of the available gains.

Second, we disclose the average expectations of the players in the network and find a match between contributions and disclosed values in 42 percent of decisions. Matches become more likely when we lower the *monetary cost* of contributions. For high-understanding players,[5] the frequency of matches between contributions and group expectations increases significantly to 53 percent. This confirms the predic-

[5]These are players who make at most 2 mistakes in the initial understanding questions.

tion of the model of guilt aversion and suggests that individuals are influenced by the expectations of the other players.

In addition, the effect is moderated by the average degree in the real-life network that connects individuals in a session. Farmers in sessions characterised by more connected real-life networks match their contributions to the expectations of the group more frequently.

Third, farmers on average expect the centre of the star to be a conditional cooperator. In sessions where we do not disclose expectations, matches between contributions and group expectations are 9 percentage points less likely. The level of contributions, however, is not affected by disclosure of group expectations.

The results hold two main implications for policy. Interventions that require central individuals to take costly actions for the benefits of others, for example 'model farmer' extension programmes (BenYishay and Mobarak, 2014; Kondylis et al., 2014), should not rely on the intrinsic motivation of central actors alone. Moreover, players in central positions care about the expectations of others, but expectations reflect current levels of pro-sociality. Institutions that enable the disclosure of expectations, such as village or cooperative meetings, will not be effective unless expectations are independently manipulated.

Our study contributes to several strands of the literature. First, it impacts on the experimental and theoretical literature on public good games played over networks. The games that have been proposed differ on at least three dimensions: whether the network determines the reach of contributions or the observability of players' actions; whether the payoff function implies an interior optimum or a corner solution; whether play is one-shot or repeated. Bramoulle and Kranton (2007) study equilibria and welfare in a game where the network determines the reach of contributions, the optimum is interior and players have no social preferences. Their seminal analysis highlights the potential for specialisation in networks. Rosenkranz and Weitzel (2012) play a repeated version of their game and document that coordination on theoretical equilibria is infrequent and unstable.[6] On the other hand, for a game of strategic complements, Charness et al. (2014) document behaviour that closely follows the theoretical predictions. Other studies find that when links determine observability, the structure of the network influences the level of cooperation (Fatas et al., 2010; Carpenter et al., 2012).[7]

We contribute to this literature by offering a design that is particularly amenable to a study of other-regarding preferences. The strategy method, which to our knowledge we apply for the first time to a public good game played over a network, removes uncertainty about the distributional consequences of actions and the history of play. The payoff function determines a corner solution at zero for rational selfish players, making deviations from selfish best response transparent and easy to analyse. The network structure creates salient asymmetries across network positions regarding the effects of contributions on the welfare of other players.

Since the widely cited study of Fischbacher et al. (2001), the strategy method has often been employed in public good games played by homogeneous groups. As we mentioned above, a widely reproduced finding is that a large fraction of players are 'conditional cooperators' (Gächter, 2006; Chaudhuri, 2011). These are defined as players whose 'contributions to the public good are positively correlated either with their ex ante beliefs about the contributions to be made by their peers or to the actual contributions made by the same' (Chaudhuri, 2011, p.56). In our study, we show that conditional cooperation is followed by the centre player of a highly asymmetric star network.[8] This extends our understanding of the settings where this behaviour occurs

[6]Only 2.4 percent of decisions are part of a theoretical equilibrium, and episodes of convergence to an equilibrium occur in 27 periods over 3360.

[7]There is also a small literature that studies prisoner dilemma games over networks, which is discussed in Kosfeld (2003).

[8]For a public goods game played over a network, Fatas et al. (2010) report a positive correlation between current contributions and the contributions of peers in previous rounds. They interpret this correlation as suggesting conditional cooperation, as peer past contributions are probably a major determinant of a player's

and highlights the importance of incorporating other-regarding motives in existing models of public good provision over networks. Bourlès et al. (2013) offer a recent theoretical analysis of altruism in networks.

We also contribute to the literature on guilt aversion. As explained above, guilt-averse players dislike to play strategies that generate a payoff for the other players that is lower than the payoff these players expect. A formal definition of guilt aversion is given by Battigalli and Dufwenberg (2007). Empirical evidence supporting this model of utility is provided in the trust games played by Dufwenberg and Gneezy (2000), Charness and Dufwenberg (2006), Bacharach et al. (2007) and Reuben et al. (2009), and in the public good game played by Dufwenberg et al. (2011). Bellemare et al. (2010) estimate that Dutch individuals drawn from a representative sample are 'willing to pay between 0.40 and 0.80 Euro to avoid letting down proposers by 1 Euro'. On the other hand, Ellingsen et al. (2010) are unable to find evidence in support of guilt aversion among Swedish students, while the results of Vanberg (2008) suggest that promise keeping is not driven by guilt aversion, as previously hypothesised in the literature.

High-degree players in networks, whose actions affect the payoffs of a large number of individuals, may be particularly concerned with guilt. Theory and evidence in social psychology suggests that guilt aversion has its roots in the fear of exclusion from a reference group (Baumeister et al., 1994). Such fear is likely to heighten as the number of disappointed individuals increases.[9]

We show that matches between contributions and group expectations are frequent, and become more common when the cost of contribution is decreased, as predicted by the model of guilt aversion. This suggests that expectations, besides distributional preferences, influence strategic behaviour in networks.

---

expectations about current contributions.

[9] It may also be, however, that central individuals are more difficult to exclude. In this case, the net effect on the fear of exclusion would be ambiguous.

Finally, our work is related to the literature on the elicitation of expectations. Economists often ask individuals to report forecasts of uncertain events (Manski, 2004). The experimental literature has imported these techniques to study expectations in strategic settings. Several studies elicit individual expectations about the *strategies* chosen by other players, while a small number of studies explore individual expectations about the *expectations* held by other players (Manski and Neri, 2013).[10] In recent years, expectation elicitation techniques have also been used successfully with populations in developing countries (Delavande et al., 2011). This strand of work has focused the attention to non-strategic environments: for example, the returns to schooling, the benefits of new technologies, the prices of agricultural products. On the other hand, evidence from developing economies about the expectations of subjects in strategic settings is scarce. An exception is Caria and Falco (2014), who report that employers in urban Accra have inaccurately pessimistic priors about the trustworthiness of a sample of employees. On the contrary, in this study we find that, on average, farmers correctly expect the centre of the star to play conditional cooperation. We contribute to this literature by showing that, within our novel strategic setting and a population of farmers who interact with each other with high frequency, expectations are not systematically biased.

In the next section we present the design of the experiment. Section 4.3 develops a number of predictions and discusses how we will formulate the related statistical tests. Section 4.44 outlines the data we use and the basic descriptives. The results are described and discussed in section 4.5. We present the concluding remarks in the final section of the chapter.

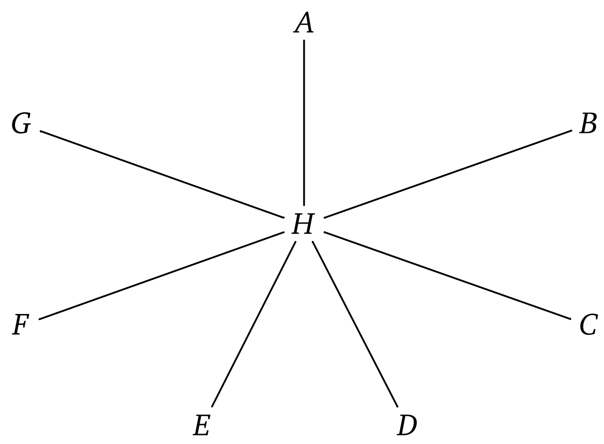
[10]The literature in behavioural game theory calls expectations about the strategies of other players 'first-order expectations'. 'Second-order expectations', on the other hand, are expectations about the first order expectations of the other players.

4.2 Design

We play a public good game between players connected by a network that determines who benefits from the public good contribution of a particular player.[11]

In each session we recruit eight participants. These will eventually be arranged over a star network like the one represented in picture 4.1 below. Links in this network cannot be changed and are undirected: if player A is linked to player H, then player H is linked to player A. There are two types of players: one centre and seven spokes. The centre benefits from the public good contributions of the seven spokes. Further, his own contribution reaches each of the spokes. A spoke, on the other hand, only receives the contribution of the centre and only reaches the centre with his own contribution. The position of each farmer in the network is randomly assigned *after* all contribution decisions have been made.

Figure 4.1: The star network



[11]The instructions for one of the treatments are included in the appendix. The remaining experimental materials can be found here: https://sites.google.com/site/stefanoacaria/pgindia

Contributions

Each player is endowed with three notes worth 50 INR each and has to decide how many notes to contribute for the provision of the public good.[12] As positions in the network are assigned *after* the contributions decisions are made, we ask players to specify in advance how much they would like to contribute if they will be assigned to (i) the spoke position and (ii) the centre position. Decision (i)- we call this the 'spoke contribution'- is an unconditional contribution decision. s_i indicates the 'spoke contribution' of player i. On the other hand, decision (ii)- we call this the 'centre contribution'- is conditional on the average of the spokes. There are 4 possible (rounded) average contribution levels of the spokes: 0,1,2 and 3 notes. We use the letter z to refer to the average contribution of the spokes: $z \in \{0, 1, 2, 3\}$. For each possible average contribution level z, player i has to declare how much he would like to contribute if he is assigned to the centre position *and* the seven spokes have contributed on average z. The conditional contribution decision of player i when spoke average contribution is z is called c_i^z. The vector $c_i = \left(c_i^0, c_i^1, c_i^2, c_i^3\right)$ collects the four conditional decisions of player i. We call c_i^z a contribution 'decision' and c_i a contribution 'profile'.

After positions are assigned, the enumerator calculates the (rounded) average of s_i for the seven players assigned to the spoke position. Given this average, the enumerator selects the right element from the c_i vector of the player assigned to the centre position. Let x_i indicate the actual number of notes that player i contributes to the public good: $x_i = s_i$ if player i is a spoke and $x_i = c_i^z$ if player i is the centre and the average contribution of the seven spokes is z.

[12]Players can contribute zero, one, two or three notes. Fractions of a note are not allowed. The value of the endowment- 150 INR- correspond to 7.75 USD, using an exchange rate of 0.0155 USD for one INR, and then a PPP conversion factor of 10/3. These are same same conversion values reported for the link formation experiment of chapter 1. The size of the endowment is comparable to a daily wage offered in a state employment program and is in line with those of similar experiments. For example, Breza et al. (2014) report a mean payout of about 110 INR for an experiment with Indian farmers in Karnataka. The minimum a farmer can earn in our experiment is 90 INR, the maximum 430 INR.

Using notation from Goyal (2007), we define N as the set of players in a session, and N_i^d as the subset of these players that are linked to player i. The payoff of player i at the end of the game is given by:

$$\pi_i = 50(3 - x_i) + r50\left(\sum_{j \in N_i^d} x_j + x_i\right) \quad r = \{3/5, 4/5\} \tag{4.1}$$

The rate of return r to investing in the public good can take a low (3/5) or a high (4/5) value. Experimental sessions are randomly allocated to a high or a low value of r.

Three features of this design are worth noting. First, the payoff function (4.1) resembles closely the standard payoff function of public good experiments (Camerer, 2003; Chaudhuri, 2011). The only difference is that we sum over the contributions of the direct connections N_i^d and not over the contributions of all players N. The main strategic features of a public good game are otherwise preserved: $r < 1$ and hence contributing a positive amount is a dominated strategy. Further, when player i increases his contribution by 1 note he forgoes 50 INR in private payoff, but generates a sum of individual payoffs from the public good equal to $r50(N_i^d + 1)$ INR. For all values of N_i^d in the star network, $r50(N_i^d + 1) > 50$. This implies that aggregate payoff monotonically increases in x_i and is maximised when everybody contributes the whole endowment.

Second, the impact of a note contributed by player i on the welfare of the other players- $r50N_i^d$- is proportional to the number of connections player i has. A note contributed by a spoke player increases the welfare of the other players by $r50$. A note contributed by a centre player has an impact of $r350$. The centre player is seven times more efficient than the spoke player at generating payoff for the other players. This is a very high difference in efficiency.[13]

[13]Increasing the payoff of the other players is very cheap for the centre player. When $r = 4/5$, an additional note contributed by the centre player increases the payoff each spoke by 40 INR (i.e. it increases the total

Third, the design relies twice on the strategy method. In the first instance, it allows players to specify a contribution decision for the case in which they are assigned to the spoke position and a second contribution decision for the case in which they are assigned to the centre position. Second, for the latter decision players are allowed to condition their contribution on the average contribution of the spokes. The strategy method has been employed frequently in public good games (Fischbacher et al., 2001; Brandts and Charness, 2011; Fischbacher et al., 2012). It has been shown to produce qualitatively similar results to those observed using direct elicitation methods (Fischbacher et al., 2012) and all evidence so far shows that the choice between direct elicitation and strategy method does not influence whether a treatment effect is found or not (Brandts and Charness, 2011).

Expectations

Farmers have expectations about what 'centre contribution' decisions c_i the other farmers will take. In all treatments, after the 'spoke contribution' decision s_i is taken, but before 'centre contribution' decisions c_i, we carry out two activities, which allow us to elicit expectations and, in some sessions, to manipulate them.

First, we distribute a closed envelope to each player containing a message. In each session, there are 2 messages. Messages are randomly assigned and four players get each message. Players know the distribution of messages, but only see the content of the message that is assigned to them.

Some messages prime players to increase or decrease their expectations about what c decisions other farmers will take. The message that primes players to increases their expectation reads as follows: 'Here is some information to help you with the

payoff of the seven spokes by 280 INR), while decreasing the centre's own payoff by 10 INR. This ratio is even more favourable than the ratio of the 'Barc2' and 'Berk17' games played by Charness and Rabin (2002), where the player has to sacrifice 15 units of payoff in order to generate 350 units of payoffs for the other player. In the 'Barc2' and 'Berk17' games, about 50 percent of dictators choose to pay 15 units of payoff to increase the payoff of their experimental partner.

expectation questions. Many farmers in your district have contributed 3 notes in every decision'. The message that primes players to decrease their expectation is identical, but replaces 3 notes with 0 notes: 'Here is some information to help you with the expectation questions. Many farmers in your district have contributed 0 notes in every decision'.[14]

The other messages are neutral. The first neutral messages is: 'Thank you for taking part in this experiment'. The second: 'We would like to thank your village for hosting this experiment'. Details about the distribution of messages across sessions are given in the next sub-section which describes the treatments.

Second, each individual i is asked to guess what the average of c_j^z among the other 7 players will be, for each of the four possible values of z.[15] We call this (point) expectation: α_i^z. More precisely: $\alpha_i^z = E_i\left(\sum_{j \in N \setminus i} \frac{c_j^z}{7}\right)$.[16] For example, α_i^2 records how much player i expects the other 7 players to contribute *if* they are assigned to the centre position and the spoke average is 2. The vector $\alpha_i = \left(\alpha_i^0, \alpha_i^1, \alpha_i^2, \alpha_i^3\right)$ collects the four expectations of player i. We call α_i an expectation 'profile'.

Finally, $\bar{\alpha}^z$ is the average of α_i^z over all eight players in N. In other words, $\bar{\alpha}^z$ indicates what is the contribution that individuals in the network, on average, expect from a player at the centre of the star when the spoke players have contributed an

[14]We do not quantify what we mean by 'many'. Non-trivial proportions of farmers indeed play either of these two strategies in the pilot. Hence this does not constitute a instance of lying, which is generally not allowed in economic experiments. A further concern is that the messages we distribute cannot be used by a Bayesian player to update his priors, as they do no constitute well specified signals with a known precision. While recognising this, we emphasise that our aim is to shock the beliefs of the subset of players that receive the priming messages. We conjecture that a signal that is mathematically precise may in fact include too much information to be effective for real subjects. We hence rely on the simplest message, in the hope that it will have the best chance of modifying expectations. As it will become apparent in the results section, however, the message fails to change expectations systematically.

[15]All instructions are double-translated. We are careful to ensure participants understand that we refer to expectations in the sense of 'forecasts', and not of 'demands'. For each average contribution of the spoke z, we ask: 'On average, how many notes will the other players put in the common pot when they play as player *H* and players *A* to *G* have put on average z notes in the common pot?'. The 'common pot' is a physical holder where players have to put the notes that they would like to contribute to the public good.

[16]$N \setminus i$ indicates all individuals in N excluding player i.

average of z notes. We refer to $\bar{\alpha}^z$ as the 'average group expectation', or sometimes simply as the 'group expectation'.

In selected treatments, after eliciting α_i from each player, we disclose $\bar{\alpha}$ publicly on a white board.[17] Subjects are not informed that the average of the expectations they report will later be disclosed to the group. This feature is important, as it rules out the possibility that farmers misreport their expectations in order to influence the behaviour of the other players.[18] It also ensures that, before the disclosure of $\bar{\alpha}$, the experimental protocol is identical across treatments.

Figure 4.2 summarises the order of activities during the experiment. First, decisions s_i are taken for the case where player i will be assigned to the spoke position. Second, messages are distributed. Third, expectations α_i are elicited. Then, in selected sessions, the average of expectations α_i is disclosed publicly. Finally, players take decisions c_i, for the case where they are assigned to the centre position.

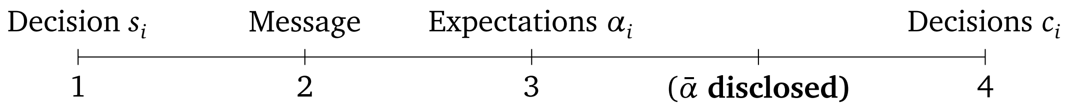
Before play, participants play a trial round of the game, which features steps 1 and 4 in figure 4.2, but does not include messages nor expectation elicitation. At the end of the trial round, the enumerator calculates the payoff that would accrue to participants given their decisions and a random draw that assigns positions in the network. This exercise reinforces participants' understanding of the game.

We choose not to incentivise the elicitation of expectations for a number of reasons. First, to keep the design simple. This is a priority given the difficulties involved in ensuring understanding of the strategy method and the expectation questions. Second,

[17]This design feature makes the average expectation of the group clear and salient. However, it could be objected that disclosing any number could influence behaviour because of anchoring effects. To ensure differences in behaviour between treatments are not driven by anchoring effects, we could have reported four random numbers on the board in the treatment where we do not disclose average expectations. We do not include this feature to avoid confusing farmers and enumerators on the purpose of the random numbers.

[18]The literature in economics has recently started analysing the strategic implications of expectation formation when interacting partners are guilt averse. The strategy that manipulates expectation to produce desired outcomes is referred to as 'guilt induction' (Cardella, 2012).

Figure 4.2: Order of activities in the experiment



because, when we disclose $\bar{\alpha}$, other-regarding farmers may align their c_i decisions to the average belief $\bar{\alpha}$ in order to ensure that the other players are awarded the expectation incentive.[19] Third, to avoid hedging strategies. For example, a player may declare to have low expectations so that he is awarded the expectation incentive in states of the world where the payoff from the centre player contribution is low.

The literature on expectations elicitations is not conclusive on the issue of incentives. Delavande et al. (2011) summarise a number of studies in development economics which elicit expectations without using monetary incentives. Gächter and Renner (2010) find that incentives reduce the dispersion of beliefs but do not change the central tendency of the distribution. In our study dispersion is not a concern as expectations can take only 4 values. Schlag et al. (2014) provide a recent review of the various methods to incentivise beliefs and the respective strengths and weaknesses.

Treatments

We have four treatments. In the baseline treatment T0 all players receive a neutral message and $\bar{\alpha}$ is not disclosed. In the first treatment T1*neutral* we disclose the true $\bar{\alpha}$ to participants, while still distributing a neutral message to each participant. In the last two treatments, we use the messages to manipulate expectations. In T1*positive* four players are given the positive priming message in order to produce an *upward* shock to their beliefs α_i, and four players are given neutral message number 2. In T1*negative* four players are given the negative priming message in order to produce a *downward* shock to α_i, and four players are given neutral message number 2.

[19] This requires (a) the player to be pivotal in determining the average of c_i and (2) that the average value of the expectation has been chosen by at least some players, which has to be true if the average is 0 and 3, but it is not necessarily the case if the average is 1 or 2.

We also manipulate the parameter r, which determines the 'cost' of contributing to the public good. Half of the sessions of each treatment are played with $r = 3/4$ and half of the sessions are played with $r = 4/5$. Table 4.1 summarises.

Table 4.1: Summary of Treatments

	T0	**T1*neutral***	**T1*positive***	**T1*negative***
Disclose $\bar{\alpha}^z$	–	✓	✓	✓
Message 1	neutral 1	neutral 1	positive	negative
Message 2	neutral 2	neutral 2	neutral 2	neutral 2

Throughout the analysis we will repeatedly perform comparisons between individuals who have been randomly assigned to receive message 'neutral 2' across treatments. Up to the point where expectations are elicited, these individuals are exposed to the same protocol irrespective of treatment. They have read the same message. They are equally uncertain about the message that the other four players have received. They do not anticipate that, in T1 sessions, the average of the expectations will be disclosed. Experimental manipulation is limited to the phase where $\bar{\alpha}$ is disclosed.

4.3 Predictions

We study the decisions that players take for the case when they are at the centre of the star.

Play in T0

The experimental literature has repeatedly found that conditional cooperation is the modal strategy in public good games played with the strategy method by homogenous groups (Chaudhuri, 2011). A conditional cooperator is somebody whose contribution correlates with the average contribution of the group, sometimes with a small self-serving bias. We define profiles that are strictly increasing in the average of the spokes as corresponding to 'strict conditional cooperation'. $c_i = (0, 1, 2, 3)$ is the only possible strictly increasing profile in our game. Strategies that are weakly increasing in the average contribution of the spokes and are not flat nor strictly increase, for example $c_i = (0, 0, 1, 2)$ or $c_i = (0, 3, 3, 3)$, are referred to as 'weak conditional cooperation'. Under this definition, a weak conditional cooperator in the centre of the star can be somebody who contributes (weakly) more than the spoke average in every decision, somebody who always contributes (weakly) less than the spoke average in every decision, or neither.

Strict conditional cooperation can be the result of an independent social preference, or, in standard public good games, may derive from an underlying preferences for equality of payoffs or for reciprocity. In our game, however, the centre of the star may have a number of reasons to contribute above the level of strict conditional cooperation.

First, contributions by the centre of the star *reach* more players and have a much higher effect on aggregate payoff than contributions by the spokes. For same cost in individual payoff terms, contributions by the centre of star generate an effect on the payoff of the other players that is seven times larger than that generated by the contributions of a spoke. Individuals may consider that a player who is more effective than others at increasing group welfare should contribute more to the public good. If so, the centre of the star will choose a profile that has a higher intercept than $c_i = (0, 1, 2, 3)$, a steeper slope, or both. For example, motivated by his high relative efficiency, the centre of star may decide to contribute proportionally more than what the spokes contribute. This would result in a profile with a steeper slope. Alternatively, he may decide to exceed conditional cooperation by a fixed absolute amount, for example, the average of the spokes plus one. This would raise the intercept of the profile.

Second, when all the spokes are contributing the same positive amount, higher contributions by the centre of the star unambiguously reduce inequality in payoff among players. When all the spokes are contributing zero, on the other hand, positive contributions by the centre of the star worsen inequality. Inequality averse players dislike payoff differences of both types (Fehr and Schmidt, 1999).[20] Under the linear specification proposed by Fehr and Schmidt (1999), reducing inequality gives constant marginal utility. The inequality-averse player may find reducing inequality too costly and so contributes nothing. Otherwise, if the utility gain from lower inequality is higher than the utility loss from a unit of public good contribution, he is prepared to contribute all of his endowment to minimise inequality. A sufficiently inequality-averse player would thus choose the profile $c_i = (0, 3, 3, 3)$.

Third, other players may expect the centre of the star to contribute a higher amount than everybody else, based on the considerations of relative efficiency and equality that we presented above. This can create a certain 'social pressure' on the central player, which is captured by the model of guilt aversion which we present below. The exact profile that follows in this case depends on the shape of the profile expected by the group and hence cannot be determined a priori.

Treatment effects

We hypothesise that farmers are guilt-averse and that this is an important determinant of the behaviour of the farmer at the centre of the star network. The model of guilt aversion makes a number of specific predictions on how individuals will respond to our treatments. In this subsection, we present the predictions. In the next subsection, we will discuss how we use the experimental data to perform the related statistical tests.

Guilt-averse players dislike to 'let other players down'. More precisely, they dislike to play strategies which determine a lower payoff for other players than what these players expect to get. In our game, for example, farmers may expect the centre of the star to contribute generously to the public good. As a consequence, they may expect to earn a high payoff even when they are assigned to the spoke position. In this scenario,

[20] For simplicity, we assume that when the average contribution of the spokes is x, the player at the centre of the star assumes that each spoke has contributed exactly x.

a guilt-averse centre of the star will feel he is letting the other players down if his contribution does not match the high expectations of these players.

In order to quantify the extent to which his actions 'let other players down', a guilt-averse player has to form an expectation about the payoff other players expect to get. We define β_i as player i's belief about what contribution, on average, other players expect him to make if he is assigned to the centre position: $\beta_i^z = E_i(\bar{\alpha}^z)$. In the language of psychological games, β_i is a 'second-order belief': a belief about the beliefs of other players.

A guilt-averse player at centre of the star maximises the following utility function:

$$u_i(c_i^z, \beta_i^z|z) = \pi_i(c_i^z|z) - g \max\left(\bar{\pi}_j(\beta_i^z|z) - \bar{\pi}_j\left(c_i^z|z\right), 0\right) \tag{4.2}$$

where z, as usual, indicates the average contribution of the spokes. The first element in utility function (4.2) reflects a concern for monetary payoffs. The second element is a utility penalty for contribution choices that determine an average payoff for the spoke players that is $\bar{\pi}_j(\beta_i^z|z) - \bar{\pi}_j(c_i^z|z)$ units lower than what the centre player thinks the spoke players expect.[21] For simplicity, we assume g is linear and β_i^z is a point belief. We assume that the guilt-averse player reacts to the the average value of the expectation distribution among the other players. It is conceivable that other moments of the distribution are also salient and flag this as an area for further research.

Suppose $c_i^z < \beta_i^z$. In this case, player i thinks that the other players are earning a lower payoff than the payoff they expect to get- he feels *guilty* about this. Contributing one more note decreases guilt by $gr50$, while decreasing the monetary payoff by $(1-r)50$. When $g > \frac{1-r}{r}$, the reduction in guilt outweighs the loss of monetary payoff and the centre of the star finds it optimal to contribute what he thinks other players expect

[21]Note that the value of the difference between $\bar{\pi}_j(\beta_i^z|z)$ and $\bar{\pi}_j(c_i^z|z)$ does not depend on z. Thus player i does not need to form a belief about z.

him to contribute: $c_i^{z*} = \beta_i^z$.[22]

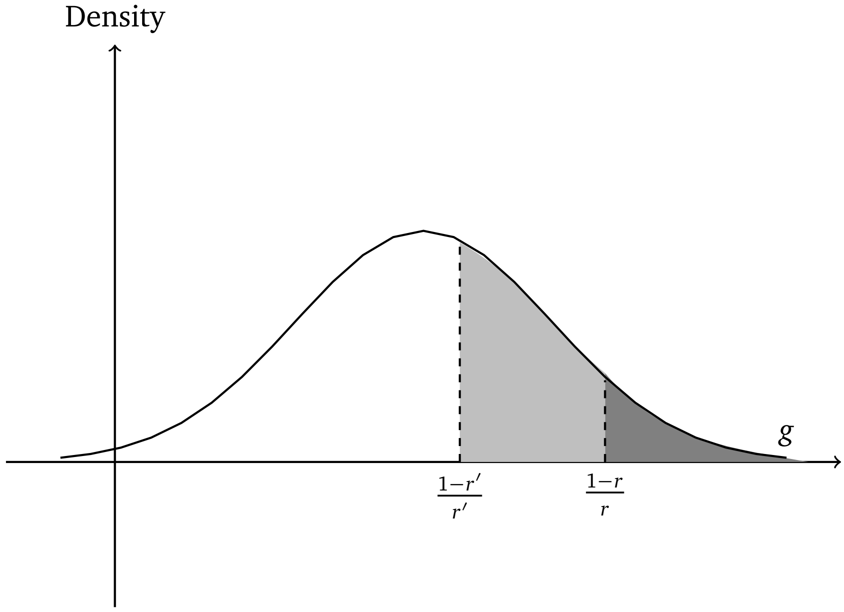
In treatments T1*neutral*, T1*positive*, and T1*negative*, we disclose the true value of $\bar{\alpha}^z$, for each value of z. Thus, when players take the 'centre contribution' decisions c_i^z, the belief β_i^z has been updated to reflect the true $\bar{\alpha}^z$. In these treatments, for sufficiently guilt-averse players with $g > \frac{1-r}{r}$, $c_i^{z*} = \bar{\alpha}^z$ for all z. When players set $c_i^z = \bar{\alpha}^z$, we say that there is a 'match' between contributions and group expectations.

Variation in r across sessions allows us to formulate the first testable prediction of the model of guilt aversion. The parameter r determines the monetary cost of contributing one more note to the public good. The higher r, the lower the cost of increasing contributions, and of reducing guilt when positive contributions are expected. Given a non-degenerate distribution of g in the population, as r gets higher more people will match their contribution to the disclosed $\bar{\alpha}^z$. Figure 4.3 illustrates. In the figure, we assume g is normally distributed in the population. Integration from $\frac{1-r}{r}$ to infinity gives the fraction of players who set $c_i^{z*} = \bar{\alpha}^z$ when the rate of return is r. This is represented by the dark grey area in the figure. Suppose now we switch to a different rate of return $r' > r$. Notice $\frac{1-r}{r} > \frac{1-r'}{r'}$. The fraction of players who set $c_i^{z*} = \bar{\alpha}^z$ is now given by the sum of the light and dark grey areas and is larger. This shows that matches between contributions and group expectations will be more frequent in sessions randomly assigned to a high level of r:

Prediction 4.1. *Players who receive message neutral 2 in a T1 session assigned to* $r = 4/5$ *are more likely to choose contributions* c_i^z *that are equal to* $\bar{\alpha}^z$ *than players who receive the same message in a T1 session assigned to* $r = 3/5$.

[22]Contributions above β_i^z, on the other hand, are always dominated by contributions matching β_i^z, as they determine an additional reduction in monetary payoff and no further reduction in guilt.

Figure 4.3: An increase in the rate of return to public good contributions



In treatments T1*positive* and T1*negative* we introduce random shocks to the level of group expectations. If the positive message in T1*positive* succeeds in raising α_i for the four players who receive the priming message, $\bar{\alpha}$ will be higher in T1*positive* sessions than in T1*neutral* sessions. Players who receive message neutral 2 now have to contribute higher amounts to minimise their guilt. As long as $g > \frac{1-r}{r}$ for at least some players who receive message neutral 2, contributions will be higher in T1*positive* than in T1*neutral*. A symmetric argument applies to T1*negative*. We hence make the following prediction:

Prediction 4.2. *The contributions c_i^z of players who receive message neutral 2 in treatment T1*positive *and T1*negative *are, respectively, higher and lower than those of players who receive message neutral 2 in treatment T1*neutral.

This increase (decrease) will be proportional to the difference between the average $\bar{\alpha}$ disclosed in T1*neutral* sessions and the average $\bar{\alpha}$ disclosed in T1*positive* (T1*negative*) sessions. This implies that if our experimental manipulation fails to affect average expectations no treatment effects will be found.

Finally, under the assumption that $g > \frac{1-r}{r}$ for at least some players, we can learn whether individuals hold correct β_i^z expectations by comparing decisions in T1*neutral* and in T0. If baseline β_i^z expectations are frequently inaccurate, guilt-averse individuals in T0 will often fail to match their contributions to the true $\bar{\alpha}$. Disclosure of $\bar{\alpha}$ in T1*neutral* will then increase the frequency of $c_i^z = \bar{\alpha}^z$ matches.[23] Some inaccuracy in expectations is likely, and we hence predict that the frequency of matches will be higher in T1*neutral*.

If match frequency increases, the effect on average contributions is ambiguous and depends on whether guilt-averse players with inaccurate priors revise these priors upwards or downwards, after disclosure of $\bar{\alpha}^z$.[24]

Prediction 4.3. *Players in T1*neutral *are more likely to choose contributions c_i^z that are equal to $\bar{\alpha}^z$ than players in T0.*

Analysis

We analyse contribution and expectation profiles in two ways. First, we study the average intercept and slope of contribution and expectation profiles with regression analysis. We pool the four decisions or expectations of each player and create a small panel with four observations per player. Suppose a profile takes the following linear form:

[23]This argument rests on the assumption that farmers are certain about the value of $\bar{\alpha}^z$- β_i^z is a point belief and not a distribution- and that g is linear. If farmers are uncertain about the beliefs of their peers and g is concave, disclosure of $\bar{\alpha}$ can increase contributions even when the mean of the distribution of β^z is correct.

[24]A simple example illustrates. Suppose there are three types of players, each occurring in the population with equal frequency: (i) guilt-indifferent and selfish, (ii) guilt-averse with accurate priors $\beta_i = \bar{\alpha}$, (iii) guilt-averse with inaccurate priors $\beta_i = p$. In T0, the average level of contributions when spoke average is z will be: $(\frac{1}{3} * 0) + (\frac{1}{3} * \bar{\alpha^z}) + (\frac{1}{3} * p^z)$. In T1*neutral*, players with inaccurate priors revise these and now contribute $\bar{\alpha}^z$. The new average will be: $(\frac{1}{3} * 0) + (\frac{2}{3} * \bar{\alpha^z})$. The difference between average contributions in T1*neutral* and in T0 is given by: $\frac{1}{3}(\bar{\alpha^z} - p^z)$, which is positive when $\bar{\alpha^z} > p^z$. Contributions are higher in T1*neutral* only when guilt-averse players with inaccurate priors underestimate the expectations of others.

$$x^*_{iz} = \kappa + \beta_1 z + u_{iz} \tag{4.3}$$

where x^*_{iz} can be either the contribution decision c^z_i, or the expectation α^z_i. The intercept κ measures the level of x^* when spoke average contribution z is 0, while β_1 captures the increase in x^* when z increases by one unit. Under strict conditional cooperation $\kappa = 0$ and $\beta_1 = 1$. Other values of κ and β_1 are also possible. However, participants are endowed with only three notes in the game. Hence, what we observe is:

$$x_{iz} = \min\left(\max(0, x^*_{iz}), 3\right) \tag{4.4}$$

In our data corner solutions at both 0 and 3 occur frequently. We hence estimate the values of κ and β_1 using a tobit model with a lower limit at 0 and an upper limit at 3. We then provide two-sided Wald tests of the hypotheses $\kappa = 0$ and $\beta_1 = 1$ and study the direction of any deviation. To separately analyse the intercept and slopes for the T0 treatment, we introduce a dummy for being in a T1 treatment and an interaction term capturing any additional effect of z in T1 sessions:

$$x^*_{iz} = \kappa + \beta_1 z + T1 + \beta_2(T1 * z) + u_{iz} \tag{4.5}$$

In model (4.5), κ and β_1 identify the intercept and slope of profiles in T0 sessions.

A potential problem of the estimation strategy above is that it assumes variable x^*_{iz} is continuous. We can show the general point that 'centre-contribution' decisions increase with the average investment of the spokes using other estimation models that do not depend on this assumption. Ordered logit, for example, requires only that data is available in ordinal form. This is satisfied in our case. The ordered logit estimate of β_1 can be given the same interpretation as above: it captures the change in x^*_{iz} that results from an increase in average spoke contribution z of 1 unit. This is the effect we are interested in and thus it is the effect we report. One drawback of ordered logit is that we cannot easily estimate and perform statistical analysis on the intercept of the profile.

Second, we categorise each individual profile in terms of its archetypal shape and report the relative frequency of each shape. The archetypal shapes we consider are:

1. **Strictly increasing**: $c_i^{z+1} > c_i^z$, for $z \in \{0,1,2\}$
2. **Flat**: $c_i^{z+1} = c_i^z$, for $z \in \{0,1,2\}$
3. **Weakly increasing**: $c_i^{z+1} \geq c_i^z$, for $z \in \{0,1,2\}$ and the profile is not strictly increasing and not flat
4. **Decreasing**: $c_i^{z+1} \leq c_i^z$, for $z \in \{0,1,2\}$ and profile is not flat
5. **Peak at 1**: $c_i^1 > c_i^0$, and $c_i^{z+1} < c_i^z$ for $z \in \{1,2\}$
6. **Peak at 2**: $c_i^{z+1} > c_i^z$ for $z \in \{0,1\}$ and $c_i^3 < c_i^2$, and

The only strictly increasing profile possible in our game is $c_i = (0,1,2,3)$. As explained before, we define this as 'strict conditional cooperation'. We define weakly increasing profiles as 'weak conditional cooperation'.

To investigate prediction 4.1 we estimate the following linear probability model:

$$match\left(c_i^z = \bar{\alpha^z}\right)_{iz} = \delta + High\,Rate\;of\;\;Return + e_{iz} \tag{4.6}$$

$match\left(c_i^z = \bar{\alpha^z}\right)_{iz}$ is a dummy variable that takes a value of 1 if $c_i^z = \bar{\alpha^z}$, that is, if the contribution decision matches the group expectation. Variable $High\,Rate\;of\;\;Return$

is a second dummy which indicates whether the session-level rate of return to investing in the public good is $\frac{4}{5}$. We estimate model (4.6) using OLS over the sample of individuals who receive message neutral 2 in T1 treatments. We include dummy controls for the values of average spoke contribution z and for the treatment in which the decision is taken.

A positive and significant coefficient on $High\,Rate\ of\ Return$ would confirm prediction 4.1. The model of guilt aversion we have presented suggests that this effect is the result of players' desire to align their contribution profiles to the average group expectations which we disclose. However, a higher frequency of matches between c_i^z and $\bar{\alpha}^z$ can also come about in two other ways. First, when players align contributions c_i^z to first-order expectation α_i^z more frequently, and the distribution of α_i^z has significant weight on the mean.[25] Individuals in the centre of the star may have different reasons to conform to the decisions they expect others to take. For example, they could be motivated by a wish to abide to respected social norms. We will check whether this effect is at work using a regression model of this form:

$$match\left(c_i^z = \alpha_i^z\right)_{iz} = \delta + High\,Rate\ of\ Return + e_{iz} \tag{4.7}$$

Second, when r is high, players may hold more realistic forecasts about what the other players will do. We will not be able to offer an independent test of this second mechanism. However, there are no strong reasons to suspect that expectations will be significantly more precise for a higher value of r.

In some specifications of regression model (4.6), we also include controls for factors that may moderate the effect of group expectations on behaviour and we interact these

[25]If this was not the case, such alignment could actually determine a decrease in the proportion of matches between c_i^z and $\bar{\alpha}^z$. Suppose for example that half of the players set $\alpha_i^z = 0$ and the other half sets $\alpha_i^z = 2$. In this case, $\bar{\alpha}^z = 1$. If everybody aligns c_i^z to α_i^z, the number of matches between c_i^z and $\bar{\alpha}^z$ will be zero. On the other hand, suppose all players set $\alpha_i^z = 1$. Again, $\bar{\alpha}^z = 1$. Aligning c_i^z to α_i^z will now bring the fraction of the matches between c_i^z and $\bar{\alpha}^z$ to 100 percent

with the treatment dummy $High\,Rate\ of\ Return$. We are particularly interested in three variables: individual and average degree in the real-life network that links the participants of the experiment, and self-reported 'oneness'.

We hypothesise that farmers will respond more readily to group expectations when they are linked to many of the group members, and when the average number of links within the group is high. For this purposes, we rely on dyadic data which we collect at the end of each session. This data describes the bilateral relationship of each player with the other seven players. We consider that a link exists between i and j when they have spoken at least once in the past 30 days. The literature in behavioural economics has argued that individuals have stronger other-regarding concerns for peers who are close in the network (Goeree et al., 2010; Leider et al., 2009; Ligon and Schechter, 2012). Social psychologists have also put forward the hypothesis that guilt is stronger for close ties (Baumeister et al., 1994). Both of these strand of work make predictions at the dyadic level. In our experiment, on the other hand, the player is confronted with the average expectation of a set of players. Both individual and the session-level network statistics may make group expectations more salient for decisions makers in our experiment. We test for both possibilities.

We also hypothesise that a feeling of connection with the other farmers in the group would make a player particularly responsive to group expectations. This feeling is embodied in the construct of 'oneness' developed in the literature in social psychology. The feeling of oneness is defined as 'a sense of shared, merged, or interconnected personal identity' (Cialdini et al., 1997). Recent experimental evidence in economics points to the importance of oneness as predictor of behaviour in strategic environments (Tufano et al., 2012). We obtain a self-reported measure of oneness by including in the end-questionnaire the same visual survey items developed by Aron et al. (1992) and deployed in the subsequent literature in social psychology. We report this items in figure 4.7 in the appendix.

To test prediction 4.2, we estimate the following variations of model (4.5) over the sample of individuals in sessions T1*neutral*, T1*positive* and T1*negative* that have received message neutral 2:

$$x^*_{iz} = \kappa + T1positive + T1negative + u_{iz} \tag{4.8}$$

$$x^*_{iz} = \kappa + \beta_1 z + T1positive + T1negative + \beta_3(T1positive * z) + \beta_4(T1negative * z) + u_{iz} \tag{4.9}$$

Now the excluded category is the T1*neutral* treatment. Our main prediction is that the coefficient on the T1*positive* dummy in model (4.8) is positive, and that the coefficient on the the T1*negative* dummy is negative. These coefficients measure differences in average contributions across treatments, pooling over all four decisions. In model (4.9), we test separately whether effects identified in model (4.8) are produced by a shift in the intercept or a change in the slope of the contribution profiles.

Finally, to investigate prediction 4.3 we restrict attention to the the sample of individuals who have received message neutral 2 in T1*neutral* and T0 and estimate the following regression models:

$$match\left(c^z_i = \bar{\alpha}^z\right)_{iz} = \delta + T1neutral + e_{iz} \tag{4.10}$$

$$x^*_{iz} = \kappa + T1neutral + u_{iz} \tag{4.11}$$

$$x^*_{iz} = \kappa + \beta_1 z + T1neutral + \beta_2(T1neutral * z) + u_{iz} \tag{4.12}$$

We will estimate regression model (4.10) using OLS and models (4.11) and (4.12) using tobit. Model (4.10) will test whether disclosure of group expectations makes matches between contributions and group expectations more frequent in T1*neutral* compared to T0. Models (4.11) and (4.12) will explore whether the level of contributions is affected by disclosure of group expectations, and, if so, whether this happens through a change of the slope or of the intercept of the contribution profile.

To account for within-session correlation in the error terms, we correct standard errors for clustering at the session level in all models presented in this section. With 98 sessions, we have a sufficient number of clusters to apply this correction.

4.4 Data

We run our field experiment in 4 'talukas' (sub-districts) of the Indian state of Maharashtra in January and February 2014. These are the same 4 'talukas' where, in September and October 2013, we ran the link formation experiment which is presented in a previous chapter of this thesis.

As in the previous chapter, study participants are selected through door-to-door random sampling. Male adult farmers who are encountered in this exercise are invited to join the experiment. We run only one session per selected village. All our participants are fluent Marathi speakers. We translate all instructions in Marathi. A third person translates all materials back into English, to enable us to check the quality of the Marathi translation.

We run 98 sessions with 765 subjects. We have 24 sessions of T0 and T1*negative* and 25 sessions of T1*neutral* and T1*positive*. In 11 sessions we played the game with seven participants and in 4 sessions we played the game with six participants. This is mostly due to the fact that some farmers left after the beginning of the explanations.[26] In one

[26]In most cases, famers who left did so early on in the experiment, before actual decisions were made. When the game was played with seven or six participants none of the rules were changed. The main differ-

case we were not able to find eight available farmers with door-to-door sampling and run the session with seven farmers. As shown later in the balance analysis, the number of individuals per session is not correlated with treatment. Table 4.2 summarises the number of sessions and individual observations we have for each treatment.

Table 4.2: Number of observations by treatment

Treatment	**Sessions**	**Players**
T0	24	187
T1*neutral*	25	194
T1*positive*	25	195
T1*negative*	24	189
Total	98	765

At the end of the game, participants compile a short questionnaire. We hence have a small set of covariates.[27] Average age is 37 years. 76 percent of participant do not belong to a scheduled caste, tribe or an other backward caste (OBC), 32 percent of them have completed high school. We also find that average total land holdings are about 4 hectares and average land cultivated is 3.1 hectares. On average, farmers report sharing information about agriculture on a regular basis with 7 other farmers. Overall, this sample has very similar average characteristics to the sample that played the link formation game.

The farmers who take part in the experiment know each other well. In the questionnaire we ask farmer i on how many days of the previous 30 days he has had a conversation with each of the other players. The density of the within-session networks we record is very high: 61 percent of farmers have spoken with all the other

ence is that the contribution of the centre player now reaches 1 or 2 individuals less.

[27]When participants fail to answer a question or report an illegible script, we code a missing value. This explains the changing number of observations in table 4.3.

Table 4.3: Summary statistics: Individual Covariates

Variable	Mean	Std. Dev.	Min.	Max.	N
Age	36.942	10.57	19	75	760
Non backward caste	0.765	0.424	0	1	745
Completed High School	0.326	0.469	0	1	748
Land Owned	4.115	5.175	0	68	761
Land Cultivated	3.158	4.49	0	68	757
Information network size[28]	6.972	4.857	0	20	744
Oneness	5.995	1.547	1	7	747

farmers and, on average, a farmer has spoken with 6 of the other 7 farmers. Conditional on speaking, farmers have on average spoken with the other farmers in the session on 7.5 of the previous 30 days.

A second piece of descriptive evidence confirms the tight nature of the social bond between participants: self reported oneness in our sample is very high. More than 70 percent of players who answer the question choose the highest possible level of oneness. Figure 4.8 in the appendix illustrates.

Table 4.4: Summary statistics: Session Networks

Variable	Mean	Std. Dev.	Min.	Max.	N
Farmers with whom i has spoken	6.04	1.68	0	7	747
Average number of days spoken	7.5	6.62	1	30	725

The first variable reports the number of farmers with whom farmer i has spoken on a least 1 day in the last 30 days. The second variable reports the average number of days spoken with the other farmers, conditional on speaking on a strictly positive number of days.

We check participants' understanding of the game by means of a initial battery of nine questions. These cover understanding of the network map, ability to calculate payoffs, awareness of the incentives created by the payoff rule, and understanding of the strategy method. Figure 4.9 in the appendix reports the cumulative distribution of mistakes in these questions. About 48 percent of individuals make two mistakes or less in the nine questions. We call players who make two mistakes or less 'high-understanding' players. In the analysis, we explore whether results change when we estimate our regression models only on the sample of these players.

Following the understanding test, enumerators reveal the right answers to the questions and give further instructions if necessary. Hence the understanding level reported in figure 4.9 is a lower bound of the actual understanding of players at the time of play.

In tables 4.10 to 4.12 in the appendix, we present some regressions that test for covariate balance across treatments. We cannot find any statistically significant difference in average characteristics across treatments, in the number of mistakes made in the understanding questions, nor in the number of individuals who choose to leave before the end of the game.

4.5 Results

The contribution profiles of the centre of the star

Result 4.1. *The contribution profiles of the centre of the star are consistent with conditional cooperation. The spokes contribute on average half of the endowment. Farmers realise 49.7 percent of the potential gains from cooperation.*

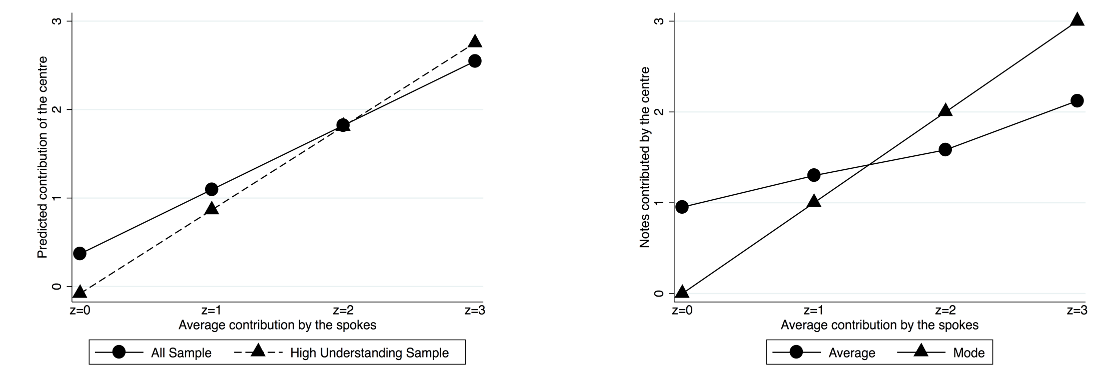
Regression analysis shows that high-understanding players choose contributions profiles consistent with strict conditional cooperation. In the first four columns of table 4.5 we report tobit estimates of the coefficients in model (4.5). When we run the regression over the whole sample of decision makers and without controls, the point estimate of coefficient β_1 is 0.73 and highly significant. A Wald test indicates that this coefficient is significantly lower than one, while the coefficient on the constant κ is significantly higher than 0 at the 10 percent level. Thus, on average, players choose contribution profiles with a higher intercept and a flatter slope than those implied by 'strict conditional cooperation'. However, when we restrict the analysis to high-understanding players, the intercept becomes statistically indistinguishable from 0, and the point estimate of β_1 is very close to 1. A Wald test cannot reject the null hypothesis that this coefficient is equal to 1. High understanding players, on average, play 'strict conditional cooperation'. This is evident when, in figure 4.4a, we plot the predicted profiles that are implied by the regression estimates.

In the tobit model the independent variable z is constrained to have the same effect on the likelihood that the dependent variable is at a corner solution and on the value of the dependent variable when this is not at the corner. We provide qualitative evidence to support this assumption by running a probit regression to investigate the effect of z on the likelihood that the dependent variable is at the corners.[29] Table 4.16 in the appendix shows the results. In both cases, the coefficient on spoke average contribution z has the same direction as in the tobit model, and has a magnitude that is lower, but roughly comparable.

The next four columns of table 4.5 report ordered logit estimates. These show that the significance of the coefficient β_1 is robust to the introduction of this alternative estimation strategy. The magnitude is also similar to that reported for the tobit model. In all four specifications, the point estimate of β_1 is statistically indistinguishable from 1.

[29]We apply both a lower and an upper limit. To study the effect of z on the upper limit, we analyse the probability that $c_i^z = 3$. To study the effect of z on the lower limit, we analyze the probability that $c_i^z > 0$. In both cases, we expect a positive coefficient, similar in magnitude to those reported in table 4.5.

Figure 4.4: Contribution profiles of the centre of the star



(a) Predicted profiles

(b) Mode and average in the data

Table 4.5: Regression: contributions of the centre player

	Tobit				Ordered Logit			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel a								
Spoke average	.726 (.108)***	.773 (.106)***	.945 (.101)***	1.003 (.104)***	.828 (.123)***	.877 (.120)***	1.134 (.119)***	1.180 (.120)***
T1	.161 (.247)	.205 (.261)	.052 (.323)	.132 (.336)	.229 (.261)	.291 (.271)	.190 (.342)	.257 (.352)
T1*Spoke average	-.103 (.119)	-.150 (.118)	-.144 (.116)	-.225 (.120)*	-.154 (.127)	-.207 (.125)*	-.224 (.125)*	-.290 (.126)**
Const.	.369 (.220)*	.141 (.354)	-.080 (.297)	-.433 (.492)				
Panel b								
H0: $\beta_1 = 1$, H1: $\beta_1 \neq 1$	6.38 (.012)**	4.54 (.033)**	0.30 (.585)	0.00 (.976)	1.98 (.16)	1.05 (.306)	1.28 (.258)	2.24 (.135)
Obs.	3060	2732	1496	1344	3060	2732	1496	1344
Cluster N	98	98	98	98	98	98	98	98
Pseudo R^2	.046	.047	.08	.081	.058	.061	.102	.104
Log-likelihood	-4581.009	-4082.093	-2144.804	-1925.252	-3985.161	-3550.656	-1847.164	-1657.385
Controls		✔		✔		✔		✔
High Understanding			✔	✔			✔	✔

The dependent variable is the number of notes contributed to the public good by player i for 'centre contribution' decision z. The first four columns present a tobit regression, with an upper limit of 3 and a lower limit of 0. The last four columns present an ordinal logit regression. Columns 3,4,7,8 restrict the analysis to 'high-understanding' players who have made 2 mistakes or less in the initial understanding questions. Columns 2, 4, 6, 8 include controls for the players' age, area of land owned, area of land cultivated, number of contacts in real information networks, self-reported oneness with the group, and dummies for having completed secondary education, for being Hindu, and for belonging to a non backward caste. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level. Panel b reports the F statistics (and p value in parenthesis) for a one-sided Wald test on estimated coefficient $\hat{\beta}$.

Graphical analysis of average and modal values of contribution decisions c_i^z confirms that these closely match the average contribution of the spoke players. Figure 4.4b illustrates.

The profile $c_i = (0, 1, 2, 3)$ is the most frequently chosen by farmers in the game. About 20 percent of them choose this profile and can hence be classified as 'strict conditional cooperators'. Tables 4.13 and 4.14 in the appendix report this data. Farmers also choose a variety of strategies that are weakly increasing in the average contribution of the spokes. These add up to 22 percent of all profiles. If we restrict the analysis to high-understanding players, weakly increasing profiles account for about 38 percent of all profiles. The majority of high-understanding players (57 percent) thus chooses strategies that increase (strictly or weakly) with the average of the spokes. Figure 4.5 illustrates.

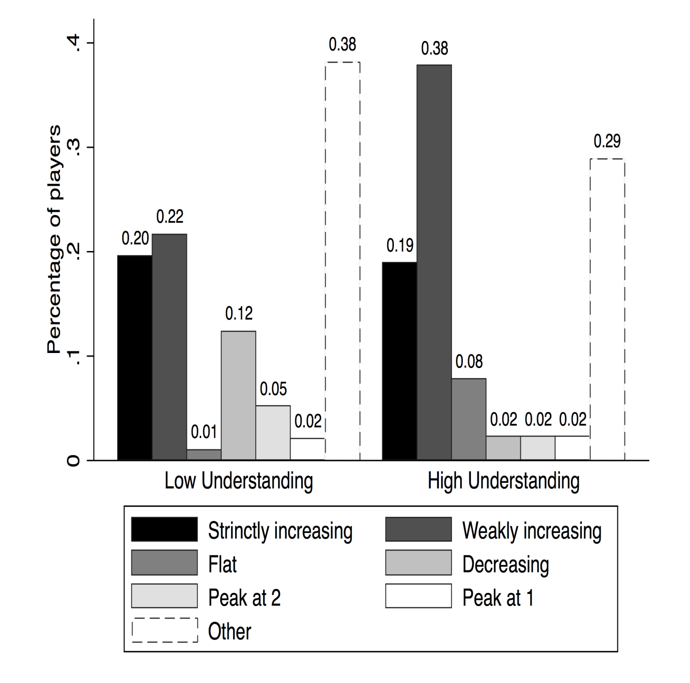
The three most common weakly increasing strategies are: $(0, 0, 1, 3)$, $(0, 0, 1, 2)$, and $(0, 0, 2, 2)$. These are all profiles in which the centre of star contributes weakly less than the average contribution of the spokes in every decision.

The remaining profiles show a large degree of heterogeneity. A small group of players do not condition their contributions on the average of the spokes. This group is composed both of players who never contribute anything (4.4 percent of high-understanding players) and players who always contribute the full amount (3.3 percent of high-understanding players). A second group chooses profiles where contributions weakly or strictly decrease with the average contribution of the spokes. However, this group is mostly composed of low-understanding players. A third group chooses profiles that peak when spoke average contribution is 1 or 2.[30] Finally, about 38 percent of low-understanding players and 29 percent of high-understanding players choose profiles that are not consistent with the archetypal candidates we have listed.

We have focused so far on the profile of four decisions taken by the centre of the star. Only one of these decisions is implemented at the end of the game. The average contribution of the spokes determines which of the decisions is implemented. Spokes

[30] This type of profile was documented also in the study of Fischbacher et al. (2001).

Figure 4.5: Archetypal contribution profiles, by understanding

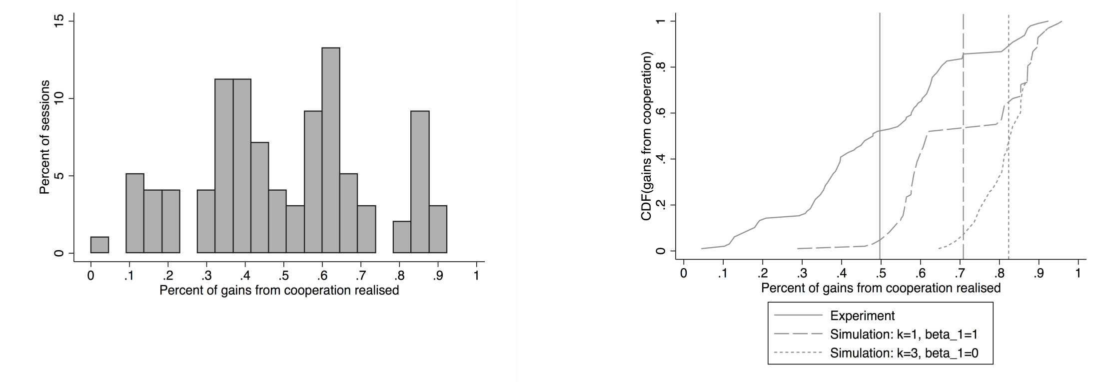


contribute on average 1.496 notes. After rounding, 51 percent of sessions have a spoke average of 1 and 47 percent a spoke average of 2. Given this, the centre player on average contributes 1.49 notes. Figure 4.10 in the appendix illustrates.

The sum of group payoffs is maximised when every player contributes the maximum number of notes to the public good. In our data, the combination of conditional cooperation and a relatively low average contribution level of the spokes determine that only about 50 percent of the potential gains from cooperation are realised.[31] The left panel of figure 4.6 shows the histogram of the realised gains from cooperation in the 98 experimental sessions. The right panel plots the cumulative distribution function of the same variable for the experimental sessions, against the CDFs of two simulated distributions. The vertical lines indicate the averages of the three distribu-

[31]Let $\Pi(3,r)$ be the sum of payoffs that would accrue to each player if every player contributes 3 notes to the public good when the return to investing in the public good is r. Define $\Pi(0)$ as the sum of payoffs when every player contributes 0 notes to the public good. $\Pi(3,r)-\Pi(0)$ represents the increase in aggregate payoff that is achieved when players make the maximum contributions to the public good. These are the potential 'gains from cooperation'. Let $\Pi_{s|r}$ be the sum of individual payoffs in session s, with rate of return r. $\frac{\Pi_{s|r}-\Pi(0)}{\Pi(3,r)-\Pi(0)}$ indicates the fraction of the potential gains from cooperation that is realised in session s.

Figure 4.6: Aggregate Efficiency



(a) Histogram (b) Comparison with two stylised profiles

See note 31 for a mathematical definition of 'gains from cooperation'.

tions. In the two simulations, we hold the spoke contributions constant and change the centre contribution profile to either (i) $\kappa = 1$, $\beta_1 = 1$, or (ii) $\kappa = 3$, $\beta_1 = 0$. Profile (i) is that of a strict conditional cooperator plus a positive shock of one to the intercept. If all centres of the star played this profile, realised efficiency would increase to 70 percent. Profile (ii) is a flat profile where the centre of the star always contributes the maximum amount. If all centres of the star played this profile, realised efficiency would increase to 82 percent. These results confirm the important role of the centre player in securing the gain from cooperation.

Are players influenced by the expectations of others?

We now move to investigation of the two predictions regarding our treatments. Our first prediction is that, when r increases to $\frac{4}{5}$, the frequency of matches between group expectations and contributions of players who receive message 2 should also increase.

Result 4.2. *In T1 treatments with $r = \frac{3}{5}$, among players who receive message neutral 2, 42 percent of the contribution decisions c_i^z match the group expectation $\bar{\alpha}^z$. Matches become more frequent when the rate of return to investing in the public good increases to $r = \frac{4}{5}$. For high understanding players, match frequency significantly increases by 11* *percentage points (31 percent).*

This is evidence in support of prediction 4.1. When we look at the whole sample of players who receive message neutral 2 in T1 treatments, 42 percent of decisions match the group expectation that has been disclosed on the board. The frequency of matches increases by 3 percentage points when the rate of return to investing in the public good is raised from $\frac{3}{5}$ to $\frac{4}{5}$. If we restrict the analysis to high understanding players, 48 percent of decisions match the group expectation. For these players, the higher rate of return generates a significant 11 percentage points increase in the frequency of matches in the regression without controls, and a highly significant 15.6 percentage points increase in the regression with controls. These are large effects: 31 and 58 percent of the frequency when $r = \frac{3}{5}$, respectively. Furthermore, consistently with our model, the larger part of this effect (77 percent) comes from a reduction in the frequency of decisions where $c_i^z < \bar{\alpha}^z$. Figure 4.16 in the appendix illustrates.

Table 4.6: Linear Probability Model: match between contribution c_i^z and group expectations $\bar{\alpha}^z$

	(1)	(2)	(3)	(4)
High rate of return	.028 (.043)	.043 (.044)	.110 (.058)*	.156 (.059)***
Const.	.376 (.050)***	.376 (.132)***	.352 (.072)***	.268 (.187)
Obs.	1152	1036	592	532
Cluster N	74	73	66	62
Controls		✔		✔
High Understanding			✔	✔

OLS regression. The dependent variable is a dummy variable that takes a value of 1 if $c_i^z = \bar{\alpha}^z$. 'High rate of return' is a dummy for whether the session value of r is $\frac{4}{5}$. The sample includes all players who have received message neutral 2 in treatments T1*neutral*, T1*positive*, T1*negative*. Columns 3 and 4 restrict the analysis to players who have made 2 mistakes or less in the initial understanding questions. Columns 2 and 4 include controls for the players' age, area of land owned, area of land cultivated, number of contacts in real information networks, self-reported oneness with the group, and dummies for having completed secondary education, for being Hindu, and for belonging to a non backward caste. All regressions include dummies for whether spoke average z is equal to 1, 2, or 3 and for whether treatment is T1*positive* or T1*negative*. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level.

A high rate of return to investments in the public good has no comparable effect on the frequency of matches between contributions c_i^z and individual expectations α_i^z. Table 4.17 in the appendix reports regression results. Over the whole sample, the frequency of these matches increases by an insignificant 1.9 percentage point. Restriction to high-understanding players does not change this coefficient appreciably. This confirms that the higher coincidence of contributions and group expectations is not the indirect result of an increase in the frequency of matches between contributions and individual expectations..

Furthermore, the change in the rate of return to contributions has only a mild effect on the *level* of individual expectations, and has no statistically significant effect on the level of $\bar{\alpha^z}$, nor on the level of contributions. Table 4.18 in the appendix reports all regression estimates. Thus, while a higher value of r significantly affects the frequency of matches between contributions and group expectations, it does not cause contemporaneous changes in other choice variables.

We investigate a number of factors that may moderate the effect of group expectations and hence determine the frequency of matches. We focus in particular on individual degree, average session-level degree and on the self-reported feeling of oneness. Table 4.19 in the appendix report results of regressions where we do not include the interaction between the variable of interest and the treatment dummy. When we restrict the sample to high-understanding players, the coefficients on average degree and on oneness are positive but small. The other coefficients are very close to zero. We are not able to find statistically significant effects in any of these regressions.

On the other hand, the interaction between a high rate of return and a high level of average session degree is positive, significant and of a large magnitude. Figure 4.18 plots the predicted treatment effect for different percentiles of the distribution of average degree. At the tenth percentile, the effect is slightly negative. At the ninetieth percentile the effect is close to a 0.3, which corresponds to a thirty percentage point increase in the likelihood of a match. The interaction with individual degree, on the other hand, is not significant. The interaction with oneness has (surprisingly) a negative coefficient and is significant at the 10 percent level. Table 4.20 presents these results.

We now move to prediction 4.2 about the contributions of players who receive message 2 in sessions T1*positive* and T1*negative*.

Result 4.3. *The manipulation of expectations is weak. We cannot offer a test of prediction 4.2.*

The manipulation of expectations is weak. Table 4.7 shows regression estimates that illustrate this point. The $\bar{\alpha}^z$ values which we disclose publicly are not significantly higher (lower) in T1*positive* (T1*negative*) sessions compared to T1*neutral* sessions. This is true both if we pool the four values together in a small panel of sessions, or if we analyse each average expectation value separately. We shed some light on how this comes about by comparing the expectations of individuals who received message neutral 1 in T1*neutral*, to the expectations of individuals who received the positive message in T1*positive*, and individuals who received the negative message in T1*negative*. This analysis, reported in tables 4.21 and 4.22 in the appendix, shows that the positive message fails to significantly affect expectations, while the negative message reduces them by about half a unit, an effect which we can detect with some statistical precision. The reduction of expectations in T1*negative* is not of a sufficiently large magnitude to modify the group average in a significant way. As average session-level expectations are not affected by the treatment, we cannot offer a convincing test of prediction 4.2. Such prediction, in fact, rests on the premise that expectations have been experimentally manipulated in the hypothesised direction.

In table 4.23 we show that there are no treatment effects on contributions c_i^z. Tables 4.24 and 4.25 in the appendix confirm that we are equally unable to find treatment effects when we analyse each contribution decision separately, or when we estimate model specification (4.9). As argued above, however, this *does not* constitute evidence

Table 4.7: Ordered logit regression over session-level average expectations $\bar{\alpha}^z$

	All $\bar{\alpha}^z$	$\bar{\alpha}^0$	$\bar{\alpha}^1$	$\bar{\alpha}^2$	$\bar{\alpha}^3$
	(1)	(2)	(3)	(4)	(5)
T1*positive*	-.366	-.638	.228	-.486	-.356
	(.410)	(.558)	(.608)	(.572)	(.648)
T1*negative*	-.498	-.638	-1.002	-.238	-.184
	(.358)	(.558)	(.706)	(.578)	(.653)
Obs.	296	74	74	74	74
Cluster N	74	74	74	74	74
Pseudo R^2	.237	.012	.034	.007	.003
Log-likelihood	-235.049	-69.817	-49.86	-50.686	-54.934

Ordered logit regression. The dependent variable is the session level average of expectations: $\bar{\alpha}^z$. The sample includes all sessions in treatments T1*neutral*, T1*positive* and T1*negative*. Column 1 pools the four values of $\bar{\alpha}^z$ for each session. Columns 2-5 analyse separately the values of $\bar{\alpha}^1$, $\bar{\alpha}^2$, $\bar{\alpha}^3$, and $\bar{\alpha}^4$, respectively. Confidence: *** $\longleftrightarrow$ 99%, ** $\longleftrightarrow$ 95%, * $\longleftrightarrow$ 90%. Standard errors are reported in parenthesis. Standard errors are corrected for clustering at the session level in column 1.

against the hypothesis that guilt aversion influences public good contributions in networks.

Do players expect conditional cooperation?

We now turn to the analysis of expectation profiles α_i. Our main result is the following:

Result 4.4. *Players expect conditional cooperation from the centre of the star.*

Our strongest piece of evidence is given by estimation of model (4.5) with expectations α_i^z as the dependent variable. Coefficient β_1 now measures the extent to which players *expect* others to increase their centre of the star contributions when the average spoke contribution increases. Table 4.8 reports the coefficient estimates. The average expectations profiles reported for the unrestricted sample have a positive value of κ, significantly higher than 0, and a positive value of β_1. When we restrict the sample to high-understanding players in column 3, κ is not significantly different from 0 and β_1 is not significantly different from 1. These results are similar to those for contribution profiles reported in table 4.5. High-understanding farmers in our sample expect strict conditional cooperation.

Table 4.8: Regression: expectations about the contribution of the centre player

	Tobit				Orderd Logit			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel a								
Spoke average	.670 (.089)***	.679 (.095)***	.946 (.084)***	1.039 (.108)***	.688 (.091)***	.687 (.096)***	1.003 (.094)***	1.056 (.114)***
T1	-.220 (.215)	-.175 (.238)	-.342 (.311)	-.104 (.365)	-.230 (.209)	-.175 (.226)	-.351 (.298)	-.124 (.332)
T1*Spoke average	.088 (.106)	.089 (.112)	.108 (.107)	.00008 (.133)	.093 (.101)	.090 (.106)	.127 (.103)	.025 (.122)
Const.	.427 (.175)**	-.179 (.328)	-.113 (.269)	-1.145 (.468)**				
Panel b								
H0: $\beta_1 = 1$, H1: $\beta_1 \neq 1$	13.72 (.000)***	11.38 (.001)***	0.41 (.524)	0.13 (.719)	11.78 (.001)***	10.71 (.001)***	0.00 (.971)	0.24 (.626)
Obs.	3060	2732	1496	1344	3060	2732	1496	1344
Cluster N	98	98	98	98	98	98	98	98
Pseudo-R^2	.053	.055	.101	.1	.066	.067	.126	.123
Log-likelihood	-4534.488	-4038.437	-2084.425	-1873.702	-3955.106	-3527.313	-1799.159	-1622.836
Controls		✔		✔		✔		✔
High Understanding			✔	✔			✔	✔

The dependent variable is expectation α_i^z. The first four columns present a tobit regression, with an upper limit of 3 and a lower limit of 0. The last four columns present an ordinal logit regression. Columns 3,4,7,8 restrict the analysis to players who have made 2 mistakes or less in the initial understanding questions. Columns 2, 4, 6, 8 include controls for the players' age, area of land owned, area of land cultivated, number of contacts in real information networks, self-reported oneness with the group, and dummies for having completed secondary education, for being Hindu, and for belonging to a non backward caste. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level. Panel b reports the F statistics (and p value in parenthesis) for a one-sided Wald test on estimated coefficient $\hat{\beta}$.

The most frequent expectation profile is the strictly increasing profile $\alpha_i = (0, 1, 2, 3)$. As before however, the combined category of weakly increasing profiles occurs more frequently than the strictly increasing profile. The most common weakly increasing profiles are: $(0, 0, 1, 3)$, $(0, 0, 1, 2)$, and $(0, 1, 2, 2)$. Again, these are all profiles where the centre of star contributes weakly less than the spokes for every decision. These results are reported in figure 4.11 and table 4.15.

Expectations are on average correct, as shown in figure 4.12 and as the similarity of the estimated regression coefficients κ and β_1 in tables 4.5 and 4.8 indicates. However, expectations are not particularly precise. In figure 4.13 we compute, for each decision z, the probability that player i's expectation α_i^z is equal to the average value of c_i^z among the other seven players in the session. In other words, we calculate the fraction of times in which farmers correctly guess the behaviour of the other farmers in the group. We calculate a confidence interval around this probability and test whether it lies above the probability of having an accurate expectation when this expectation is randomly chosen.[32] For α_i^1, α_i^2 and α_i^3 farmers' predictions are correct significantly more often than random predictions. However, for α_i^1 and α_i^3 , the confidence interval is actually very close to including the value under random prediction. Furthermore, even for decision c_i^2, which farmers are best at predicting, mistaken predictions are more common than correct guesses.

There is a correlation between expectations α and strategies, but this is by no means perfect.[33] 60 percent of players who expect strict conditional cooperation from others are strict conditional cooperators themselves, while 13 percent of players who do not expect strict expect conditional cooperation from others choose a profile consistent with strict conditional cooperation. The respective numbers for farmers who choose weakly increasing profiles are 50 and 20 percent. Similarly, 50 percent of players who choose a flat contribution profile also expect others to choose a flat profile, while only 4 percent of players who do not choose a flat profile expect others to choose a flat profile. Figure 4.14 illustrates. These figures are high, however, in only 15 percent of cases $c_i^z = \alpha_i^z \; \forall z$, as shown in figure 4.15.

Finally, we study prediction 4.3. In table 4.9 we show that matches between contributions and average group expectations are more likely in T1*neutral* than in T0 by about 9 percentage points (27 percent), an effect significant at the 5 percent level. When we add controls or restrict to high-understanding players, the coefficient stays close to 0.09. This result is consistent with inaccuracy in second order beliefs: players with an inaccurate prior fail to match true group expectations when these are not disclosed in T0.

Figure 4.17 in the appendix shows that the higher frequency of matches comes

[32] As there are four possible values of c_i^z, the probability of picking the right value when guessing at random is 0.25.

[33] This is a correlation between what player i expects the other seven players to do and his own decision. It could be driven, for example, by conformism. Alternatively, it could reflect a bias in expectations: farmers disproportionately expect others to act like oneself.

Table 4.9: Linear probability model: match between c_i^z and $\bar{\alpha}^z$ in T1*neutral* and T0

	(1)	(2)	(3)	(4)
T1*neutral*	.091 (.037)**	.086 (.038)**	.092 (.056)*	.090 (.060)
Const.	.334 (.020)***	.391 (.089)***	.364 (.029)***	.469 (.141)***
Obs.	1524	1332	700	616
Cluster N	49	49	48	47
Controls		✔		✔
High understanding			✔	✔

OLS regression. The dependent variable is a dummy variable that takes a value of 1 if $c_i^z = \bar{\alpha}^z$. The sample includes all subjects in T1*neutral* and T0. Columns 3 and 4 restrict the analysis to players who have made 2 mistakes or less in the initial understanding questions. Columns 2 and 4 include controls for the players' age, area of land owned, area of land cultivated, number of contacts in real information networks, self-reported oneness with the group, and dummies for having completed secondary education, for being Hindu, and for belonging to a non backward caste. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parenthesis.

about mainly through a reduction in contribution decisions below the average of the spokes. This effect is too weak, however, to influence overall contributions significantly. Tobit regressions reported in table 4.26 show that the level of contributions is not significantly higher in T1*neutral* than in T0.

When we try to separately look at changes in the intercept and slope of the profiles, we find that in T1*neutral* the intercept is somewhat higher and the slope less steep. Neither of these effects is however measured with sufficient statistical precision.

Result 4.5. *In T1*neutral*, players are 9 percentage points more likely to set* $c_i^z = \bar{\alpha}^z$ *than players in T0.*

4.6 Conclusion

We play a one-shot public good game over an exogenous star network with farmers in rural India. The network determines who benefits from the public good contributions of each player. We use the strategy method to obtain from each player a contribution decision for the case where the player is assigned to the spoke position *and* for the case where the player is assigned to the centre of the star. The centre of the star is further allowed to condition his contribution decision on the (rounded) average contribution of the spokes.

The experimental literature has long recognised that heterogeneity in individual characteristics affects the level and dynamics of cooperation in human groups (Ledyard, 1995; Reuben and Riedl, 2013). Social networks where individuals are asymmetrically connected create an important dimension of heterogeneity. The star network, in particular, is characterised by a strong asymmetry: contributions by the centre of the star benefit each of the spokes, while contributions by the spokes benefit only the centre player. This makes the centre player particularly effective at raising aggregate welfare. A player at the centre of the star concerned with relative efficiency would thus contribute more than the average of the spokes. Further, because his payoff is higher than that of the spokes when everybody contributes the same positive amount, payoff equality requires higher contributions from the player at the centre.

We find that, when in the position of the centre of the star, farmers choose contribution profiles that match the average contribution of the spokes. This is despite the fact that considerations of relative efficiency and equality militate in favour of higher contributions. Players in experiments with homogeneous groups often contribute just as much as other players contribute. This is usually referred to as 'conditional cooperation'. A key contribution of this study is to show that the player at the centre of a star network also behaves as a conditional cooperator.

Second, we present evidence suggesting that the centre of the star responds to the expectations other players hold about his behaviour. For this purpose, we elicit players' expectations about the contribution profile chosen by the centre of the star and disclose average expectations in randomly chosen sessions. Players match their contributions with the disclosed values in 42 percent of the cases. For players who have understood the rules of the game well, a decrease in the monetary cost of contributions is associated with a large (11 percentage points) increase in the frequency of matches. This effect is predicted by the model of guilt aversion. The average degree in the real-world network that connects farmers in the same session is a moderating factor: at the 90th percentile of average degree, the effect grows to almost 30 percentage points.

Third, we find that farmers expect the centre of the star to contribute as much as the spokes. They match their contributions to group expectations less frequently (by 9 percentage points) when group expectations are not disclosed. This is consistent with farmers holding inaccurate beliefs about what players in the network expect from the centre of the star. Disclosure of group expectations is not associated with an increase in average public good investments by the centre player.

Our results carry implications for policy and several leads for future research. First, when players are conditionally cooperative, public good games admit only equilibria characterised by symmetric contribution levels. Our findings suggest that individuals are likely to deviate from asymmetric contribution configurations even when they occupy the high-efficiency central position in a network. This is an important point for the design of public policies that require cooperation from selected, socially central individuals in the field, such as injection points for the diffusion of innovations (BenYishay and Mobarak, 2014; Berg et al., 2013; Kondylis et al., 2014).

A second important lesson is about the provision of information about the expectations of other players in the network. This treatment generates more matches between contributions and average expectations, but does not lead to higher contribution levels. In a widely cited study, Jensen (2010) documents significant improvements in schooling attainment following an intervention that informs students with inaccurately low priors about average returns to schooling. On the contrary, information provision is not sufficient to improve welfare in our experiment. The expectations we disclose reflect prevalent levels of pro-sociality. For example, they do not forecast that the player at the centre of star will contribute above strict conditional cooperation. If the policy maker aims to induce contributions above conditional cooperation from central players, separate interventions to incentivise the centre of the star are required. Disclosure of information about expectations can be best used to entrench behavioural change initially generated by means of incentives.

In terms of future research, we believe that this study illustrates the potential for using the strategy method to study public good contributions in heterogeneous groups. We envisage further exploration of specific dimensions of network heterogeneity. For example, as the sociological literature has long emphasised the importance of brokers for aggregate network outcomes, the effects of between-group centrality can be separated from those of degree centrality. Through the strategy method the researcher can both investigate how the behaviour of a specific individual varies when his position in an exogenous network changes, and how the contribution profiles of individuals correlate with the position they occupy in real-world networks. Other dimensions of heterogeneity that are not related to features of the network can also be explored.

Further, in our study we find evidence in support of a role of guilt aversion in determining public good contributions in networks. We also uncover the moderating influence of the structure of the real-world network that links study participants. These results lend weight to models of decision utility that incorporate expectations and socially-determined moderating factors. The development and empirical validation of such models is a particularly promising direction for future research that wants to understand within-individual variation in pro-social behaviour across contexts. This effort will help determine the scope for welfare-improving policy interventions that promote pro-social behaviour in different settings. More research is needed, including, in particular, direct measurement of first- and second-order expectations and further laboratory tests of the manipulability of expectations and their causal influence on behaviour. Field settings where participants are linked through a rich and varied web of connections are particularly appropriate to study how social structure moderates these effects.

Appendix

A.1 Figures

Figure 4.7: Oneness question

Q 16. ...Which one of these pictures best describes your relationship with the group?   Please circle the desired picture

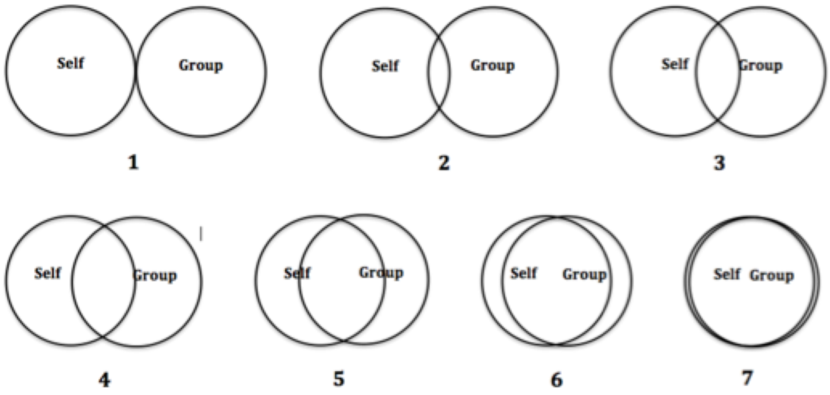


Figure 4.8: Histogram of self-reported oneness

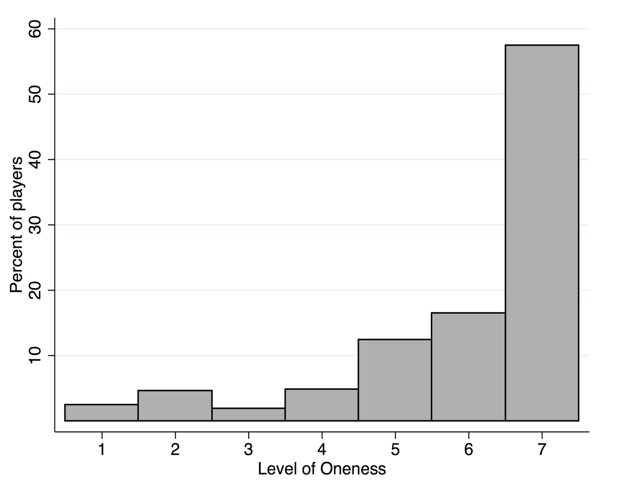


Figure 4.9: Cumulative distribution of mistakes

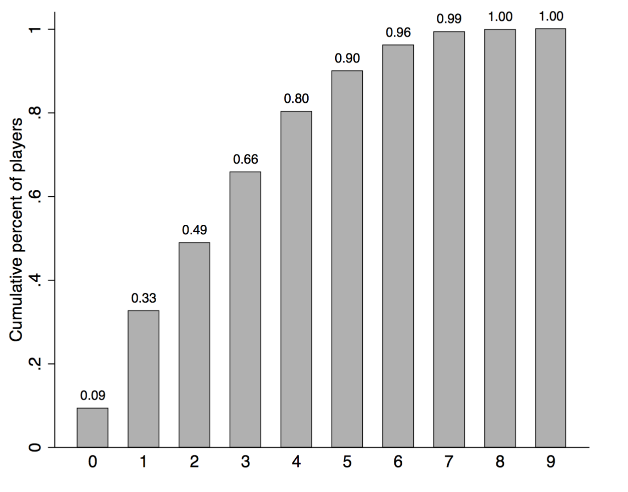
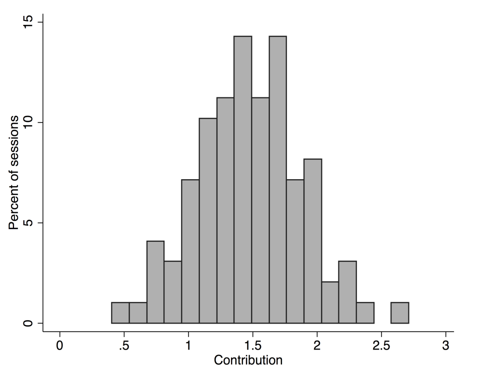
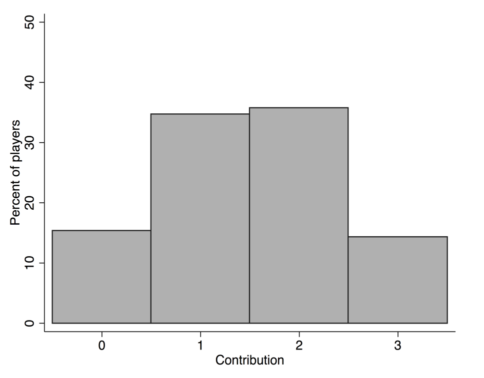


Figure 4.10: Final contributions



(a) Session average of spoke players contribution



(b) Centre player contribution

Figure 4.11: Archetypal expectation profiles, by understanding

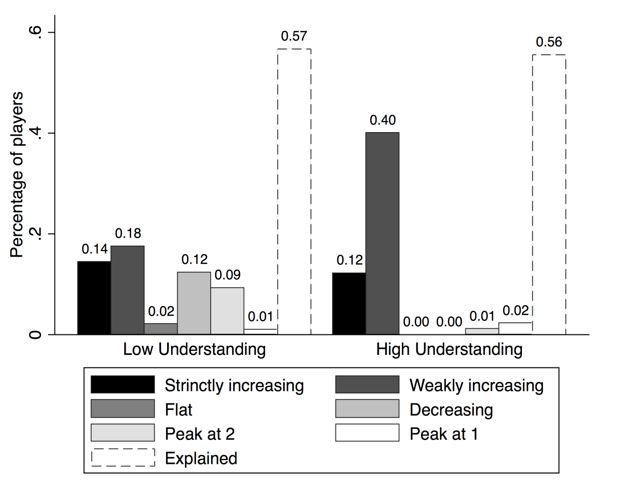


Figure 4.12: Average values of contributions and expectations

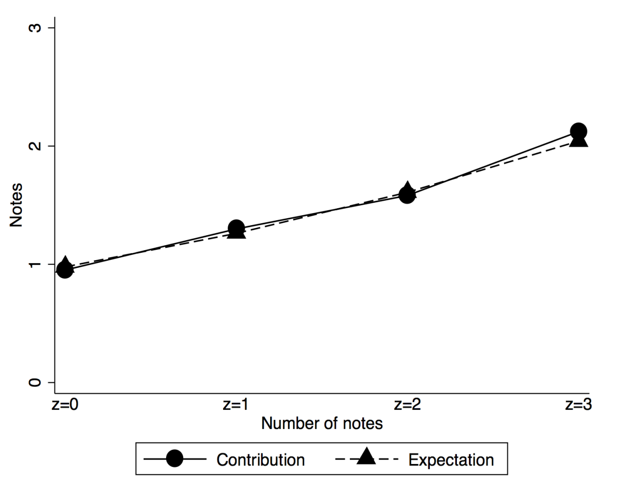


Figure 4.13: Probability $\alpha^z_{i,s} = \sum_{j \in N_s \setminus i} \frac{c^z_{j,s}}{7}$

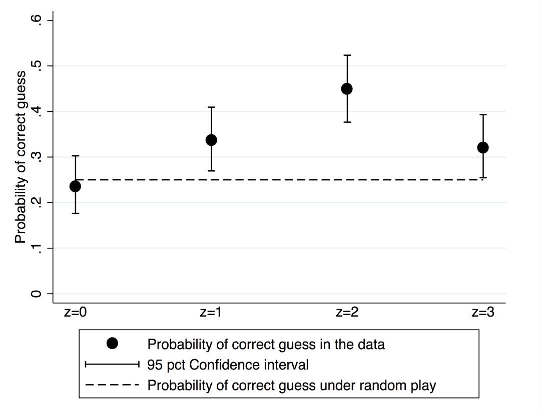
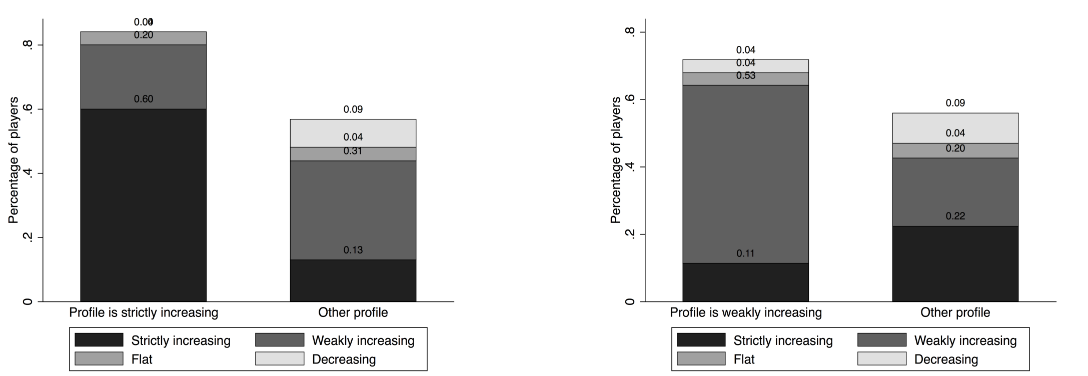


Figure 4.14: Expectation profiles by player's contribution profile



(a) By strictly increasing contribution profile (b) By weakly increasing contribution profile

Figure 4.15: Cumulative distribution of $e_i = \sum_{z=0}^{3} I\left(c_{i,s}^z = \alpha_{i,s}^z\right)$

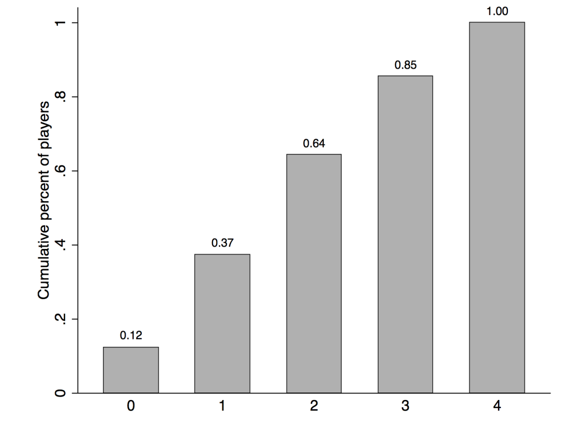
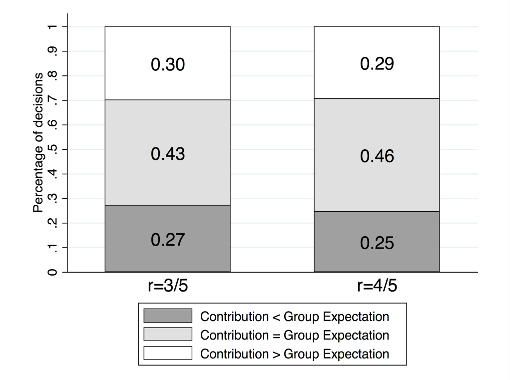
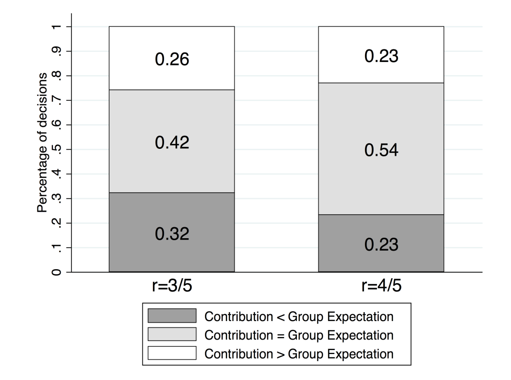


Figure 4.16: Match between contribution c_i^z and group expectations $\bar{\alpha}^z$. Comparison of T1 treatments.



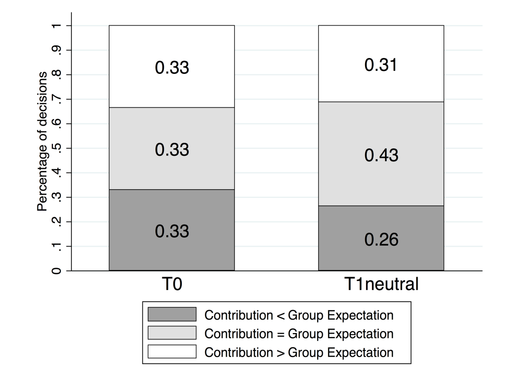
(a) Full sample



(b) High-understanding players

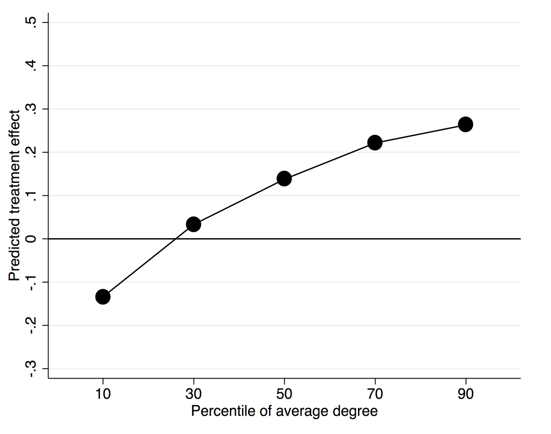
Panel (a) includes all individuals who received message 2 in T1 sessions. Panel (b) restricts the sample to all individuals who received message 2 in T1 sessions *and* made at most 2 mistakes in the initial understanding questions.

Figure 4.17: Match between contribution c_i^z and group expectations $\bar{\alpha}^z$. Comparison of T0 with T1*neutral*



The sample includes all individuals in treatments T0 and T1*neutral*.

Figure 4.18: Treatment effect at different percentiles of average session-level degree



A.1 Tables

Table 4.10: Balance test 1

	Age	UpperCaste	HigherEdu	LandOwned	LandCult	NetSize	Oneness	Understanding	SessionN
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
T1	-.605	.099	-.019	-.785	-.177	-.624	-.169	.099	-.032
	(1.421)	(.072)	(.069)	(.597)	(.551)	(.706)	(.228)	(.253)	(.159)
T2p	.320	.013	-.021	.161	-.137	.505	-.180	-.080	.008
	(1.358)	(.087)	(.070)	(.752)	(.592)	(.781)	(.217)	(.265)	(.156)
T2n	-.447	.037	-.016	-.518	-.183	-.164	-.090	-.173	.083
	(1.382)	(.078)	(.062)	(.608)	(.547)	(.660)	(.181)	(.250)	(.138)
Obs.	760	745	748	761	757	744	747	765	98

OLS regressions. The dependent variable is indicated in the row's name. 'HigherEdu' is a dummy that takes the value of 1 if the respondent has completed secondary school. 'Upper caste' is a variable that takes value of 1 if respondent is not from a schedule caste, a scheduled tribe or an Other Backward Caste. 'LandCult' is the area of land cultivated in hectares. 'NetSize' is the self reported number of peers with whom the farmer exchanges advice on agricultural matters. 'Oneness' is a number from 1 to 7. Higher numbers reflect an increasing feeling of oneness. 'Understanding' refers to the number of mistakes in the initial understanding questions. The last column is a regression over a session level outcome-'SessionN'- the number of participants in each session. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors clustered at the session level reported in parentheses in columns 1-8. Robust standard errors are reported for the regression in column 9.

Table 4.11: Balance test 2

	Age	UpperCaste	HigherEdu	LandOwned	LandCult	NetSize	Oneness	Understanding	SessionN
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
T0	.605	-.099	.019	.785	.177	.624	.169	-.099	.032
	(1.421)	(.072)	(.069)	(.597)	(.551)	(.706)	(.228)	(.253)	(.159)
T2p	.925	-.086	-.002	.946	.040	1.129	-.011	-.179	.040
	(1.296)	(.082)	(.069)	(.641)	(.521)	(.801)	(.234)	(.251)	(.145)
T2n	.157	-.062	.003	.267	-.007	.460	.079	-.272	.115
	(1.322)	(.073)	(.062)	(.464)	(.470)	(.683)	(.201)	(.234)	(.125)
Obs.	760	745	748	761	757	744	747	765	98

OLS regressions. The dependent variable is indicated in the row's name. 'HigherEdu' is a dummy that takes the value of 1 if the respondent has completed secondary school. 'Upper caste' is a variable that takes value of 1 if respondent is not from a schedule caste, a scheduled tribe or an Other Backward Caste. 'LandCult' is the area of land cultivated in hectares. 'NetSize' is the self reported number of peers with whom the farmer exchanges advice on agricultural matters. 'Oneness' is a number from 1 to 7. Higher numbers reflect an increasing feeling of oneness. 'Understanding' refers to the number of mistakes in the initial understanding questions. The last column is a regression over a session level outcome-'SessionN'- the number of participants in each session. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors clustered at the session level reported in parentheses in columns 1-8. Robust standard errors are reported for the regression in column 9.

Table 4.12: Balance test 3

	Age	UpperCaste	HigherEdu	LandOwned	LandCult	NetSize	Oneness	Understanding	SessionN
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
T0	-.320	-.013	.021	-.161	.137	-.505	.180	.080	-.008
	(1.358)	(.087)	(.070)	(.752)	(.592)	(.781)	(.217)	(.265)	(.156)
T1	-.925	.086	.002	-.946	-.040	-1.129	.011	.179	-.040
	(1.296)	(.082)	(.069)	(.641)	(.521)	(.801)	(.234)	(.251)	(.145)
T2n	-.768	.024	.005	-.679	-.046	-.669	.090	-.093	.075
	(1.254)	(.087)	(.062)	(.651)	(.517)	(.761)	(.189)	(.247)	(.121)
Obs.	760	745	748	761	757	744	747	765	98

OLS regressions. The dependent variable is indicated in the row's name. 'HigherEdu' is a dummy that takes the value of 1 if the respondent has completed secondary school. 'Upper caste' is a variable that takes value of 1 if respondent is not from a schedule caste, a scheduled tribe or an Other Backward Caste. 'LandCult' is the area of land cultivated in hectares. 'NetSize' is the self reported number of peers with whom the farmer exchanges advice on agricultural matters. 'Oneness' is a number from 1 to 7. Higher numbers reflect an increasing feeling of oneness. 'Understanding' refers to the number of mistakes in the initial understanding questions. The last column is a regression over a session level outcome-'SessionN'- the number of participants in each session. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors clustered at the session level reported in parentheses in columns 1-8. Robust standard errors are reported for the regression in column 9.

Table 4.13: Most frequently chosen contribution profiles in T0. All players

Contribution profile c_i	**Percentage**
0123	19.3
0013	5.9
3210	3.7
0122	3.2
0000	2.7
0012	2.7
0223	2.1
1233	2.1
3123	2.1

A strategy is indicated by a four digit code. Code 0123, for example, indicates the strategy where player i chooses: $c_i^0 = 0$, $c_i^1 = 1$, $c_i^2 = 2$ and $c_i^3 = 3$. We only include strategies played by at least 2 percent of the players in T0.

Table 4.14: Most frequently chosen contribution profiles in T0. High-understanding players

Contribution profile c_i	**Percentage**
0123	18.9
0013	10
0000	4.4
0012	4.4
0122	3.3
0333	3.3
1233	3.3
3333	3.3
0022	2.2
0112	2.2
0323	2.2
3123	2.2
3223	2.2

A strategy is indicated by a four digit code. Code 0123, for example, indicates the strategy where player i chooses: $c_i^0 = 0$, $c_i^1 = 1$, $c_i^2 = 2$ and $c_i^3 = 3$. We restrict the analysis to players who have made 2 mistakes or less in the initial understanding questions. We only include strategies played by at least 2 percent of the high-understanding players.

Table 4.15: Most frequently chosen expectation profiles in T0. All players

Expectation profile α_i	**Percentage**
0123	13.4
0013	4.8
0012	4.3
0112	2.7
0212	2.1
0213	2.1
3122	2.1
3123	2.1
3210	2.1
3223	2.1

An expectation profile is indicated by a four digit code. Code 0123, for example, indicates the expectation profile where player i chooses: $\alpha_i^0 = 0$, $\alpha_i^1 = 1$, $\alpha_i^2 = 2$ and $\alpha_i^3 = 3$. We only include expectation profiles chosen by at least 2 percent of players.

Table 4.16: Probit regression: robustness of tobit assumption

	$c_i^z > 0$	$c_i^z > 0$	$c_i^z = 3$	$c_i^z = 3$
	(1)	(2)	(3)	(4)
Spoke average	.450 (.065)***	.566 (.081)***	.263 (.059)***	.375 (.059)***
T1	.172 (.125)	.151 (.166)	-.082 (.138)	-.347 (.210)*
T1*Spoke average	.034 (.075)	-.007 (.097)	-.047 (.065)	-.012 (.075)
Const.	-.016 (.106)	-.265 (.145)*	-1.052 (.120)***	-1.237 (.170)***
Obs.	3060	1496	3060	1496
Cluster N	98	97	98	97
Pseudo R^2	.125	.163	.034	.088
Log-likelihood	-1456.069	-742.678	-1583.694	-667.209
High understanding		✔		✔

Probit regression. The dependent variable is dummy variable indicated on the top of each column. Columns 2 and 4 restrict the analysis to players who have made 2 mistakes or less in the initial understanding questions. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parentheses.

Table 4.17: Linear probability model: match between c_i^z and α_i^z

	(1)	(2)	(3)	(4)
High Rate of Return	.019 (.036)	.005 (.037)	.024 (.049)	.023 (.051)
Const.	.596 (.044)***	.520 (.129)***	.607 (.071)***	.583 (.204)***
Obs.	1152	1036	592	532
Cluster N	74	73	66	62
Controls		✔		✔
High Understanding			✔	✔

OLS regression.The dependent variable is a dummy variable that takes a value of 1 if $c_i^z = \alpha_i^z$. 'High rate of return' is a dummy for whether the session value of r is $\frac{4}{5}$. The sample includes all players who have received message neutral 2 in treatments T1*neutral*, T1*positive*, T1*negative*. Columns 3 and 4 restrict the analysis to players who have made 2 mistakes or less in the initial understanding questions. Columns 2 and 4 include controls for the players' age, area of land owned, area of land cultivated, number of contacts in real information networks, self-reported oneness with the group, and dummies for having completed secondary education, for being Hindu, and for belonging to a non backward caste. All regressions include dummies for whether spoke average z is equal to 1, 2, or 3 and for whether treatment is T1*positive* or T1*negative*. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level.

Table 4.18: The effect of a high rate of return on the level of contributions and expectations

	Contributions c_i^z	Contributions c_i^z	Expectations α_i^z	Expectations α_i^z	Group expectations $\bar{\alpha}_i^z$
	(1)	(2)	(3)	(4)	(5)
Spoke average	.678 (.058)***	.850 (.076)***	.748 (.073)***	1.149 (.087)***	1.498 (.181)***
High Rate of Return	.112 (.150)	.193 (.177)	.244 (.138)*	.221 (.185)	.183 (.320)
Const.	.341 (.153)**	-.214 (.185)	.131 (.162)	-.668 (.203)***	
Obs.	1152	592	1152	592	296
Cluster N	74	74	66	66	74
Pseudo R^2	.055	.092	.055	.128	.234
Log-likelihood	-1707.425	-835.488	-1703.895	-799.434	-236.215
Controls					
High Understanding		✔		✔	

The dependent variable is indicated on top of each column. Columns 1-4 report estimates from a tobit regression, with an upper limit of 3 and a lower limit of 0. The sample includes all players who have received neutral message 2 in treatments T1*neutral*, T1*positive*, T1*negative*. Columns 3 and 4 restrict the analysis to players who have made 2 mistakes or less in the initial understanding questions. Column 5 reports the results of an ordered logit regression over session-level expectation averages $\bar{\alpha}^z$. 'High rate of return' is a dummy for whether the session value of r is $\frac{4}{5}$. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors are reported in parenthesis. Standard errors are corrected for clustering at session level in columns 1-4.

Table 4.19: Linear probability model: match between c_i^z and $\bar{\alpha}^z$. Moderating factors

	(1)	(2)	(3)	(4)	(5)	(6)
High Rate of Return	.023 (.042)	.100 (.057)*	.023 (.043)	.104 (.057)*	.024 (.043)	.111 (.058)*
Average session degree	.006 (.028)	.027 (.039)				
Degree			-.001 (.011)	.007 (.016)		
Oneness					.003 (.013)	.008 (.018)
Const.	.333 (.189)*	.176 (.264)	.375 (.084)***	.301 (.121)**	.372 (.083)***	.310 (.107)***
Obs.	1144	584	1144	584	1132	588
Cluster N	73	65	73	65	74	66
High understanding		✔		✔		✔

OLS regression.The dependent variable is a dummy variable that takes a value of 1 if $c_i^z = \bar{\alpha^z}$. 'Degree' is a variable that reports the number of other farmers in the session that the player knows. 'Average degree' is the session-level average of degree. 'Oneness' is the self-reported value of oneness. The sample includes all players who have received message neutral 2 in treatments T1*neutral*, T1*positive*, T1*negative*. In columns 2, 4 and 6 the analysis is restricted to players who have made 2 mistakes or less in the initial understanding questions. All regressions include dummies for whether spoke average z is equal to 1, 2, or 3, for whether treatment is T1*positive* or T1*negative* and for whether $r = \frac{4}{5}$. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parenthesis.

Table 4.20: Linear probability model: match between c_i^z and $\bar{\alpha}^z$. Heterogeneous treatment effects

	(1)	(2)	(3)	(4)	(5)	(6)
High Rate of Return	-.471 (.311)	-.908 (.401)**	.071 (.136)	.100 (.188)	.172 (.141)	.426 (.170)**
Average session degree	-.033 (.038)	-.050 (.051)				
High Rate of Return * average degree	.082 (.052)	.167 (.067)**				
Degree			.003 (.012)	.006 (.022)		
High Rate of Return * degree			-.008 (.022)	.0005 (.032)		
Oneness					.015 (.017)	.039 (.021)*
High Rate of Return * oneness					-.025 (.024)	-.053 (.031)*
Const.	.564 (.238)**	.631 (.325)*	.352 (.089)***	.302 (.145)**	.301 (.099)***	.125 (.105)
Obs.	1144	584	1144	584	1132	588
Cluster N	73	65	73	65	74	66
High understanding		✔		✔		✔

OLS regression.The dependent variable is a dummy variable that takes a value of 1 if $c_i^z = \bar{\alpha^z}$. 'Degree' is a variable that reports the number of other farmers in the session that the player knows. 'Average degree' is the session-level average of degree. 'Oneness' is the self-reported value of oneness.The sample includes all players who have received message neutral 2 in treatments T1*neutral*, T1*positive*, T1*negative*. In columns 2, 4 and 6 the analysis is restricted to players who have made 2 mistakes or less in the initial understanding questions. All regressions include dummies for whether spoke average z is equal to 1, 2, or 3, for whether treatment is T1*positive* or T1*negative* and for whether $r = \frac{4}{5}$. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parenthesis.

Table 4.21: Ordered logit regression over expectations α_i^z

	(1)	(2)	(3)	(4)
T1*positive*	-.173 (.145)	-.167 (.158)	-.073 (.179)	-.050 (.211)
T1*negative*	-.387 (.169)**	-.446 (.182)**	-.208 (.251)	-.229 (.303)
Obs.	1160	1040	544	488
Cluster N	74	74	66	64
Pseudo R^2	.003	.009	.0008	.006
Log-likelihood	-1598.649	-1425.325	-742.76	-663.543
Controls		✔		✔
High Understanding			✔	✔

Ordered logit regression. The dependent variable is expectation α_i^z. The sample includes all subjects in T1*neutral*, T1*positive* and T1*negative* who have received message 1 (neutral in T1*neutral*, positive in T1*positive* and negative in T1*negative*). Columns 3 and 4 restrict the analysis to players who have made 2 mistakes or less in the initial understanding questions. Columns 2 and 4 include controls for the players' age, area of land owned, area of land cultivated, number of contacts in real information networks, self-reported oneness with the group, and dummies for having completed secondary education, for being Hindu, and for belonging to a non backward caste. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parenthesis.

Table 4.22: Ordered logit regression over expectations α_i^z, by spoke average

	α_i^0	α_i^1	α_i^2	α_i^3
	(1)	(2)	(3)	(4)
T1*positive*	-.183	-.301	-.172	-.007
	(.281)	(.277)	(.272)	(.247)
T1*negative*	-.326	-.589	-.554	-.217
	(.313)	(.326)*	(.282)**	(.253)
Obs.	290	290	290	290
Cluster N	74	74	74	74
Pseudo R^2	.002	.006	.006	.001
Log-likelihood	-324.074	-381.098	-346.641	-362.719

Ordered logit regression. The dependent variable is expectation α_i^z. The sample includes all subjects in T1*neutral*, T1*positive* and T1*negative* who have received message 1 (neutral in T1*neutral*, positive in T1*positive* and negative in T1*negative*). Columns 3 and 4 restrict the analysis to players who have made 2 mistakes or less in the initial understanding questions. Columns 2 and 4 include controls for the players' age, area of land owned, area of land cultivated, number of contacts in real information networks, self-reported oneness with the group, and dummies for having completed secondary education, for being Hindu, and for belonging to a non backward caste. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parenthesis.

Table 4.23: Tobit regression model (4.8) over contributions c_i^z

	(1)	(2)	(3)	(4)
T1*positive*	-.044	-.061	-.013	-.144
	(.192)	(.203)	(.228)	(.248)
T1*negative*	.090	.116	.237	.171
	(.160)	(.175)	(.214)	(.250)
Const.	1.413	1.479	1.089	1.566
	(.109)***	(.387)***	(.167)***	(.718)**
Obs.	1152	1036	592	532
Cluster N	74	73	66	62
Pseudo R^2	.0003	.003	.001	.01
Log likelihood	-1805.97	-1620.165	-919.444	-819.762
Controls		✔		✔
High Understanding			✔	✔

Tobit regression, with an upper limit of 3 and a lower limit of 0. The dependent variable is contribution decision c_i^z. The sample includes all subjects in T1*neutral*, T1*positive* and T1*negative* who have received message 2 neutral 2. Columns 3 and 4 restrict the analysis to players who have made 2 mistakes or less in the initial understanding questions. Columns 2 and 4 include controls for the players' age, area of land owned, area of land cultivated, number of contacts in real information networks, self-reported oneness with the group, and dummies for having completed secondary education, for being Hindu, and for belonging to a non backward caste. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parentheses.

Table 4.24: Tobit regression model (4.8) over contributions c_i^z, by spoke average

	c_i^0 (1)	c_i^1 (2)	c_i^2 (3)	c_i^3 (4)
T1*positive*	.207 (.667)	.026 (.233)	-.078 (.158)	-.301 (.314)
T1*negative*	.043 (.588)	.231 (.203)	-.005 (.168)	-.011 (.297)
Const.	-.566 (.470)	1.192 (.133)***	1.609 (.115)***	2.643 (.219)***
Obs.	288	288	288	288
Cluster N	74	74	74	74
Pseudo R^2	.0002	.002	.0004	.002
Log-likelihood	-366.395	-438.4	-397.619	-412.407

Tobit regression, with an upper limit of 3 and a lower limit of 0. The dependent variable is contribution decision c_i^z. The sample subjects in T1*neutral*, T1*positive* and T1*negative* who have received neutral message 2. Columns 1-4 analyse separately the values of c_i^1, c_i^2, c_i^3, and c_i^4, respectively. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parenthesis.

Table 4.25: Tobit regression model (4.9) over contributions c_i^z

	(1)	(2)	(3)	(4)
T1*positive*	.147 (.333)	.057 (.342)	-.107 (.430)	-.445 (.447)
T1*negative*	.134 (.285)	.170 (.292)	-.087 (.375)	-.240 (.353)
Spoke average	.734 (.102)***	.703 (.100)***	.758 (.148)***	.660 (.148)***
T1*positive* * spoke average	-.132 (.145)	-.086 (.146)	.055 (.210)	.185 (.219)
T1*negative* * spoke average	-.037 (.135)	-.044 (.135)	.201 (.180)	.257 (.180)
Const.	.303 (.205)	.408 (.385)	-.045 (.297)	.576 (.717)
Obs.	1152	1036	592	532
Cluster N	74	73	66	62
Pseudo R^2	.055	.054	.094	.094
Log-likelihood	-1706.765	-1537.37	-834.164	-749.935
Controls		✔		✔
High Understanding			✔	✔

Tobit regression, with an upper limit of 3 and a lower limit of 0. The dependent variable is contribution decision c_i^z. The sample includes all subjects in T1*neutral*, T1*positive* and T1*negative* who have received neutral message 2. Columns 3 and 4 restrict the analysis to players who have made 2 mistakes or less in the initial understanding questions. Columns 2 and 4 include controls for the players' age, area of land owned, area of land cultivated, number of contacts in real information networks, self-reported oneness with the group, and dummies for having completed secondary education, for being Hindu, and for belonging to a non backward caste. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parenthesis.

Table 4.26: Tobit regression over contributions c_i^z in T1*neutral* and T0

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T1*neutral*	.059 (.158)	.076 (.167)	-.180 (.270)	-.077 (.303)	.204 (.302)	.288 (.314)	.139 (.402)	.381 (.408)
Spoke average					.755 (.113)***	.807 (.111)***	1.035 (.115)***	1.100 (.116)***
T1*neutral* * spoke average					-.096 (.145)	-.139 (.149)	-.200 (.163)	-.289 (.165)*
Const.	1.468 (.133)***	1.109 (.428)***	1.342 (.246)***	.822 (.671)	.323 (.230)	-.135 (.496)	-.231 (.333)	-.856 (.745)
Obs.	1524	1332	700	616	1524	1332	700	616
Cluster N	49	49	48	47	49	49	48	47
Pseudo R^2	.00006	.002	.0005	.007	.047	.052	.074	.084
Log-likelihood	-2386.04	-2079.048	-1085.009	-948.697	-2274.715	-1975.799	-1005.009	-874.526
Controls		✔		✔		✔		✔
High understanding			✔	✔			✔	✔

Tobit regression, with an upper limit of 3 and a lower limit of 0. The dependent variable is contribution c_i^z. The sample includes all subjects in T1*neutral* and T0. Columns 3, 4, 7 and 8 restrict the analysis to players who have made 2 mistakes or less in the initial understanding questions. Columns 2 , 4, 6 and 8 include controls for the players' age, area of land owned, area of land cultivated, number of contacts in real information networks, self-reported oneness with the group, and dummies for having completed secondary education, for being Hindu, and for belonging to a non backward caste. Confidence: *** ⟷ 99%, ** ⟷ 95%, * ⟷ 90%. Standard errors corrected for clustering at session level are reported in parenthesis.

Instructions

Good evening! Thank you for joining us. Tonight you will be taking part in an experiment that will last about 1,30 h.

In this experiment, you have the chance to win real money. How much money you win depends on the way you and the other participants play the game. Your decisions will also affect how much other people win.

During the experiment you are not allowed to talk. Also, please switch off your mobile phones. You will have a chance to ask questions after I finish reading each part of this explanation. But while the experiment is on, we ask you please not to talk. You are given a paper copy of the instructions I am about to read. You can refer to that at any point in the game.

You are of course allowed to leave the game whenever you want. However, payments will be made only at the end of the whole game. Hence if you choose to leave, you will only collect the show-up fee and will not receive any further payment.

Before we start, please fill in your name, surname, and telephone number in the appropriate boxes at the beginning of your game sheet.

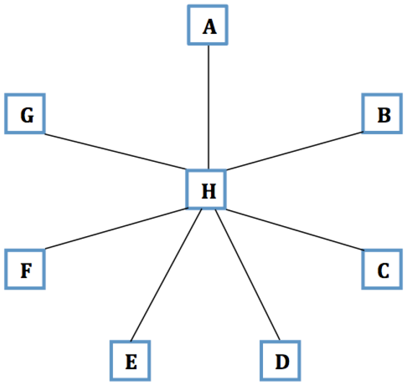
The game

During the game, a letter of the alphabet will identify each player. Later on in the game, this letter will be assigned by drawing a card from an urn.

For the purposes of the game, you will be friends with some of the other participants. We have assigned friendships in advance. Figure 1 below shows this. For example, if you are player A, you are friend with player H. If you are player B, you are friend with player H. If you are player H, you are friend with all the other players: A, B, C, D, E, F, and G.

Notice that in this game H is friend with all other seven players. Each of the other seven players, on the other hand, has only one friend: player H.

Figure 1 The Friendship Map



You have 150 INR at the start of this game, divided in three notes of 50 INR each. You have to choose how many notes you want to put in a pot called the "common pot". Money in the common pot generates a payoff for both you and your friends. That is, if you put a sum of money in the common pot, that sum of money will be collected at the end of the game both by yourself and by your friends. Notice that the money is not divided among yourselves: the same sum becomes available for everyone.

You will also collect money from the common pot of your friends.

At the end of the game, your winnings will be the sum of the money you kept for yourself, plus the money that is available in your common pot, plus the money that is available in the common pot of your friends.

Putting money into the common pot has a cost. For every 50 INR you put in the common pot, 20 will be taken away. So if you put 50 INR in the common pot, both yourself and your friends will collect from this pot 30 INR at the end of the game.

***

Let's discuss a few examples. Imagine you are player A. You have only one friend: player H.

Example 1. Imagine neither A nor H put any money in their common pot. The common pots remain empty. In this case, at the end of the game, player A's winnings will be the sum of money he has kept for himself (150), plus the amount in his common pot (0), plus the amount in the common pot of H (0). The total winnings of A will be 150.

Player H's winnings will be the sum of money he has kept for himself (150), plus the amount in his common pot (0), plus the amount in the common pot of A (0), plus the amount in the common pots of his other friends. The total winnings of H will be 150 plus the amount in the common pots of his other friends.

Example 2. Imagine now that both A and H put all of their money in the common pot. Thus there are 90 INR in the common pot of A and 90 INR in the common pot of H. At the end of the game, player A's winnings will be the sum of money he has kept for himself (0), plus the amount in his common pot (90), plus the amount in the common pot of H (90). The total winnings of A will be 180.

Player H's winnings will be the sum of money he has kept for himself (0), plus the amount in his common pot (90), plus the amount in the common pot of A (90), plus the amount in the common pots of his other friends. The total winnings of H will be 180 plus the amount in the common pots of his other friends.

Example 3. Imagine now that A puts 2 notes in the common pot and keeps 1 note for himself, while H keeps all of his money for himself. Player A's winnings will be the sum of money he has kept for himself (50), plus the amount in his common pot (60), plus the amount in the common pot of H (0). So the total winnings of A will be 110.

Player H's winnings will be the sum of money he has kept for himself (150), plus the amount in his common pot (0), plus the amount in the common pot of A (60), plus the amount in the common pots of his other friends. The total winnings of H will be 210 plus the amount in the common pots of his other friends.

***

You can think of this game in terms of deciding whether to invest in an activity with private benefits, or investing in an activity whose benefits spill-over to your friends. For example, experimenting with new agricultural technologies and sharing the new knowledge you generate with your friends.

You can now ask any questions you may have.

Decisions

You will have to make your decisions before you pick the ID card.

First, you will decide how many notes you would like to put in the common pot if you pick the ID of a player with only one friend (A,B,C,D,E,F,G).

Second, you will decide how many notes you would like to put in the common pot if you pick the ID of player H, who is friend with the other seven players.

Player H can choose how many notes he would like to put in his common pot on the basis of the average number of notes that his friends have put in their common pots.

This means you will decide how many notes you would like to put in your common pot if you pick the ID of H and your friends have put on average 0, 1, 2 or 3 notes in their common pot.

All numbers will be rounded. So if the average is 0.5 or below, it counts as 0. If the average is between 0.5 and 1.5, it counts as 1. If the average is between 1.5 and 2.5, it counts as 2. If the average is between 2.5 and 3, it counts as 3.

Let me give you two examples about a fictitious player called "Sachin". When his friends put on average 0 notes in their common pot, Sachin chooses to put 0 notes in his common pot. When they put 1 note in, Sachin chooses to put 0 notes in the pot. When they put 2 notes in, Sachin chooses to put 1 note in the common pot. When they put 3 notes in, Sachin chooses to put 3 notes in the common pot.

Suppose the average number of notes the friends put in the common pot is 3. In this case, if Sachin draws ID letter H, he will put 3 notes in his common pot.

Please ask any questions you may have.

Trial round

We will now play a trial round of the game. Each player will take a decision and we will calculate the resulting prizes in front of you. This will make sure you understand fully how prizes are calculated. However, no prizes will actually be distributed for the winnings in the trial round.

You will have to take one decision for the case in which you pick the ID of player A-G, and four decisions for the case in which you pick the ID of player H. For each one of these, please put the desired number of notes in the pot in front of you. Then circle the number corresponding to your decision on page X of the game sheet.

Notice that, as you have three notes, you can decide to put into the common pot 0, 1, 2, or 3 notes.

Now, please pick from this urn an ID card. Notice that this ID card will only be valid for the trial round. You will have to pick again for the main game.

I will now show you how the prizes are calculated.

Please ask any questions you may have.

I would now like to make sure you have understood the rules of the game. Please look at your game sheet and answer questions U.1 to U.6.

Task 1

We will now play the actual game. As in the trial round, you will have to take your decisions before you pick the ID card.

Please go to page X of the game sheet. For your decision, you should both physically put the notes in the common pot and circle the right answer in the game sheet.

Task 2

As in the trial round, you are all about to make four decisions about how many notes to put in the pot if you are player H. Before we proceed to these decisions, I will ask you some expectation questions. Please tell me how many notes on average you expect the other players to put in the pot for each of these 4 decisions.

Notice this question is about what decisions the other players will take. So I will not ask you to put the notes in the pot in front of you. Just circle the number on the space provided on page X of the game sheet.

You can now ask any questions you have.

I would like to give to each one of you a message, which is contained in a closed envelope. There are two messages. Four envelopes contain the first message, and four envelopes contain the second message.

You will now draw an envelope from this urn. Please open it and read the message. Do not share this message with any other player.

I will now collect the game sheets and record the expectations of each player.

This way, I will know what your friends expect you to do if you pick player H.

I will write the average of these expectations on the board.

This information gives you a clear idea about what your friends are expecting. It is up to you whether you want to use this information or not.

Task 3

Now you will have to take four decisions for the case in which you pick the ID of player H. Please go to page X of the game sheet

ID Letter

I will now assign ID letters. Please pick one card from this urn. The letter printed on the card will be your ID letter.

Remember that you are not allowed to inform anyone in the room of which letter you have picked. If you are found doing so, you will not be allowed to continue the game and will not collect any winnings.

Before I distribute the prizes, I would like to ask you some final questions about yourself and about your knowledge of other participants in the game. We guarantee that this data will only be used for research purposes and that your information will not be revealed to any third party.

Conclusion

In this thesis I study how networks are formed and I analyse the strategies that well-connected individuals adopt in public good games played over a network.

A central finding is that network-formation decisions can be predicted using simple models of strategic behaviour: individuals target their links to players in advantageous positions in the network and take into account partners' valuable characteristics, for example, low relative risk aversion. However, optimal strategies are chosen too infrequently to guarantee the overall efficiency of the network. The link-formation game we play with Indian farmers directly documents a large shortfall in network efficiency. Similarly, in the experiment in Ethiopia, only 47 percent of decisions are optimally targeted.

A second key finding is that other-regarding preferences and group identity play an important role in networks. Indian farmers frequently choose links that maximise overall welfare or the payoff of the least-connected players in the network. Disclosure of group identity is associated with an increase in in-group links. When contributing to public goods, well-connected players largely follow the principle of conditional cooperation. Surprisingly, however, in a game where they have no possibility to influence their personal payoff, Ethiopian individuals do not link with the least-connected, worse-off players significantly more often than with other players in the network.

I conclude by discussing three possible ideas for future research that build on these findings.

The convergence of beliefs in a network

Theoretical models suggest that, in the long run, individuals who are embedded in a network and observe each others' expectations or actions hold the same beliefs (Bala and Goyal, 1998; DeMarzo et al., 2003; Gallo, 2014). Under specific conditions, for example the absence of individuals with excessive influence, long-run beliefs correspond to true states of the world.

Documenting the extent of belief convergence in real networks is important to fully determine the scope of policies to support information diffusion. The results of this thesis point to the inefficiency of the networks formed by farmers in India and urban dwellers in Ethiopia. It remains an open question whether inefficient link-formation decisions carry over to real networks and whether these are so severe as to compromise the convergence of beliefs, for example, because they partition the network in components. Do beliefs converge to the true state of the world in real networks?

I propose the design of a new study of Indian farmers to answer this question. This is based on an observational and an experimental component. First, I will sample a number of villages and select twenty farmers in each village. I will collect *observational information* about (i) their connections with each other and (ii) farmers's beliefs about technologies, prices and available government programmes. With observational data I can study whether individuals in these small communities hold similar beliefs about features of the world that are relevant for the economic decisions they take. However, I cannot test whether beliefs are correct, as I largely do not observe the true state of the world, nor can I observe the process of belief formation.

To address the last two limitations of the observational data, I will carry out a *framed field experiment* (Harrison and List, 2004). In the first phase of the experiment, which lasts five days, ten of the twenty farmers sampled in a village receive a signal about a hidden number.[34] The signal is communicated at a randomly chosen date and time during the five-days window, through a short mobile phone call. The hidden number can take values from one to twenty. A signal is also a number from one to twenty. Farmers are informed that the hidden number always corresponds to the signal that is

[34]Each farmer is informed about the identity of the other 19 farmers who participate in the experiment.

observed most frequently in their village. Other numbers are observed with different frequencies, with no meaningful pattern. Hence, the farmers' task is to aggregate all signals in their village and determine which signal occurs with the highest frequency.

After the first phase is over, on a given day, all farmers in a village are asked about the hidden number and about the individuals with whom they shared signals, again through individual phone calls. Those who report the right number win a significant monetary prize. The day when the phone calls are made is determined through a lottery, creating variation in the time allowed for information aggregation.[35]

The data from this experiment will allow me to study both whether individuals hold the same beliefs, and whether individuals hold the correct belief about the hidden number. I can calculate the frequency of villages where beliefs have converged to the same number (within some acceptable bounds) and test whether this is significantly different from the probability of observing such convergence when players report random numbers. In addition, I can study whether individuals report the right number by testing the joint hypothesis that $\alpha = 0$ and $\beta = 1$ in the following model for person i in village v:

$$\text{reported number}_{iv} = \alpha + \beta \text{hidden number}_{v} + u_{iv}$$

Information on the network position and individual characteristics of the farmers who receive the signals will allow me to investigate mechanisms that potentially prevent convergence to the true number. For example, I can test whether the probability of learning decreases when socially central individuals receive a wrong signal, or when individuals who receive the correct signal are socially isolated. Moreover, trends in the correlation of beliefs over time can be used to make out-of-sample predictions. Finally,

[35]To ensure the continuation probability of the game does not decline with time, every day in this second phase we roll a fair die: on the first day when number six shows up, farmers are called and the game comes to an end.

I can explore the external validity of these results by studying whether convergence of beliefs in the experiment is more frequent in villages where individuals' beliefs about technologies, prices and policies are more correlated.

Group categorisation and the convergence of beliefs

In chapter one I show that Indian farmers respond to the disclosure of players' group membership by creating more in-group links. This result obtains for groups that have been arbitrarily formed and in an experiment that maintains personal anonymity. Arguably, however, individuals feel a stronger sense of affiliation to real groups. Moreover, when their personal identity is known, individuals who restrict interaction to the in-group may improve their reputation and derive economic benefits. The results in chapter one hence give a lower-bound estimate of the effect of group identity on information networks.

In the area of India that I visited for the experiment, caste divisions in social interaction are highly visible. This personal observation is consistent with recent work by Indian scholars (Guha, 2008) and with network analysis. Using dyadic data from a village in the state of Karnataka, Jackson (2014) shows that the frequency of links among households belonging to the same caste is more than ten times higher than the frequency of links among households of different castes.

The experimental design I proposed above can be extended to study how to maximise the convergence of beliefs to the true state of the world in a community characterised by group divisions. The results of this study would be of clear policy relevance.

In particular, I plan two adjustments of the basic design. First, sampling has to take into account the spatial segregation of castes. My observation suggests that a village often consists of different hamlets, where individuals of the same caste reside. The classic study of Beteille and Srinivas (1964) makes a similar point for a different area of India. In the sampling procedure deployed for the experiments with Indian farmers in this thesis, I bring enumerators to a single starting point in the village and then ask them to walk in randomly chosen directions, selecting all farmers they encounter. Given geographical segregation, this procedure is likely to produce samples that disproportionately belong to the same caste. I thus plan to set different starting points in the same village, chosen to capture different caste clusters.

Second, I can manipulate the distribution of signals. These can be either (i) assigned to castes in proportion to group size (treatment 1) or (ii) assigned exclusively to a single caste (treatment 2). If information exchange across castes is weak, treatment 2 may have a stronger impact on the overall diffusion of correct beliefs, albeit at the cost of high inequality across groups. If communication between individuals of different castes is unproblematic, on the other hand, the relative effectiveness of the two treatments depends on a tradeoff: individuals in dense clusters find it easier to aggregate the information to obtain the right answer, but this answer has to travel a long distance to diffuse to other parts of the network (Centola, 2010).

The cost of having few connections

Without payoff incentives, Ethiopian individuals do not link with the least-connected players in a job-referral network more often than with other players. Consistently with this finding, I observe among the same subjects a heavily skewed distribution of connections in residential-block job-contact networks. Is skewness an artefact of the fact that connections to individuals outside of the residential block are not captured in my data? And, if not, what is the real economic cost of having few connections?

I will tackle these questions through the analysis of an ongoing intervention that assists young jobseekers in Addis Ababa, Ethiopia.[36] Treated individuals are offered a

[36] This study is joint work with Girum Abebe (Ethiopian Development Research Institute), Marcel Fafchamps (Stanford University), Paolo Falco (OECD), Simon Franklin (University of Oxford) and Simon Quinn (University of Oxford).

weekly transport subsidy to travel to the centre of town in search for employment. A pilot version of the program has reported positive effects on the intensity and duration of job search, and on the probability of being permanently employed by the end of the study (Franklin, 2014). In the current, scaled-up version of the project we interview untreated individuals in treated areas to study the spillover effects of the policy. Treated friends may be better sources of information and referrals and may be more effective at motivating untreated participants to search with higher intensity.

Two features make this research project particularly suitable to answer my questions. First, we collect detailed data on participants' networks. Jobseekers are sampled from selected, small geographical units. Within a unit, we record information describing the relationship and interactions participants have with each other. We further ask questions about social interactions outside of the neighbourhood and interview some of the friends that do not reside in the same area. This allows us to construct a rich picture of job-contact networks and crucially, to identify individuals who are, relative to others, poorly connected.

Second, the proportion of individuals who receive the subsidy is randomly varied across localities. Designs of this type have recently been introduced for the study of interpersonal influence under the name of 'randomised saturation' (Crépon et al., 2013; Baird et al., 2014). A comparison of the outcomes of untreated job seekers across areas assigned to different levels of saturation allows causal identification of the spillover effects of the policy. Of particular interest is the interaction between spillover effects and integration in job-contact networks: do poorly-connected participants benefit less than others from the positive spillovers?

We will answer this question and quantify the economic cost of being poorly connected with a model of the following form:

$$y_{iv} = \alpha + \beta_1 \text{saturation}_v + \beta_2 \text{connections}_{iv} + \beta_3 \left(\text{saturation}_v * \text{connections}_{iv}\right) + u_{iv}$$

On the left hand side we will include variables that capture the decisions of participants, for example, the intensity of job search, and their labour market outcomes. The variable 'saturation$_v$' indicates the proportion of treated individuals in area v, while 'connections$_{iv}$' records the z-score of the number of connections in local job-contact networks. The coefficient β_1 thus captures the spillover effect of the policy for individuals with the average number of connections. β_3 measures the effect of a standard deviation increase in the number of connections on the intensity of the spillover.

References

Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *The American Economic Review*, 101(2):470–492.

Akerlof, G. and Kranton, R. (2010). *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton University Press.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753.

Anderson, S. (2011). Caste as an impediment to trade. *American Economic Journal: Applied Economics*, 3(1):239–63.

Andreoni, J. and Miller, J. (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2):737–753.

Aron, A., Aron, E. N., and Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, 63(4):596.

Ashraf, N., Bandiera, O., and Jack, K. (2012). No margin, no mission? A field experiment on incentives for pro-social tasks. Working paper.

Ashraf, N., Karlan, D., and Yin, W. (2006). Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *The Quarterly Journal of Economics*, 121(2):635–672.

Attanasio, O., Barr, A., Cardenas, J. C., Genicot, G., and Meghir, C. (2012). Risk pooling, risk preferences, and social networks. *American Economic Journal: Applied Economics*, 4(2):134–67.

Bacharach, M., Guerra, G., and Zizzo, D. J. (2007). The self-fulfilling property of trust: An experimental study. *Theory and Decision*, 63(4):349–388.

Baird, S., Bohren, A., McIntosh, C., and Özler, B. (2014). Designing experiments to measure spillover effects. Working paper.

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

Bala, V. and Goyal, S. (2000). A noncooperative model of network formation. *Econometrica*, 68(5):1181–1230.

Bandiera, O. and Rasul, I. (2006). Social networks and technology adoption in northern Mozambique. *Economic Journal*, 116(514):869–902.

Banerjee, A., Breza, E., Chandrasekhar, A., and Jackson, M. O. (2012). Come play with me: Experimental evidence of information diffusion about rival goods. Working paper.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144).

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

Barr, A. and Genicot, G. (2008). Risk sharing, commitment, and information: An experimental analysis. *Journal of the European Economic Association*, 6(6):1151–1185.

Barr, A. and Stein, M. (2008). Status and egalitarianism in traditional communities: An analysis of funeral attendance in six zimbabwean villages. Csae working paper 2008-26.

Battigalli, P. and Dufwenberg, M. (2007). Guilt in games. *The American Economic Review*, 97(2):170–176.

Baumeister, R. F., Stillwell, A. M., and Heatherton, T. F. (1994). Guilt: an interpersonal approach. *Psychological Bulletin*, 115(2):243.

Bayer, P., Ross, S. L., and Topa, G. (2008). Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy*, 116(6):1150–1196.

Beaman, L. and Magruder, J. (2012). Who gets the job referral? Evidence from a social networks experiment. *American Economic Review*, 102(7):3574–93.

Beaman, L. A. (2012). Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the U.S. *Review of Economic Studies*, 79(1):128–161.

Bellemare, C., Sebald, A., and Strobel, M. (2010). Measuring the willingness to pay to avoid guilt: Estimation using equilibrium and stated belief models. IZA Discussion Papers 4803, Institute for the Study of Labor (IZA).

Belot, M. and Fafchamps, M. (2014). Good Samaritans and the Market: Experimental Evidence on Other-Regarding Preferences. Working paper.

Benjamin, D. J., Choi, J. J., and Strickland, A. J. (2010). Social identity and preferences. *American Economic Review*, 100(4):1913–28.

BenYishay, A. and Mobarak, A. M. (2014). Social learning and communication. NBER Working Paper 20139.

Berg, E., Ghatak, M., Manjula, R., Rajasekhar, D., and Roy, S. (2013). Motivating knowledge agents: Can incentive pay overcome social distance? STICERD - Economic Organisation and Public Policy Discussion Papers Series 42.

Berninghaus, S., Ehrhart, K.-M., and Ott, M. (2006). A network experiment in continuous time: The influence of link costs. *Experimental Economics*, 9(3):237–251.

Beteille, A. and Srinivas, M. (1964). Networks in Indian social structure. *Man*, 64:165–168.

Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press.

Binswager, H. P. (1980). Attitudes toward risk: Experimental measurement in rural India. *American Journal of Agricultural Economics*, 62(3):395–407.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, pages 1170–1182.

Bourlès, R., Bramoullé, Y., et al. (2013). Altruism in networks. Working paper.

Bramoulle, Y. and Kranton, R. (2007). Public goods in networks. *Journal of Economic Theory*, 135(1):478–494.

Brandts, J. and Charness, G. (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics*, 14(3):375–398.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues*, 55(3):429–444.

Breza, E., Chandrasekhar, A. G., and Larreguy, H. (2014). Social structure and institutional design: evidence from a lab experiment in the field. NBER Working Paper 20309.

Bulte, E., Kontoleon, A., List, J., Turley, T., and Voors, M. (2012). When economics meets hierarchy: A field experiment on the workings of the invisible hand. Working paper.

Burt, R. S., Jannotta, J. E., and Mahoney, J. T. (1998). Personality correlates of structural holes. *Social Networks*, 20(1):63–87.

Callander, S. and Plott, C. R. (2005). Principles of network development and evolution: An experimental study. *Journal of Public Economics*, 89(8):1469–1495.

Calvo-Armengol, A. (2004). Job contact networks. *Journal of Economic Theory*, 115(1):191–206.

Camerer, C. F. (2003). *Behavioral Game Theory*. Russell Sage Foundation, New York, NY.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427.

Cardella, E. (2012). Exploiting the guilt aversion of others (do agents do it and is it effective? Working paper.

Cardenas, J. C. and Carpenter, J. (2008). Behavioural development economics: Lessons from field labs in the developing world. *The Journal of Development Studies*, 44(3):311–338.

Caria, S. and Falco, P. (2014). Do employers trust workers too little? An experimental study of trust in the labour market. Working paper.

Carpenter, J., Kariv, S., and Schotter, A. (2012). Network architecture, cooperation and punishment in public good experiments. *Review of Economic Design*, 16(2):93–118.

Casaburi, L., Kremer, M., Mullainathan, S., and Ramrattan, R. (2014). Harnessing ICT to increase agricultural production: Evidence from Kenya. Working paper.

Centola, D. (2010). The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–1197.

Centola, D. (2011). An experimental study of homophily in the adoption of health behavior. *Science*, 334(6060):1269–1272.

Chandrasekhar, A. and Lewis, R. (2011). Econometrics of sampled networks. Working paper.

Charness, G. and Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, 74(6):1579–1601.

Charness, G., Feri, F., Melandez-Jimenez, M. A., and Sutter, M. (2014). Experimental games on networks: Underpinnings of behavior and equilibrium selection. *Econometrica*, 82(5):1615–1670.

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869.

Charness, G., Rigotti, L., and Rustichini, A. (2007). Individual behavior and group membership. *American Economic Review*, 97(4):1340–1352.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, 14(1):47–83.

Chen, Y. and Li, S. X. (2009). Group identity and social preferences. *The American Economic Review*, 99(1):431–457.

Cialdini, R. B., Brown, S. L., Lewis, B. P., Luce, C., and Neuberg, S. L. (1997). Reinterpreting the empathy–altruism relationship: When one into one equals oneness. *Journal of Personality and Social Psychology*, 73(3):481–494.

Cingano, F. and Rosolia, A. (2012). People I know: Job search and social networks. *Journal of Labor Economics*, 30(2):291 – 332.

Cole, S. A. and Fernando, A. N. (2012). The value of advice: Evidence from mobile phone-based agricultural extension. Harvard Business School Working Papers 13-047.

Comola, M. (2010). The network structure of mutual support links: Evidence from rural tanzania. PSE Working Paper 2008–73.

Comola, M. and Fafchamps, M. (2013). Testing unilateral and bilateral link formation. *The Economic Journal*, 144(579):954–976.

Conley, T. G. and Udry, C. R. (2004). Social networks in Ghana. Economic Growth Center Working Paper no. 888.

Conley, T. G. and Udry, C. R. (2010). Learning about a new technology: pineapple in Ghana. *American Economic Review*, 100(1):35–69.

Conte, A., Cagno, D. D., and Sciubba, E. (2009). Strategies in social network formation. Birkbeck Working Papers in Economics and Finance 0905.

Crawford, V., Costa-Gomes, M., and Iriberri, N. (2013). Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature*, 51(1):5–62.

Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., and Zamora, P. (2013). Do labor market policies have displacement effects? Evidence from a clustered randomized experiment. *The Quarterly Journal of Economics*, 128(2):531–580.

Currarini, S., Jackson, M. O., and Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045.

Currarini, S. and Menge, F. (2012). Identity, homophily and in-group bias. Working Papers 2012.37, Fondazione Eni Enrico Mattei.

de Brauw, A. and Eozenou, P. (2014). Measuring risk attitudes among Mozambican farmers. *Journal of Development Economics*, 111(0):61 – 74. Special Issue: Imbalances in Economic Development.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. Mc Graw-Hill, New York.

Delavande, A., Giné, X., and McKenzie, D. (2011). Measuring subjective expectations in developing countries: A critical review and new evidence. *Journal of Development Economics*, 94(2):151–163.

DeMarzo, P. M., Vayanos, D., and Zwiebel, J. (2003). Persuasion bias, social influence, and unidimensional opinions. *The Quarterly Journal of Economics*, 118(3):909–968.

Dessi, R., Gallo, E., and Goyal, S. (2012). Network cognition. Working paper.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Duflo, E., Kremer, M., and Robinson, J. (2011). Nudging farmers to use fertilizer: Theory and experimental evidence from kenya. *American Economic Review*, 101(6):2350–90.

Dufwenberg, M., Gächter, S., and Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior*, 73(2):459–478.

Dufwenberg, M. and Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. *Games and Economic Behavior*, 30(2):163–182.

Dunning, T. and Nilekani, J. (2013). Ethnic quotas and political mobilization: Caste, parties, and distribution in Indian village councils. *American Political Science Review*, 107:35–56.

Eckel, C. C. and Grossman, P. J. (2005). Managing diversity by creating team identity. *Journal of Economic Behavior & Organization*, 58(3):371–392.

Ellingsen, T., Johannesson, M., Tjøtta, S., and Torsvik, G. (2010). Testing guilt aversion. *Games and Economic Behavior*, 68(1):95–107.

Fafchamps, M. (2004). Market institutions in sub-Saharan Africa: Theory and evidence. *MIT Press Books*.

Fafchamps, M. (2011). Risk sharing between households. Handbook of Social Economics. North-Holland.

Fafchamps, M. and Gubert, F. (2007). The formation of risk sharing networks. *Journal of Development Economics*, 83(2):326–350.

Falco, P. (2014). Does risk matter for occupational choices? Experimental evidence from an African labour market. *Labour Economics*, 28:96–109.

Falk and Kosfeld (2012). It's all about connections: Evidence on network formation. *Review of Network Economics*, 11(3):1–36.

Falk, A. and Szech, N. (2013). Morals and markets. *Science*, 340(6133):707–711.

Fatas, E., Meléndez-Jiménez, M. A., and Solaz, H. (2010). An experimental analysis of team production in networks. *Experimental Economics*, 13(4):399–411.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Finan, F. and Schechter, L. (2012). Vote-buying and reciprocity. *Econometrica*, 80(2):863–881.

Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3):397–404.

Fischbacher, U., Gächter, S., and Quercia, S. (2012). The behavioral validity of the strategy method in public good experiments. *Journal of Economic Psychology*, 33(4):897–913.

Foster, A. D. and Rosenzweig, M. R. (1995). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy*, 103(6):1176–1209.

Franklin, S. (2014). Location, search costs and youth unemployment: The impact of a randomized transport subsidy in urban ethiopia. Working paper.

Gächter, S. (2006). Conditional cooperation: Behavioral regularities from the lab and the field and their policy implications. Discussion Paper 2006-03, The Centre for Decision Research and Experimental Economics, School of Economics, University of Nottingham.

Gächter, S. and Renner, E. (2010). The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics*, 13(3):364–377.

Gaertner, W. and Schokkaert, E. (2011). *Empirical Social Choice*. Cambridge University Press.

Galeotti, A. and Goyal, S. (2010). The law of the few. *The American Economic Review*, 100(4):1468–1492.

Gallo, E. (2014). Social learning by chit-chat. *Journal of Economic Theory*, 153(C):313–343.

Goeree, J. K., McConnell, M. A., Mitchell, T., Tromp, T., and Yariv, L. (2010). The $\frac{1}{d}$ law of giving. *American Economic Journal: Microeconomics*, 2(1):183–203.

Goeree, J. K., Riedl, A., and Ule, A. (2009). In search of stars: Network formation among heterogeneous agents. *Games and Economic Behavior*, 67(2):445–466.

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

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. (2007). *Connections: An Introduction to the Economics of Networks*. Princeton University Press.

Granovetter, M. (1995). *Getting a Job. A Study of Contacts and Careers*. The University of Chicago Press, 2nd edition.

Grossman, G. and Baldassarri, D. (2012). The impact of elections on cooperation: Evidence from a lab-in-the-Field experiment in Uganda. *American Journal of Political Science*, 56(4):964–985.

Guha, R. (2008). *India After Gandhi: The History of the World's Largest Democracy*. Ecco book. HarperCollins.

Harrison, G. W. and List, J. A. (2004). Field Experiments. *Journal of Economic Literature*, 42(4):1009–1055.

Hedstrom, P., Kolm, A., and Aberg, Y. (2003). Social interactions and unemployment. Working Paper Series 2003:15, IFAU - Institute for Labour Market Policy Evaluation.

Hitsch, G. J., Hortacsu, A., and Ariely, D. (2006). What makes you click? Mate preferences and matching outcomes in online dating. MIT Sloan Research Paper No. 4603-06.

Hoff, K. and Pandey, P. (2006). Discrimination, social identity, and durable inequalities. *American Economic Review*, 96(2):206–211.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *The American Economic Review*, 92(5):1644–1655.

Ioannides, Y. M. and Loury, L. D. (2004). Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature*, 42(4):1056–1093.

J-PAL (2013). J-PAL Youth Initiative review paper. Technical report.

Jack, K. B. and Recalde, M. (2013). Local leadership and the voluntary provision of public goods: Field evidence from Bolivia . Working paper.

Jackson, M. O. (2010). *Social and Economic Networks*. Princeton University Press.

Jackson, M. O. (2014). Networks in the understanding of economic behaviors. *The Journal of Economic Perspectives*, 28(4):3–22.

Jackson, M. O., Rodriguez-Barraquer, T., and Tan, X. (2012). Social capital and social quilts: Network patterns of favor exchange. *American Economic Review*, 102(5):1857–97.

Jackson, M. O. and Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44–74.

Jakiela, P. (2011). Social preferences and fairness norms as informal institutions: Experimental evidence. *American Economic Review*, 101(3):509–13.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*, 125(2):515–548.

Karlan, D., Mobius, M., Rosenblat, T., and Szeidl, A. (2009). Trust and social collateral. *The Quarterly Journal of Economics*, 124(3):1307–1361.

Karlan, D., Osei, R. D., Osei-Akoto, I., and Udry, C. (2012). Agricultural decisions after relaxing credit and risk constraints. NBER Working Paper 18463.

Kirchsteiger, G., Mantovani, M., Mauleon, A., and Vannetelbosch, V. (2011). Myopic or Farsighted? An Experiment on Network Formation. CEPR Discussion Papers 8263.

Kondylis, F., Mueller, V., and Zhu, S. J. (2014). Seeing is believing? Evidence from an extension network experiment. World Bank Policy Research Working Paper Series 7000.

Kosfeld, M. (2003). Network Experiments. IEW Working Paper.

Kranton, R., Pease, M., Sanders, S., and Huetter, S. (2012). Identity, group conflict, and social preferences. Working paper.

Kremer, M. and Miguel, E. (2007). The illusion of sustainability. *The Quarterly journal of economics*, 122(3):1007–1065.

Krishnan, P. and Patnam, M. (2012). Neighbors and extension agents in Ethiopia: Who matters more for technology diffusion? IGC Working Paper.

Krishnan, P. and Sciubba, E. (2009). Links and architecture in village networks. *Economic Journal*, 119(537):917–949.

Ledyard, J. O. (1995). Public goods. a survey of experimental research. Handbook of Experimental Economics. Princeton, NJ: Princeton University Press.

Leider, S., Mobius, M. M., Rosenblat, T., and Do, Q.-A. (2009). Directed altruism and enforced reciprocity in social networks. *The Quarterly Journal of Economics*, 124(4):1815–1851.

Levitt, S. and List, J. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2):153–174.

Ligon, E. and Schechter, L. (2012). Motives for sharing in social networks. *Journal of Development Economics*, 99(1):13–26.

List, J. A. (2004). Testing neoclassical competitive theory in multilateral decentralized markets. *Journal of Political Economy*, 112(5):1131–1156.

Liu, E. M. (2013). Time to change what to sow: Risk preferences and technology adoption decisions of cotton farmers in China. *Review of Economics and Statistics*, 95(4):1386–1403.

Magruder, J. R. (2010). Intergenerational networks, unemployment, and persistent inequality in South Africa. *American Economic Journal: Applied Economics*, 2(1):62–85.

Mano, Y., Yamanops, T., Suzuky, A., and Matsumotol, T. (2010). Local and personal networks in employment and the development of the labour market: Evidence from the cut flowers industry in Ethiopia. GRIPS Policy research Centre Discussion Paper 10-29.

Manski, C. F. (2004). Measuring expectations. *Econometrica*, 72(5):1329–1376.

Manski, C. F. and Neri, C. (2013). First- and second-order subjective expectations in strategic decision-making: Experimental evidence. *Games and Economic Behavior*, 81(C):232–254.

Marmaros, D. and Sacerdote, B. (2006). How do friendships form? *The Quarterly Journal of Economics*, 121(1):79–119.

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.

Miller Moya, L. M., Barr, A., Burns, J., and Shaw, I. (2011). Individual notions of distributive justice and relative economic status. DFAEII Working Papers 2011-03.

Mueller, A. (2012). In the public eye. norms of distributive justice and sharing behaviour under asymmetric information: Evidence from rural malawi. Working paper.

Munshi, K. (2004). Social learning in a heterogeneous population: Technology diffusion in the Indian Green Revolution. *Journal of Development Economics*, 73(1):185–213.

Nikiforakis, N., Noussair, C. N., and Wilkening, T. (2012). Normative conflict and feuds: The limits of self-enforcement. *Journal of Public Economics*, 96(9):797–807.

Oster, E. and Thornton, R. (2012). Determinants of technology adoption: Peer effects in menstrual cup take-up. *Journal of the European Economic Association*, 10(6):1263–1293.

Reuben, E. and Riedl, A. (2009). Public goods provision and sanctioning in privileged groups. *Journal of Conflict Resolution*, 53(1):72–93.

Reuben, E. and Riedl, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, 77(1):122–137.

Reuben, E., Sapienza, P., and Zingales, L. (2009). Is mistrust self-fulfilling? *Economics Letters*, 104(2):89–91.

Rosenkranz, S. and Weitzel, U. (2012). Network structure and strategic investments: An experimental analysis. *Games and Economic Behavior*, 75(2):898–920.

Santos, P. and Barrett, C. B. (2010). Identity, interest and information search in a dynamic rural economy. *World Development*, 38(12):1788–1796.

Schlag, K. H., Tremewan, J., and van der Weele, J. (2014). A penny for your thoughts: A survey of methods for eliciting beliefs. Vienna economics papers.

Schmutte, I. M. (2015). Job referral networks and the determination of earnings in local labor markets. *Journal of Labor Economics*, 33(1):1–32.

Serneels, P. (2007). The nature of unemployment among young men in urban Ethiopia. *Review of Development Economics*, 11(1):170–186.

Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology*. CUP Archive.

Tanaka, T., Camerer, C. F., and Nguyen, Q. (2010). Risk and time preferences: Linking experimental and household survey data from Vietnam. *The American Economic Review*, 100(1):557–571.

Tarbush, B. and Teytelboym, A. (2014). Friending. Working paper.

Topa, G. (2001). Social interactions, local spillovers and unemployment. *Review of Economic Studies*, 68(2):261–95.

Topa, G. (2011). Labor markets and referrals. volume 1 of *Handbook of Social Economics*, pages 1193 – 1221. North-Holland.

Tufano, F., Gächter, S., and Starmer, C. (2012). The power of social relations for coordination: The magic of 'oneness'. Conference presentation, TIBER XI, Tillburg University.

Van Dolder, D. and Buskens, V. (2009). Social motives in network formation: An experimental study. *International Conference on Game Theory for Networks. Papers and Proceedings*, (1):593–602.

Vanberg, C. (2008). Why do people keep their promises? an experimental test of two explanations. *Econometrica*, 76(6):1467–1480.

Vasilaky, K. and Leonard, K. (2013). As good as the networks They keep? Improving farmers' social networks via randomized information exchange in rural Uganda. Working paper.

Viceisza, A. (2012). Treating the field as a lab: A basic guide to conducting economics experiments for policymaking. Food Security in Practice series 7, International Food Policy Research Institute (IFPRI).

Vigneri, M. (2005). *Trade Liberalisation and Agricultural Performance: Contrasting Macro and Micro Evidence from Ghana*. PhD thesis, University of Oxford.

Wooldridge, J. (2012). *Introductory Econometrics: A Modern Approach*. Cengage Learning.

Yaari, M. and Bar-Hillel, M. (1984). On dividing justly. *Social Choice and Welfare*, 1(1):1–22.

Yamagishi, T. and Kiyonari, T. (2000). The group as the container of generalized reciprocity. *Social Psychology Quarterly*, 63(2):116–132.

Zeitlin, A. Market institutions and productivity: Evidence from Ghanaian cocoa, type=Working Paper year=2006. Technical report.

Zizzo, D. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1):75–98.