Workshop / *Atelier*

Learning for adopting

Apprendre pour adopter

Wednesday-Thursday, June 1st-2nd, 2016
Mercredi-Jeudi 1er-2 Juin 2016

Disconnected

Experimental Evidence on Network Efficiency among Indian Farmers

by

A. Stefano Caria, Centre for the Study of African Economies, Department of Economics, University of Oxford

Marcel Fafchamps, Stanford University

Disconnected

Experimental Evidence on Network Efficiency among Indian Farmers[*]

A. Stefano Caria[†] Marcel Fafchamps[‡]

May 2016

Abstract

In an artefactual field experiment in rural India, we find that subjects form inefficient networks and lose 35 percent of payoffs as a result. the game is designed so that the efficient network can be reached if all players choose strategies consistent with selfish rationality and with a concern for overall welfare. These strategies are chosen frequently. However, in 20 percent of the cases, subjects target the 'most popular' player in the network and this causes large efficiency losses. When information about group membership is disclosed, there are more connections between subjects of the same group, but networks do not become significantly less efficient. Networks play an important role in the diffusion of information and in the sharing of risk and favours. If they are inefficiently structured, there is scope for policies that shape social connections.

1 Introduction

Social networks play an influential role in economic life. Individuals who share a social connection often exchange information about technologies and opportunities (Goyal, 2007; Conley and Udry, 2010; Banerjee et al., 2013), share risk and support (Fafchamps and Gubert, 2007; Munshi, 2011; Ambrus et al., 2014), and help each other set goals (Field et al., 2015; Breza and Chandrasekhar, 2015). These behaviours have tangible economic returns and can be exploited by policy makers to increase the impacts of social interventions (Feigenberg et al., 2013; BenYishay and Mobarak, 2015). The magnitude of these returns, however, is influenced by the structure of the network. For example, information diffuses slowly in networks where most connections occur within small groups of similar individuals (Golub and Jackson, 2012). There is very limited evidence on whether, in practice, individuals form networks that maximise the benefits of social interaction and on what factors may prevent network efficiency.

In this paper, we report the results of an experiment designed to study network efficiency. Assessing efficiency using observational data would be challenging, as the researcher has to quantify the costs and benefits of each connection. We thus study efficiency in a game where players sequentially create connections with each other. The benefits of a link are proportional to the number of players that are 'reached' directly or indirectly through this link. Links are created unilaterally and the benefits flow to only one of the two players, as in the model of Bala and Goyal (2000). Each player can only form one link. In this simple game, the efficient network is a cycle where every player is connected, directly or indirectly, to everybody else.

We find that players systematically form inefficient networks. On average, a player is connected with only 3.2 of the 5 other players in the game and payoffs are about 35 percent lower than those in the efficient network. This is surprising because we design the experiment so that players converge to the efficient network if they choose links to maximise their own payoffs (in the first experimental condition) or to maximise the payoff of others (in the second experimental condition). The efficient network is also an equilibrium of the game and determines equal payoffs for all players.

A relatively small group of players that adopt an inefficient strategy is mostly responsible for aggregate inefficiency. This finding highlights that networks are fragile, in the sense that small groups that deviate from efficient strategies can compromise overall welfare. Pooling all decisions together, we find that players choose links consistent with payoff maximisation about half of the times. They also often connect to the least well-off player in the network. Lastly, in 20 percent of the cases, links target the 'most popular' player – the player with the largest number of direct connections at

that point in the game. Simulation analysis shows that this strategy is responsible for the largest efficiency losses. The efficiency loss would be reduced to 5 percent if links to the most popular player were re-wired according to the strategy that maximises payoffs.

We also provide evidence on a key factor that is believed to affect network structure: group identity. Previous research has highlighted how group categorisation generates in-group favoritism (Tajfel, 1981; Brewer, 1999; Akerlof and Kranton, 2010) and influences behaviour in strategic environments (Yamagishi and Kiyonari, 2000; Charness et al., 2007). Empirical studies on networks have also documented a tendency of socially similar individuals to interact with each other with disproportionate frequency (McPherson et al., 2001; Currarini et al., 2009; Golub and Jackson, 2012). To investigate the role of group identity, we randomly assign players to groups at the beginning of the experiment. We then vary whether players are informed of the group affiliation of others during the network formation game. Restricting links to in-group partners would result in large efficiency losses in our game.

We find that when we disclose information about group membership, individuals form more links within their group, but the resulting networks are not less efficient. These effects can occur simultaneously if players (i) create connections on the basis of the network position of the partner, and (ii) whenever there are both in-group and out-group partners with the desired network position, players prefer to connect with an in-group partner. We present evidence consistent with this mechanism.

To enhance external validity, we run the experiment in a population that heavily relies on social networks for the diffusion of information: farmers from the Indian state of Maharashtra. With the many social identities based on caste, religion and class, India also offers an ideal setting to study the effects of group membership (Beteille and Srinivas, 1964; Guha, 2008; Hoff and Pandey, 2006). Recent evidence on information agents in rural communities shows that social distance decreases the probability that information is shared between individuals (Berg et al., 2013).

We contribute to the literature that studies the formation of networks (Jackson and Wolinsky, 1996; Bala and Goyal, 2000). One strand of this literature has used dyadic data from rural areas of developing countries to explore what factors determine 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; Comola and Fafchamps, 2013). These networks are often characterised by homophily (McPherson et al., 2001; Rogers, 2003). Another strand of the literature has studied network formation in the lab, exploring 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 re-

sponses or far-sighted strategies (Callander and Plott, 2005; Conte et al., 2009; Kirchsteiger et al., 2011). In a related experiment, Belot and Fafchamps (2012) compare unilateral partnership formation decisions to dictator game allocations with equivalent payoff consequences. All of these experiments use western subjects, typically university students. 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 affects the diffusion of information and behaviour.

Falk and Kosfeld (2012) – who study a game of unilateral, one-way-flow link formation based on Bala and Goyal (2000) – is the experiment most closely related to ours. The two designs differ because of a number of features that we introduce to limit coordination and computation problems: in our experiment links are added to the network one at a time and players are allowed only one link, so that the only cost of a connection is the opportunity cost of not forming another connection. Falk and Kosfeld (2012) find that efficient networks are achieved in about half of the periods of the game. By comparison, in the last period of our game, the efficient network is reached in less than 10 percent of the sessions. The difference in efficiency may arise from the fact that information networks play a larger role in the lives of the subjects in our experiment. These subjects may hence have developed rules of thumb to deal with complicated network formation problems. Targeting the most popular player in the network may be an effective rule of thumb in games where the efficient network is a star. In many games, however, star networks are not desirable (Bala and Goyal, 1998). Our findings suggest that some subjects inefficiently use the same rule of thumb across different games. This also highlight the importance of studying behaviour in games across different populations.

We also contribute to a growing literature that studies the preferences and strategies of subjects in non-Western countries (Henrich et al., 2010). Using both survey and experimental data, Becker et al. (2015) and l'Haridon et al. (2016) document large differences in social, risk and time preferences across countries. Efficiency in market exchange also varies widely across societies (List, 2004; Bulte et al., 2012; Haushofer and Zurlinden, 2013).

Finally, our study is related to identity economics (Akerlof and Kranton, 2000) and a rich literature in economics and social psychology, referenced above, that studies how group categorisation generates in-group bias and modifies behaviour.

The paper is organised as follows. Section 2 presents the design. Section 3 develops predictions and testable hypotheses. Section 4 describes the data. Section 5 reports the results of the analysis. Section 6 concludes.

2   Design

2.1   The link-formation game

We play a sequential link-formation game with groups of six players.[1] In this game links are valuable because they increase players' chances of winning a monetary prize: one of the six players is randomly drawn at the end of the game to receive the prize and those players who are *directly or indirectly* connected to the winner receive a prize of equivalent value. The prize can be thought of as a valuable non-rival good – like information – which can be shared among socially connected individuals. It is worth 100 Indian Rupees (5.2 USD at purchasing power parity), which is about 85 percent of the average daily wage paid by the NREGA public work program in that area.[2]

At the start the game the network has no links. Players add links sequentially, over two rounds of six turns each. We randomly assign each player to one turn per round.[3] In the first round, players add a link to the network. In the second round, they rewire their existing link if they wish.[4] Links are formed unilaterally – without requiring the consent of the other player. They are recorded on a network map which is drawn on a white board in front of all players and is updated after every turn. A number of design features ensure sequential updating takes place without breaking anonymity.[5] We use the whiteboard and a simple Java application to remind the player with the turn of the total number of (direct and indirect) connections that every other player has at that particular point in the game.[6]

---

[1]All experimental materials can be found here: `https://sites.google.com/site/stefanoacaria/linkformationindia`.

[2]In 2012-2013 the National Rural Employment Guarantee Scheme paid an average daily wage to his workers of 121 INR (`http://nrega.nic.in/netnrega/home.aspx`).
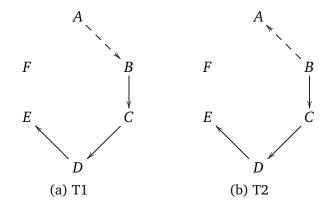
[3]Participants are informed of that the order of play is chosen randomly, but do not know the particular order of play which has been drawn for their session.

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

[5]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.

[6] The counting of connections is done by means of a Java application running on a small laptop operated

Figure I: Links in T1 and T2

(a) T1　　　　　　　　(b) T2

The benefits of a link flow only in one direction. We vary the direction of the flow of benefits across sessions of the game. In treatment 1 sessions (henceforth T1), players form links that let them *reach* other individuals. This means that if player A links with player B, then player A will receive the monetary prize whenever B wins the prize, but not vice versa. In the left panel of figure I this is represented with an arrow pointing from A to B. In the network of figure I, when player A connects to player B, he also reaches farmers C and D, and hence receives the prize if C or D win the prize draw. Thus player A reaches three players in the network. In treatment 2 sessions (henceforth T2), links *let other players reach* the player who proposes the link. If A links with B, then B will receive the monetary prize whenever A wins it, but not vice versa. The right panel of figure I illustrates.[7]

2.2 Group identity

At the beginning of the experiment we assign players to one of two groups. We then disclose information about group membership in randomly chosen sessions of the link formation game. This is cross-cut with T1 and T2, resulting in four experimental conditions, as shown in table I. In the first two conditions, T1no and T2no, individuals have no information about the group affiliation of the other players. In the last two conditions, T1id and T2id, players' group identity is disclosed publicly, by using different symbols to identify players from different groups on the network map.

---

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.

[7]In T2, we report on the whiteboard both the number of individuals who reach a particular player and the number of individuals who are reached by that player.

Table I: Summary of treatments

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

Figure II: Order of activities in the experiment

Group and ID assigned	2	Allocation task	4	Questionnaire
1	Knowledge game	3	Link-formation game	5

We assign group membership randomly – players pick a card from an urn which determines their group affiliation. We do not use art or sport preferences to create the groups because these preferences may correlate with other player characteristics. To strengthen group identity, the groups compete in a game testing agricultural knowledge. Contests of this kind have been used in previous experimental studies to make group affiliation more salient (Eckel and Grossman, 2005). In our game, players answer three questions related to local weeds, pesticides, etc. If all players in a group answer all questions correctly, the group obtains one "point", which is added to those accumulated by the same group in previous sessions. Every player can thus have a strong marginal impact on the group's outcome: if the player fails to answer one question correctly, the whole group fails to gain the point. Players answer the agricultural knowledge questions before the link formation game, but the results are publicly revealed only after all parts of the experiment are completed. Our objective is to make the groups salient from the start of the experiment, while ensuring that we do not disclose information that can influence the other tasks.

We measure whether players are biased in favour of other players of the same group in an incentivised allocation task. In this task, players have to divide 30 Rupees between an in-group and an out-group recipient. Recipients are randomly drawn from the participants in the *following session* of the experiment. This task is played right after the questions on agricultural knowledge. We also ask a number of unincentivised questions about norms and expectations at the end of the experiment. The order of activities is summarised in figure II.

2.3 Instructions

Instructions are framed in terms of the local context. The link-formation game is presented as a game where one farmer will receive a valuable 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 get the information. In T2 the choice is about which other player you want to help in case you get the valuable information. The groups are called the mango and the pineapple group, an implicit reference to the producer groups which farmers typically form in the areas of our study.

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).

2.4 Discussion

Understanding. We test players' understanding before the game starts. We ask subjects in T1no and T1id 8 questions, while we ask subjects in T2no and T2id 7 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 to correct mistakes. Hence the answers give a low bound to the level of understanding of the players in the game. Figure XVIII in the appendix presents the cumulative distribution of mistakes. In both T1 and T2 more than 50 percent of players made at most 1 mistake, and about 80 percent of players made at most 2 mistakes. To further increase understanding, we run a trial round of the link-formation game before the main game is played. We also explore the sensitivity of our results to the level of understanding of the players in the session.

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 network-formation decisions will be targeted towards individuals from whom side payments can be more easily extracted.

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 identity may suggest to players that we expect them to use this information. To minimize 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 identity in the trial round, to avoid making unintended suggestions about how we expect players to use this information.

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.

3.1   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:

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

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.[8] 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 randomly chooses one of these. 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 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).[9] 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:

---

[8]In the appendix we show this formally.

[9]Notice that player i's strategy in T1 also has an impact on the payoffs of the individuals who have a path towards i in g. In future work, we will extend this section to include an the analysis of how other-regarding preferences affect behaviour in T1.

$$u_i = \pi_i + \gamma f(\pi_{-i}) \qquad (2)$$

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

$$u_i = \pi_i + \gamma \sum_{j \in N \backslash i} \pi_j \qquad (3)$$

Utility function 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.[10] We predict that a fraction of players in T2 have other-regarding preferences expressed by 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.*

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:

$$u_i = \pi_i + \gamma \min_{j \in N \backslash i} \pi_j \qquad (4)$$

Utility function 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 e player with the lowest expected payoff. We predict that a fraction of players in T2 have other-regarding preferences expressed by 4 and will thus play according to the following link formation rule:

---

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

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 XIV 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 inequity aversion, a player feels guilt towards players with a lower expected payoff and envy towards players with a higher expected payoff. An inequity 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 inequity aversion any further.

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

Prediction 1. *In T1 players will form links to partners with the maximum reach. In T2 players will 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.

3.2   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.

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

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 III 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.[11] When all players in T2 play according to rule 3, on the other hand, network efficiency is 67 percent. Figure IV 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 XV shows this graphically. We can thus formulate the following prediction on session level efficiency.

---

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

Figure III: Network evolution under rule 1 in T1

(a) Empty network

(b) Turn 1

(c) Turn 2

(d) Turn 3

(e) Turn 4

(f) Turn 5

(g) Turn 6

(h) Turn 7

(i) Turn 11: **The cycle!**

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 IV: 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 value of reach after 12 turns of play for 500 simulated sessions. The vertical line indicates average efficiency achieved by a given rule. The rule used in each set of simulations is indicated below the panel.

Prediction 2. *Network efficiency in T1no is close to 96 percent. In T2no, it is between 96 and 67 percent.*

Figure XVI 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.

3.3   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:

$$u_i = \pi_i + \gamma f(\pi_{-i}) + P_i \qquad (6)$$

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 randomly chosen 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 player with the highest in-reach overall. 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 XVII 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.[12] These considerations motivate a final prediction:

Prediction 3. *Disclosure of group identity generates networks characterised by (i) more in-group links and, depending on the magnitude of c, (ii) lower efficiency.*

3.4 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 ingroup 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*[13] than the other.

---

[12]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.

[13]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.

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

$$\text{link}_{ij,r} = \alpha + \textbf{Network Position}_{j,r}\beta + \textbf{D}_{ij}\gamma + \delta\,\text{round}_r + e_{ij,r} \qquad (7)$$

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. $\text{link}_{ij,r}$ 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 each treatment, these include the variables specified in the link formation rules we propose above. 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.[14] The matrix D_{ij} contains these variables. Furthermore, we control for round specific effects.

Model (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, $E[e_{ij,r}e_{ik,r}] \neq 0$. This inference problem is typical in dyadic regression analysis (Fafchamps and Gubert, 2007). We correct for intra-session

---

[14]For example, see figure III. 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 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 7 is not confounded by correlations with the initial position in the network map.

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).

4 Data

We run the 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.[15] Villages in these sub-dstricts 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.

We select study participants through door-to-door random sampling. On alternating days, we start sampling from the periphery of the village or from the center of the village.[16] 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, and corresponding dyadic dataset with 4,800 dyads.[17] Table II summarises the number of observations we have for each treatment.

At the end of the game, participants compile a short questionnaire. We hence have a small set of covariates.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

---

[15]We exclude from the sampling large towns on the main highway of the district.

[16]We identity the centre by asking village dwellers. This is typically a small square in front of the village temple.

[17]We create the dyadic dataset in the following way. For a player i in round r we create an observation for each possible player j to which player i can connect. We then stack these observations across players and turns. As a result, when a session has six individuals, we have five dyads per turn for each player. When a session has 5 individuals, we have four dyads per turn for each player.

Table II: 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

average, farmers report regularly sharing information about agriculture with 11 other farmers.[18]

Table III: 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,[19] 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.

In tables VI to IX in the appendix, we present a set of regressions that test whether covariates are balanced across treatments. We cannot find any statistically significant imbalance.

---

[18]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 III.

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

Table IV: Summary statistics: session networks

Variable	Obs	Mean	Std. Dev.	Min	Max
How may other players do you know?	438	4.78	.76	0	5
Average days spoken with known players	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.

5 Results

5.1 Network efficiency

We first investigate overall efficiency. Table X summarises treatment-level averages of player reach and the related measure of 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 V below:

Result 1. *Network efficiency in T1no and T2no is lower (higher) than average efficiency under 'efficiency-minded' link-formation (random link-formation) by a significant 31 percentage points (13 percentage points).*

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 the average efficiency which random play would have achieved (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 VI below 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
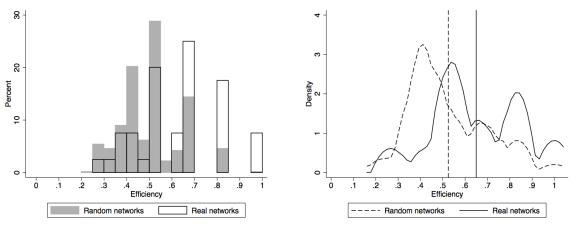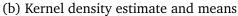
Figure V: Efficiency in no-identity sessions and in random networks
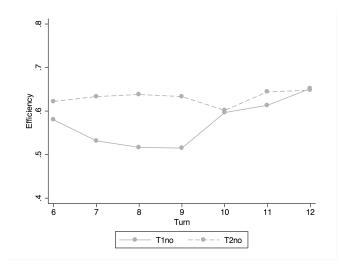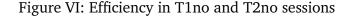

(a) Histogram

(b) Kernel density estimate and means

prediction for T1 is clearly rejected.

Result 2. *Efficiency in T2no sessions is not significantly different from efficiency in T1no sessions.*

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 it was at turn 6. Figure VII illustrates. Falk and Kosfeld (2012), on the other hand, document strong learning dynamics and positive efficiency trends in their experiment.

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

21

Figure VI: Efficiency in T1no and T2no sessions

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

5.2 Individual decisions

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 7. Results are reported in table V.

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.

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 XII 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:

Result 3. *In T1no, links to farmers who have the maximum reach in the network are*

Table V: Dyadic linear probability model (7)

	(1)	(2)	(3)	(4)
Panel a				
max reach$_j$	.132	.130		
	(.001)***	(.001)***		
min in-reach$_j$	.018	.016		
	(.461)	(.627)		
max in-reach$_j$			.111	.120
			(.006)***	(.002)***
min reach$_j$			.073	.066
			(.078)*	(.142)
Const.	.323	.367	.192	.218
	(.002)***	(.002)***	(.002)***	(.028)**
Panel b				
max reach$_j$ = min in-reach$_j$	10.34	10.81		
	(.004)***	(.004)***		
max in-reach$_j$ = min reach$_j$			0.57	1.12
			(.459)	(.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.

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 XII 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.[20] We have however some qualitative evidence in support of it. In the

---

[20]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.

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 3 is in line with prediction 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.

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.[21] Results are presented in panel

---

[21]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:
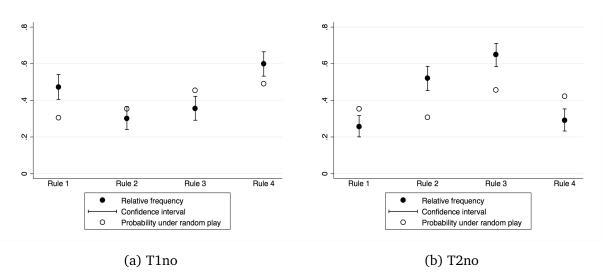
Figure VIII: Relative frequency of decisions consistent with the hypothesised rules

(a) T1no

(b) T2no

(a) of figure XX in the appendix. Overlaps are very frequent. We will hence repeat the analysis for the sample turns where the best response sets are not fully overlapping.

Figure VIII 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 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 XXII and XXIII in the appendix illustrate.

The frequency of 'efficiency-minded' and 'Rawlsian' links is similar in the T1no and T2no treatments. This explains why these two treatments have close 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 XX shows this. The majority of these decisions are consistent with both rules, while about 16 percent of links in both

---

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.

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*[22]

Rule 6. *Choose a link to the player who has chosen you in a previous round* [23]

Regression analysis reported in table XIII 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 XXI and **??** illustrate.
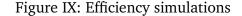
Simulation analysis shows that the largest efficiency gains can be achieved by reducing the proportion of links that are targeted to the 'most popular' player, as opposed to reducing the proportion of 'Rawlsian' links. We simulate a link formation process where 54 percent of decisions are consistent with rule 1, 16 percent with the rule 4 and the remaining 30 percent with the 'most popular' player rule.[24] 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.[25] Figure IX illustrates. In figure XXIV in the appendix we present the same simulation, with a different assumption about the baseline proportion of decisions following the various rules. Qualitatively, results are not affected.
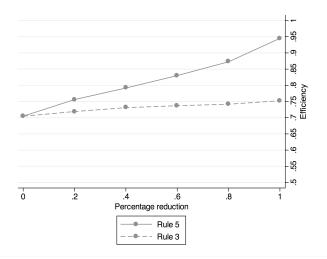
---

[22]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.

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

[24]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.

[25]We also know from the simulations reported in figure XV 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 XV shows that network efficiency in such scenario would be above 90 percent, which corresponds to 20 percentage points gain.

Figure IX: 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.

5.3 Group identity

We next turn to the the effect of group 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.
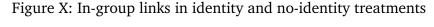
Figure XXV in the appendix shows results from the initial allocation task where a player has to divide a sum of money between an in-group and an out-group partner in the following session. The modal allocation in this task is skewed towards the in-group partner. Overall 54 percent of individuals show 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.[26] 57 percent of players answer yes to this question. Furthermore, about 70 percent of players expect at least 3 of the other players (the majority) to answer year. Table XI and figure XXVI 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.
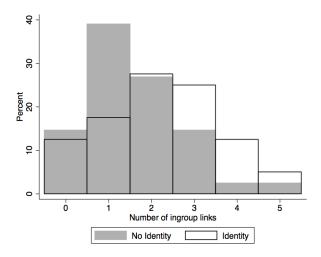
Our main result on the identity treatments is the following:

---

[26]Players in T1no and T2no also answer this question, albeit information about group affiliations was not disclosed in these treatments during the link formation game.

Result 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 3. First, in-group links increase. Figure X 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 X: In-group links in identity and no-identity treatments

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 XI.

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

$$x_{dis} = \alpha + \beta_1 \text{Identity Session}_s + e_{dis} \tag{8}$$

$$x_{dis} = \alpha + \beta_1 \text{Identity Session}_s + \beta_2 z_{is} + \beta_3 \left(z_{is} * \text{Identity Session}_s \right) + e_{dis} \tag{9}$$

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
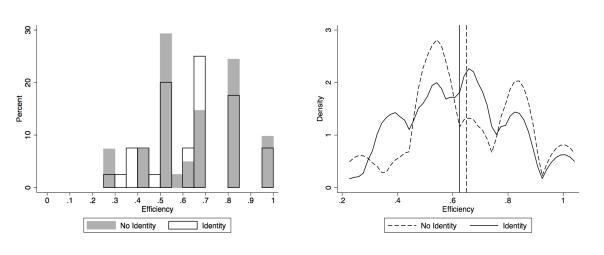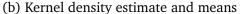
Figure XI: Efficiency in identity and no-identity sessions

(a) Histogram

(b) Kernel density estimate and means

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.[27] In model 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 XII 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 player 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

---

[27] 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.
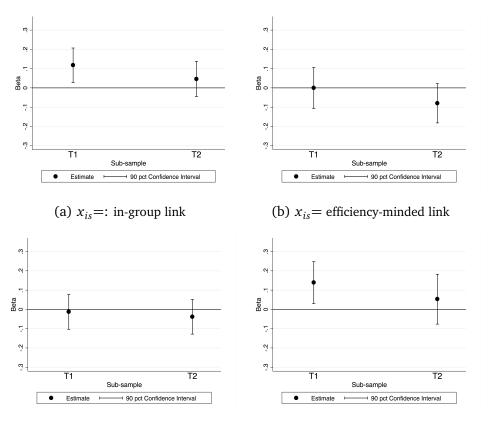
Figure XII: Linear probability model (8): coefficient estimates

(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.

Note. Coefficients estimates from linear probability model 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 XIV, XV, XVI, and XVII.

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 8, with the in-group link dummy as dependent variable. The significant coefficient of 13.8 which we obtain for T1 can be interpreted as the 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 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.[28]

5.4 Additional tests

In this section, we show 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:

$$
\begin{aligned}
x_{dis} =& \alpha + \beta_1 \text{Identity Session}_s \\
& + \beta_2 \left(understanding_{is} * \text{Identity Session=0}_s \right) \\
& + \beta_3 \left(understanding_{is} * \text{Identity Session=1}_s \right) + e_{dis}
\end{aligned}
\tag{10}
$$

$understanding_{is}$ 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 XX and reports results of separate estimations of model 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.

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).

---

[28]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.

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 (Ben Yishay and Mobarak, 2012) 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.

References

Akerlof, G. and R. Kranton (2010). *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton University Press.

Akerlof, G. A. and R. E. Kranton (2000, August). Economics and identity. *The Quarterly Journal of Economics 115*(3), 715–753.

Ambrus, A., M. Mobius, and A. Szeidl (2014). Consumption Risk-Sharing in Social Networks. *The American Economic Review 104*(1), 149–182.

Andreoni, J. and J. Miller (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica 70*(2), pp. 737–753.

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

Bala, V. and S. Goyal (2000, September). A noncooperative model of network formation. *Econometrica 68*(5), 1181–1230.

Banerjee, A., E. Breza, A. Chandrasekhar, and M. O. Jackson (2012). Come Play with Me: Experimental Evidence of Information Diffusion about Rival Goods. Working papers.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013). The diffusion of microfinance. *Science 341*(6144).

Barr, A. and G. Genicot (2008, December). Risk sharing, commitment, and information: An experimental analysis. *Journal of the European Economic Association 6*(6), 1151–1185.

Becker, A., T. J. Dohmen, B. Enke, A. Falk, D. Huffman, U. Sunde, et al. (2015). The Nature and Predictive Power of Preferences: Global Evidence. *CEPR Discussion Papers*.

Belot, M. and M. Fafchamps (2012, August). Good Samaritans and the Market: Experimental Evidence on Other-Regarding Preferences. Discussion Papers 2012001, University of Oxford, Nuffield College.

Ben Yishay, A. and M. Mobarak (2012). Communicating with farmers through social networks. *Working paper*.

BenYishay, A. and A. M. Mobarak (2015). Social learning and incentives for experimentation and communication? *Working Paper*.

Berg, E., M. Ghatak, R. Manjula, D. Rajasekhar, and S. Roy (2013, March). Motivating knowledge agents: Can incentive pay overcome social distance? STICERD - Economic Organisation and Public Policy Discussion Papers Series 42, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

Berninghaus, S., K.-M. Ehrhart, and M. Ott (2006). A network experiment in continuous time: The influence of link costs. *Experimental Economics 9*(3), 237–251.

Beteille, A. and M. Srinivas (1964). 212. networks in indian social structure. *Man 64*, pp. 165–168.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of social issues 55*(3), 429–444.

Breza, E. and A. Chandrasekhar (2015). Social Networks, Reputation and Commitment: Evidence from a Savings Monitors Field Experiment . *NBER Working Paper 21169*.

Bulte, E., A. Kontoleon, J. List, T. Turley, and M. Voors (2012). When economics meets hierarchy: A field experiment on the workings of the invisible hand. Technical report.

Callander, S. and C. R. Plott (2005, August). Principles of network development and evolution: an experimental study. *Journal of Public Economics 89*(8), 1469–1495.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008, August). Bootstrap-Based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics 90*(3), 414–427.

Casaburi, L., M. Kremer, S. Mullainathan, and R. Ramrattan (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.

Charness, G. and M. Rabin (2002, August). Understanding social preferences with simple tests. *The Quarterly Journal of Economics 117*(3), 817–869.

Charness, G., L. Rigotti, and A. Rustichini (2007, September). Individual behavior and group membership. *American Economic Review 97*(4), 1340–1352.

Cole, S. A. and A. N. Fernando (2012, November). The Value of Advice: Evidence from Mobile Phone-Based Agricultural Extension. Harvard Business School Working Papers 13-047, Harvard Business School.

Comola, M. (2010, June). The network structure of mutual support links: Evidence from rural tanzania. Pse working papers, HAL.

Comola, M. and M. Fafchamps (2013). Testing unilateral and bilateral link formation. *The Economic Journal*, n/a–n/a.

Conley, T. G. and C. R. Udry (2010, March). Learning about a new technology: Pineapple in ghana. *American Economic Review 100*(1), 35–69.

Conte, A., D. D. Cagno, and E. Sciubba (2009, June). Strategies in social network formation. Birkbeck Working Papers in Economics and Finance 0905, Birkbeck, Department of Economics, Mathematics & Statistics.

Currarini, S., M. O. Jackson, and P. Pin (2009, 07). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica 77*(4), 1003–1045.

Eckel, C. C. and P. J. Grossman (2005, November). Managing diversity by creating team identity. *Journal of Economic Behavior & Organization 58*(3), 371–392.

Fafchamps, M. and F. Gubert (2007, July). The formation of risk sharing networks. *Journal of Development Economics 83*(2), 326–350.

Falk and Kosfeld (2012, September). It's all about connections: Evidence on network formation. *Review of Network Economics 11*(3), 1–36.

Fehr, E. and K. M. Schmidt (1999, August). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics 114*(3), 817–868.

Feigenberg, B., E. Field, and R. Pande (2013). The Economic Returns to Social Interaction: Experimental Evidence from Microfinance. *The Review of Economic Studies 80*(4), 1459–1483.

Field, E., S. Jayachandran, R. Pande, and N. Rigol (2015). Friendship at Work: Can Peer Effects Catalyze Female Entrepreneurship? Technical report, NBER Working Paper 21093.

Goeree, J. K., A. Riedl, and A. Ule (2009, November). In search of stars: Network formation among heterogeneous agents. *Games and Economic Behavior 67*(2), 445–466.

Golub, B. and M. O. Jackson (2012). How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics 127*(3), 1287–1338.

Goyal, S. (2007). *Connections: An Introduction to the Economics of Networks*. Princeton University Press.

Guha, R. (2008). *India After Gandhi: The History of the World's Largest Democracy*. Ecco book. HarperCollins.

Haushofer, J. and N. Zurlinden (2013). Market convergence and equilibrium in a kenyan informal settlement.

Henrich, J., S. J. Heine, and A. Norenzayan (2010). The Weirdest People in the World? *Behavioral and Brain Sciences 33*(2-3), 61–83.

Hoff, K. and P. Pandey (2006, May). Discrimination, social identity, and durable inequalities. *American Economic Review 96*(2), 206–211.

Jackson, M. O., T. Rodriguez-Barraquer, and X. Tan (2012, August). Social capital and social quilts: Network patterns of favor exchange. *American Economic Review 102*(5), 1857–97.

Jackson, M. O. and A. Wolinsky (1996, October). A strategic model of social and economic networks. *Journal of Economic Theory 71*(1), 44–74.

Karlan, D., M. Mobius, T. Rosenblat, and A. Szeidl (2009, August). Trust and social collateral. *The Quarterly Journal of Economics 124*(3), 1307–1361.

Kirchsteiger, G., M. Mantovani, A. Mauleon, and V. Vannetelbosch (2011, February). Myopic or Farsighted? An Experiment on Network Formation. CEPR Discussion Papers 8263, C.E.P.R. Discussion Papers.

Krishnan, P. and E. Sciubba (2009, 04). Links and architecture in village networks. *Economic Journal 119*(537), 917–949.

l'Haridon, O., F. M. Vieider, et al. (2016). All Over the Map: Heterogeneity of Risk Preferences across Individuals, Prospects, and Countries. *Working Paper*.

List, J. A. (2004, October). Testing Neoclassical Competitive Theory in Multilateral Decentralized Markets. *Journal of Political Economy 112*(5), 1131–1156.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology 27*(1), 415–444.

Munshi, K. (2011). Strength in Numbers: Networks as a Solution to Occupational Traps. *Review of Economic Studies 78*(3), 1069–1101.

Rogers, E. M. (2003). *Diffusion of innovations (5. ed.)*. Free Press.

Santos, P. and C. B. Barrett (2010, December). Identity, interest and information search in a dynamic rural economy. *World Development 38*(12), 1788–1796.

Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology*. CUP Archive.

Van Dolder, D. and V. Buskens (2009, May). Social motives in network formation: An experimental study. *International Conference on Game Theory for Networks. Papers and Proceedings* (1), 593–602.

Vasilaky, K. and K. Leonard (2013, March). As good as the networks they keep?: Improving farmers' social networks via randomized information exchange in rural uganda.

Viceisza, A. (2012). Treating the field as a lab: A basic guide to conducting economics experiments for policymaking. Food security in practice technical guide series 7, International Food Policy Research Institute (IFPRI).

Yaari, M. and M. Bar-Hillel (1984). On dividing justly. *Social Choice and Welfare* (1), 1–22.

Yamagishi, T. and T. Kiyonari (2000). The group as the container of generalized reciprocity. *Social Psychology Quarterly 63*(2), pp. 116–132.

Zizzo, D. (2010, March). Experimenter demand effects in economic experiments. *Experimental Economics 13*(1), 75–98.

7 Appendix

7.1 Formal derivation of the rules

7.1.1 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 $g_i = (g_{i1}, g_{i2}, ..., g_{ii-1}, g_{ii+1}, ..., g_{in})^{29}$, 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.[30] $\Gamma = \Gamma_1 \times \Gamma_2 \times ... \times \Gamma_n$ 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 $\mu_{ji}(g)$ 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.

---

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

[30]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.

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.[31]

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:

$$\pi_j(g) = \frac{1 + \mu_j(g)}{n} \tag{11}$$

7.1.2 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:

$$\max_j \ \pi_i(g + ij) \tag{12}$$

Notice that when i has the turn $N_i(g) = \{\varnothing\}$ 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 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. *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 $\mu_{ji}(g)$. $\square$

7.1.3 Rule 2

Let player i in T1 have a utility function (3). The gain in utility obtained from a g_{ji} link can be summarised as follows:

---

[31]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.

$$u_i(g + ji) - u_i(g) = \pi_i(g + ji) - \pi_i(g) + \gamma \sum_{k \in N \setminus i} \pi_k(g + ji) - \pi_k(g)$$

$$= \gamma \sum_{k \in N \setminus i} \pi_k(g + ji) - \pi_k(g)$$

$$= \gamma f(g + ji) \tag{13}$$

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:

$$\left(\mu_i(g) + 1\right)\left(\mu_{-ji}(g) + 1\right) \tag{14}$$

To derive (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. $\square$

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 $(\mu_i(g) + 1)(\mu_{-j}(g) + 1)$. As $j \notin N_i(g)$, $\mu_{-j}(g) = \mu_{-ji}(g)$. We can express the increase in expected payoff also as: $(\mu_i(g) + 1)(\mu_{-ji}(g) + 1)$.

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 XIII 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.[32]

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 $f(g + ji)$.

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 $f(g + ji)$.
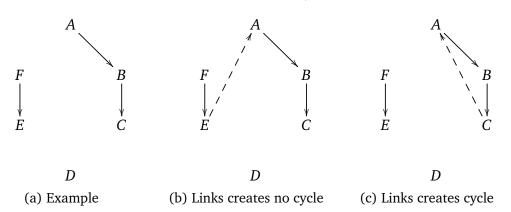
---

[32]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 $i \rightarrow^g j$ 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

$$f(g + ij) = \sum_{n=1}^{\mu_{-ji}+1} n$$
$$= \frac{(\mu_{-ji} + 1)(\mu_{-ji} + 2)}{2} \tag{15}$$

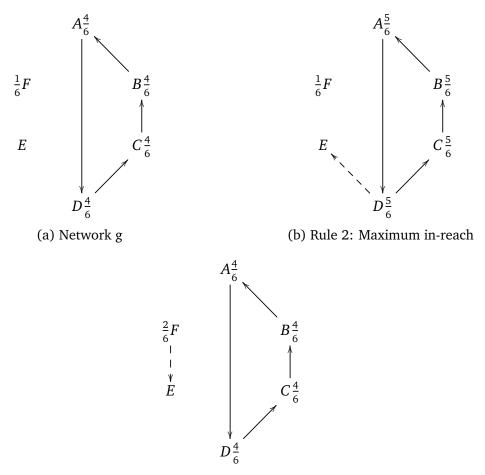If $N_i(g)$ is a tree, this expression would become more complicated.

Figure XIII: Effect of a g_{ji} link

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

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.

Figure XIV: Example of a network

$A\frac{4}{6}$

$\frac{1}{6}F$ $B\frac{4}{6}$

E $C\frac{4}{6}$

$D\frac{4}{6}$

(a) Network g

$A\frac{5}{6}$

$\frac{1}{6}F$ $B\frac{5}{6}$

E $C\frac{5}{6}$

$D\frac{5}{6}$

(b) Rule 2: Maximum in-reach

$A\frac{4}{6}$

$\frac{2}{6}F$ $B\frac{4}{6}$

E $C\frac{4}{6}$

$D\frac{4}{6}$

(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.
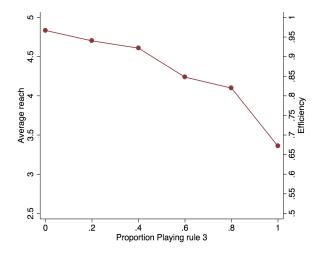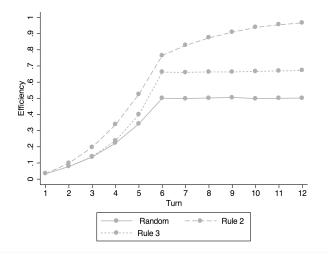
7.2   Figures

Figure XV: Link formation process with mixed rules

Simulation where rule 3 is played with probability p and rule 2 with probability 1-p. 500 simulation for each level of p.

Figure XVI: Simulated time series of average reach

Each rule is simulated 500 times.

Figure XVII: 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). We report results summarising 100 simulation for each of the 4 rules.

Figure XVIII: Cumulative distribution of mistakes in understanding questions

(a) T1

(b) T2

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

(a) Histogram

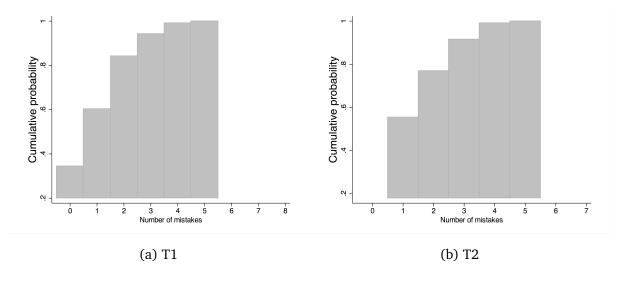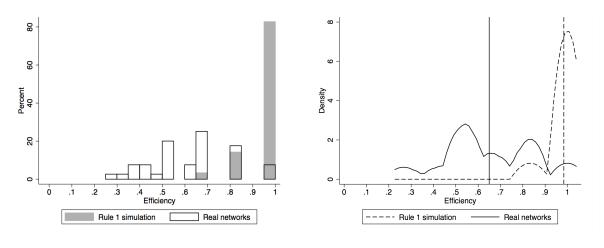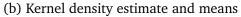(b) Kernel density estimate and means

Figure XX: Overlap in decisions and in choice sets

(a) Best response sets

(b) Decisions

46

Figure XXI: 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 XXII: 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 XXIII: Relative frequency of individuals who play twice consistently with the hypothesised rules

(a) T1

(b) T2

Figure XXIV: 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 XXV: Distribution of coin allocations to in-group partner

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

7.3 Tables

Table VI: Balance test: identity sessions

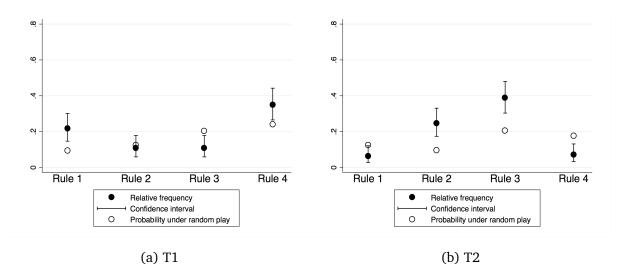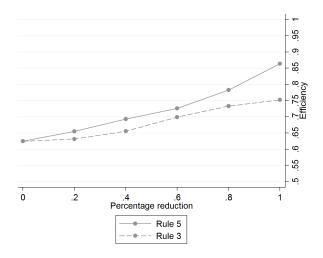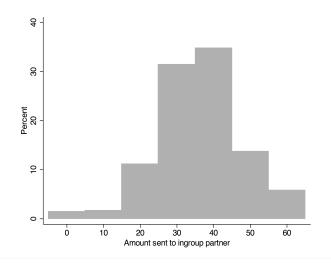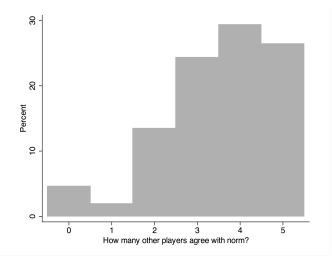|	Age	Edu	UpperCaste	LandOwned	LandCult	NetSize
	(1)	(2)	(3)	(4)	(5)	(6)
Identity	-.194	.029	-.087	.063	.101	-.201
	(1.764)	(.056)	(.067)	(.517)	(.468)	(1.100)
Obs.	479	466	433	475	470	428

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 VII: Balance test: T2 sessions

	Age	Edu	UpperCaste	LandOwned	LandCult	NetSize
	(1)	(2)	(3)	(4)	(5)	(6)
T2	-1.582	-.028	-.052	-.085	-.049	1.293
	(1.761)	(.056)	(.068)	(.514)	(.465)	(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 VIII: Balance test: identity sessions in T1

	Age	Edu	UpperCaste	LandOwned	LandCult	NetSize	Und
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Identity	-2.378	.091	-.040	.077	.141	.111	-.267
	(2.544)	(.081)	(.098)	(.737)	(.657)	(1.102)	(.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 IX: Balance test: identity sessions in T2

	Age	Edu	UpperCaste	LandOwned	LandCult	NetSize	Und
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Identity	1.879	-.033	-.135	.046	.061	-.482	-.224
	(2.400)	(.076)	(.093)	(.733)	(.673)	(1.877)	(.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 X: 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 XII: Dyadic linear probability model (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.

51

Table XI: 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 XIII: Dyadic linear probability model (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 XIV: Linear probability model (8): in-group links

	T1	T2
Identity session	.117	.046
	(.055)**	(.056)
Const.	.282	.268
	(.037)***	(.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 XV: Linear probability model (8): efficiency-minded links

	T1	T2
Identity session	-.0003	-.080
	(.065)	(.062)
Const.	.473	.519
	(.044)***	(.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 XVI: Linear probability model (8): Rawlsian links

	T1	T2
Identity session	-.013	-.038
	(.055)	(.055)
Const.	.600	.649
	(.033)***	(.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 XVII: Linear probability model (8): in-group links
Restricted sample

	T1	T2
Identity session	.139	.053
	(.066)**	(.079)
Const.	.298	.242
	(.048)***	(.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 XVIII: Linear probability model (9): in-group links T1 treatment

	(1)	(2)	(3)	(4)
Identity session$_s$	.095	.133	-.014	.191
	(.073)	(.077)*	(.115)	(.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 XIX: Linear probability model (9): in-group links T2 treatment

	(1)	(2)	(3)	(4)
Identity session$_s$	.014	.001	-.080	.024
	(.080)	(.073)	(.133)	(.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 XX: Linear probability model: understanding in T1

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