Can Farmers Create Efficient Information Networks?

Experimental Evidence from Rural India*

A. Stefano Caria† Marcel Fafchamps‡

August 2014

*We are first of all indebted to Gaurav Puntambekar for his outstanding management of the project. We would like to thank Johannes Abeler, Francesco Cecchi, Megenagna Gashaw Guade, Esteban Ortiz Ospina, Simon Quinn, Bassel Tarbush, Mateusz Ujma and seminar participants at the 2014 Pacific Development Conference in Los Angeles, US, the 2014 conference of the Centre for the Study of African Economies, the 2014 CESifo conference on social economics, and at the CSAE workshop series for useful comments and suggestions. Our gratitude also goes to the the enumerators and supervisors whose contributions made this study possible. We finally would like to thank the study participants for their goodwill and patience. All mistakes remain naturally ours. We acknowledge financial support from the Department for International Development through Phase 2 of the iiG Programme. Stefano Caria further acknowledges financial support from the Economic and Social Research Council, UK.

†DPhil Candidate, Centre for the Study of African Economies, Department of Economics, University of Oxford. Email stefano.caria@economics.ox.ac.uk. Web: https://sites.google.com/site/stefanoacaria

‡Stanford University

Abstract

We run an artefactual field experiment in rural India which tests whether farmers can create efficient networks in a repeated link formation game, and whether group categorization results in homophily and loss of network efficiency. We find that the efficiency of the networks formed in the experiment is significantly lower than the efficiency which could be achieved under selfish, rational play. Many individual decisions are consistent with selfish rationality and with a concern for overall welfare, but the tendency to link with the most “popular" farmer in the network causes large efficiency losses. When information about group membership is disclosed, social networks become more homophilous, but not significantly less efficient. Networks play an important role in the diffusion of innovations in developing countries. If they are inefficiently structured, there is scope for development policies that support diffusion.

1 Introduction

A large literature in development economics argues that information about agricultural innovations diffuses through farmers' social networks [Foster and Rosenzweig, 1995, Munshi, 2004, Bandiera and Rasul, 2006, Conley and Udry, 2010, Krishnan and Patnam, 2012]. However, information sharing does not guarantee that learning about new technologies will be complete [Bala and Goyal, 1998, Jack, 2013]. Networks do not necessarily connect all individuals in a society. Connected individuals may be linked through long average paths, slowing down diffusion. Learning may be a complex contagion that requires contact with multiple infected nodes to propagate [Centola and Macy, 2007]. In other words, networks may be inefficiently structured to allow information to diffuse extensively and fast. Indeed, two recent randomised control trials show that incentives for information agents in rural communities improve the diffusion of information in a cost-effective fashion, suggesting that un-incentivised diffusion was suboptimal [Ben Yishay and Mobarak, 2012, Berg et al., 2013].

Observational assessment of the efficiency of farmers' networks is complicated by several factors. First, a census of all farmers and all links within a village is required.[1] Such data is usually hard to obtain. Furthermore, the costs and benefits of each link have to be quantified. These are often not observed by the field researcher, and are hard to measure. Finally, with observational data alone, it is difficult to attribute inefficiency to the preferences and decision making rules of individuals, or to the constraints individuals face. In this paper, we hence rely on an artifactual laboratory experiment that allows us to quantify the efficiency of small experimental networks, with a great deal of control over the constraints imposed on decision makers [Harrison and List, 2004].

We study whether farmers can form efficient network architectures in a sequential link formation game where players' identities are private knowledge. Our set-up is that of unilateral, one-way flow link formation discussed in Bala and Goyal [2000]: links are formed unilaterally by the player with the turn and the benefits of a link flow only to one of the players. The benefits of a link to a player are proportional to the number of other players in the network that he can access directly or indirectly thanks to this link. In a first treatment the player with the turn can only add links that let him access other players. In a second treatment, only links that let other players access the

[1]For the moment, we leave aside the important issue of links to farmers outside the village.

player with the turn are allowed. In this game, the cycle architecture is efficient[2], a Nash equilibrium, and generates no inequality in payoffs across agents. Selfish agents playing best responses would converge to the cycle architecture after repeated play of the first treatment. Efficiency-minded, other-regarding players would converge to the cycle architecture in the second treatment. By ruling out tradeoffs between efficiency and equilibrium, limiting coordination requirements, and anonymising interaction we give the "best shot" to the possibility of efficient networks emerging in the game. Our main objective is to investigate whether farmers create efficient networks in a setting where this is relatively easy to achieve.

Further, we test whether common knowledge of membership in induced social groups increases the number of in-group links and decreases the efficiency of network architectures. Social differentiation may discourage links that are desirable from an efficiency point of view. Observational research on networks often points to the importance of homophily- the tendency of similar individuals to interact with each other with disproportionate frequency [McPherson et al., 2001, Currarini et al., 2009, Golub and Jackson, 2012] Homophily can be the result of a norm related to group membership [Akerlof and Kranton, 2000]. In our game, consistently restricting links to in-group partners would result in large efficiency losses.

When group identity is private knowledge, we predict that individuals will play simple, intuitive link formation rules leading to high levels of efficiency. In the first treatment, the selfish best response is to choose the player who accesses, directly or indirectly, the highest number of individuals in the network. In the second treatment, other-regarding players who want to maximise the sum of payoffs in the group will target the player who is accessed by the highest number of individuals. Players who instead want to maximise the payoff of the least well-off peer will choose the player who accesses the smallest number of individuals. We derive these predictions from standard models of strategic network formation augmented to include other-regarding preferences [Bala and Goyal, 2000, Charness and Rabin, 2002].

When group identity is common knowledge, we predict individuals will choose in-group links more frequently, possibly at the cost of establishing less efficient networks. Previous research has highlighted how group categorisation generates in-group favoritism [Tajfel, 1981, Brewer, 1999, Akerlof and Kranton, 2010], affects social, risk and time preferences [Benjamin et al., 2010, Chen and Li, 2009, Kranton et al., 2012],

[2] We define the efficient network architecture as the architecture which maximises the sum of players' payoffs.

influences behaviour in strategic environments [Yamagishi and Kiyonari, 2000, Charness et al., 2007], and modifies performance [Hoff and Pandey, 2006]. Akerlof and Kranton [2000] posit that agents receive utility from following prescriptions associated with their social categories. Self-reports suggest that many subjects in our game perceive a social prescription to restrict links to the in-group. When the player with the desired position in the network belongs to the out-group, the farmer with the turn faces a tradeoff between efficiency and conformity with the social prescription. We expect that at least some subjects will choose to conform.

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

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

Our findings can be summarised as follows. First, realised efficiency is higher than that achieved by a purely random link formation process, but lower than the level of efficiency which myopic selfish or efficiency-minded players would have achieved. In the first treatment, the simple rule to link with the player who accesses the highest number of individuals would deliver 96 percent efficiency. Realised efficiency is about 65 percent, 31 percentage point below what the simple rule could have achieved. Interestingly, the direction of the benefits of the links does not significantly affect the level of architecture efficiency achieved by players.

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

Third, when information about group membership is disclosed, the resulting networks have more in-group links, but are not significantly less efficient. Interestingly, disclosure of group identity has different effects in the two two treatments. In-group links are chosen more often in the first treatment, while the frequency of decisions consistent with the predicted rule is unaffected. In the second treatment, the effect on the frequency in-group links is less pronounced, while decisions consistent with the efficiency-minded rule are taken less often.

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

A parallel literature has used observational dyadic data from rural areas of developing countries to explore how specific networks for the sharing of risk, favours, information and labour are formed [Fafchamps and Gubert, 2007, Krishnan and Sciubba, 2009, Karlan et al., 2009, Comola, 2010, Jackson et al., 2012, Santos and Barrett, 2010, Comola and Fafchamps, 2013]. Empirical studies of naturally occurring networks typically document some degree of homophily [McPherson et al., 2001]. A recent theoretical literature distinguishes between homophily motivated by *preferences*, *opportunities* or *strategic behaviour* [Currarini et al., 2009, Currarini and Menge, 2012, Tarbush and Teytelboym, 2014].[3]

[3]Individuals may have a desire to match with in-group partners, they may simply be exposed to more

Falk and Kosfeld [2012] study a game of unilateral, one-way-flow link formation that is based on Bala and Goyal [2000] and hence is closely related to ours. The design we propose, however, differs on a number of dimensions: links are added to the network one at a time; players are allowed only one link, so that the only cost of a connection is the opportunity cost of not forming another connection; the game is played by groups of 6. The first two features limit coordination and computation problems and make the game simpler. Falk and Kosfeld [2012] find that efficient networks are achieved in about half of the periods of the game. However, they do not report an average efficiency statistics for the one-way-flow treatment, which makes it difficult to compare their results to ours.[4]

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

Finally, our study is related to identity economics and a rich literature in economics and social psychology, referenced above, that studies how group categorisation generates in-group bias and modifies behaviour. A recent experiment by Currarini and Menge [2012] shows that group categorisation produces both in-group bias in allocation and homophilous matching. Interestingly, a further treatment with exogenous matching shows that in-group bias is *lower* when subjects can choose their interacting partners.

Our contribution is threefold. First, we document low levels of network efficiency in unilateral, one-way flow link formation. This is a stark result in a simple game with limited coordination issues and clear theoretical predictions. The literature has recently started exploring how subjects in rural areas of developing countries achieve

potential matches with in-group peers, or they may link with in-group peers because it is in their material interest to do so, for example, in order to avoid sanctions associated with deviations from social norms.

[4]We can however note that in the last period of our game, the cycle architecture is achieved in less than 10 percent of the sessions, that is, much less frequently than in any of the one-way-flow treatments devised by Falk and Kosfeld [2012].

lower efficiency in trading experiments than Western players [List, 2004, Bulte et al., 2012]. Our findings suggest that networks could be a second domain of widespread inefficiency and cross-cultural difference. Second, we document heterogeneous link formation strategies. This variety of motives can inspire models of network formation that explicitly include other-regarding preferences. Third, our results document an effect of arbitrary group categorisation on networks. This expands our understanding of the settings in which group categorisation modifies behaviour. It also shows how, beyond the material payoffs of the game, the formation of networks can be influenced by features of the social world. This could be a particularly fertile area for future experimental research on networks.

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

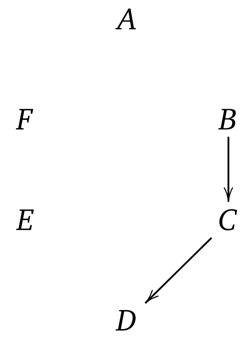
In the experiment subjects unilaterally create one-way-flow links.[5] In each session, 6 participants play the game. One of the six players is randomly drawn at the end of the experiment to receive a monetary prize.[6] Players who are directly or indirectly connected to the winner of the prize in the final network receive a prize of equivalent value. In the network represented in figure 15, for example, player B accesses players C and D, and hence receives the prize if C or D win the draw. Player C accesses only player D, while players D, E, F and A do not access any other player.

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

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

[6]The prize is worth 100 Indian Rupees, or about 5.2 USD at PPP, given an exchange rate at the time of 0.0155 USD per INR and a PPP conversion factors of 10/3 from the 2011 ICP round (http://data.worldbank.org/indicator/PA.NUS.PPPC.RF). This figure would grow if we applied a conversion factor calculated using rural prices only. For comparison, notice also that in 2012-2013 the National Rural Employment Guarantee Scheme paid an average daily wage to his workers of 121 INR (http://nrega.nic.in/netnrega/home.aspx).

Figure 1: Example



players create one link. In the second round, players can rewire their existing link.[7]

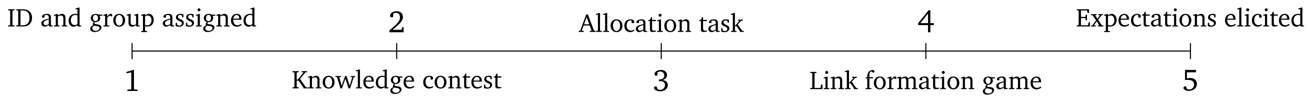
Players' decisions are recorded on a network map drawn on a white board visible to all players. The map is updated after every turn with the last decision. A number of design features ensure sequential updating takes place without breaking anonymity.[8] Furthermore, the pilot revealed that when the network map has more than a few links players find it difficult to calculate the number of direct and indirect connections of each peer. We hence remind the decision maker of the number of connections every other player has in the current network. The counting of connections is done by means of a Java application running on a small laptop operated by the game assistant. After entering a new link, the software produces a table with the number of connections each player has in the current network. This number is written next to the respective player ID on the white board before the next decision maker takes his turn.

The experimental tasks are carried out in the following order. First, players randomly draw a card from an urn which assigns them a letter ID and an experimental group. Second, participants answer three questions on agricultural knowledge, which

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

[8] 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 for reflection, the game assistant then 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.

Figure 2: Order of activities in the experiment



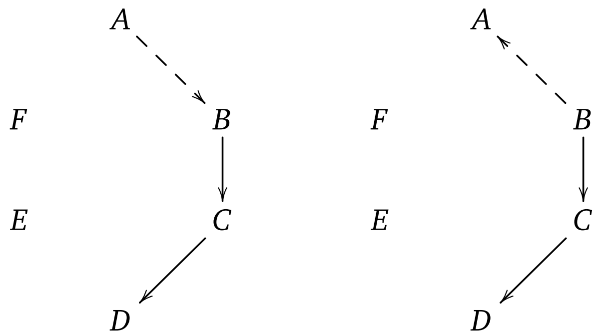
are part of an intergroup contest in agricultural knowledge. At the end of the game, if all players in a group have answered all questions right, the group receives one point and is applauded by everyone. Points are summed across sessions and participants are informed of the overall ranking between the two groups.[9] Third, participants play a simple allocation task, where they have to divide a fixed sum of money between an in-group and an out-group recipient randomly drawn from participants in the following session of the experiment. Fourth, participants play the link formation game. They are given the instructions of the game, they answer a number of questions which test their understanding of these instructions[10], and they then play a trial of the game that lasts for seven rounds. At the end of the trial, the game assistant randomly draws a participant and shows who would receive the prize if this was the actual game. After this, the actual link formation game is played. Fifth, participants are asked three questions about their expectations and beliefs and are then administered a short questionnaire, which collects information on socio-demographic variables and asks participants to explain the motivation behind their decisions in the game. At the end of this fifth phase, participants are informed of which team won the contest in agricultural knowledge and of the number of points each team has collected across sessions.

We rely on a between-subject design. We vary the direction of the flow of benefits associated with a link. In Treatment 1 (henceforth T1), players form links that let them access other individuals. This means that if player A chooses player B, then player A will receive the monetary prize whenever B wins the prize, but not vice versa. In Treatment 2 (henceforth T2), links let other players access the player who proposes the link. If A chooses a link with B, then B will receive the monetary prize whenever A wins it, but not vice versa. Figure 3 illustrates. Following our theoretical predictions

[9] Notice this information is disclosed only at the end of the game, that is, after step 5 in figure 2. So, whilst the contest creates the feeling of inter-group competition on a second, unrelated domain, it does not affect the beliefs players have about the levels of knowledge and cognitive ability of players in their group.

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

Figure 3: Links in the two treatments



(a) T1 (b) T2

Table 1: Summary of treatments

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

in the next section, as the network is updated, players are reminded of the number of individuals which each player access. In T2, players are also reminded of the number of individuals who access a particular player. For example, in panel (a) of figure 3, player C accesses one more player (player D). While two players (A and B) access player C. In T1, we report only the first of these two numbers. In T2 we report both.[11]

We also vary the information about peer group membership available to players during the link formation game. This is cross-cut with T1 and T2. In a first set of treatments, which we call T1no and T2no, individuals have no information about peers' group affiliation. Hence their link formation decisions are by design unrelated to the groups formed at the beginning of the experiment. In a second set of treatments, called T1id and T2id, group identity is common knowledge, as players belonging to different groups are identified with different symbols on the public network map on the whiteboard.[12] We hence run four treatments, as shown in table 1.

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

[11]We give precise definitions of these concepts in the following section.

[12]Mango group players are identified with a circle. Pineapple group players are identified with a triangle.

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

In our design group membership is randomly allocated. The original experiments in social psychology, on the other hand, rely on groups which are formed on the basis of a trivial preference.[13] While preference-based matching has the potential to increase the saliency of group membership, it also has two disadvantages. First, players' characteristics may be correlated to what the researchers considers as orthogonal preferences. Second, even if the chosen set of preferences is truly orthogonal, some players may believe that such correlation exists. For example, a player may (erroneously) think that people with a certain preference in art or sport are smarter. In both cases, the effects of common knowledge of social identity would be confounded by those associated with correlated categories and beliefs.

We hence opt for a design which relies on random assignment to social groups and increases the saliency of group identity by means of the contest in agricultural knowledge. This task combines four desirable features: (i) it is linked to the overall framing of the experiment, (ii) it creates a feeling of competition between the two groups on a domain, that of agricultural knowledge, which is distinct to the domain of monetary outcomes of the experiment, (iii) the relative position of the two groups in this second domain is only revealed after the link formation game has been played, (iv) every player can have a strong marginal impact on the group's outcome: if a player fails to answer one question correctly, the whole group fails to gain the point for that session. The idea of using contests to increase the salience of group identity has been successfully used before in experimental studies [Eckel and Grossman, 2005].

To ensure comparability and minimize noise factors 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].

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

2.1 Possible Confounders

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

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

Figure 19 in the appendix presents the cumulative distribution of mistakes. In both T1 and T2 more than 50 percent of players did one mistake or less, and about 80 percent of players did 2 mistakes or less. Given that these mistakes were followed by further explanations, we feel confident that the experiment was well understood.

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

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

Wealth effects. Both the allocation and link formation tasks are incentivised with monetary payments. However, there should be no "within-person" wealth effects across these two tasks, as the first involves no costs to the participants. Furthermore, in the allocation task individuals allocate money between two peers in a future session of the experiment. So, for example, somebody who has given more to the in-group in the allocation task has not affected the wealth of in-group individuals in the current session, which could in turn affect play in the link formation game. Hence "between-person" wealth effects are also ruled out.

Experimenter demand effects. These arise when subjects, in an attempt to please the experimenter, respond to implicit cues embedded in the experimental design [Zizzo, 2010]. For example, the fact that we disclose information about group identities may suggest to players that we expect them to use this information somehow. To 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 identities in the instruction phase and in the trial round, to avoid making unintended suggestions about how we expect players to use the group membership information.

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

3 Predictions

3.1 Notation

We define some basic notation following Goyal [2012]. Let N=(1,2,..,n) be the set of players. In T1, each player i chooses a (pure) strategy $g_i = (g_{i1}, g_{i2}, ..., g_{ii-1}, g_{ii+1}, ..., g_{in})$[14], 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.[15] $\Gamma = \Gamma_1 \times \Gamma_2 \times ... \times \Gamma_n$ 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 series 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 $i \to^g j$, player i receives the prize whenever player j receives it. $i \to^g j$, on the other hand, has no implication on whether player j receives the prize when player i receives it.

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

[15]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.

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 accesses through a path of links in network g. Sometimes we want to exclude from the count the path from player j to player i who has the turn. 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 accesses*** through a path of links, *excluding player i*.

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 access player j through a path of links in the network. Again, we sometimes need to exclude the path from player i, who has the turn, 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 access player j*** through a path of links, *excluding player i*.

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.[16]

Network g determines an expected payoff $\pi_i(g)$ 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 i (this time, including player i himself):

$$\pi_i(g) = \frac{1 + \mu_i(g)}{n} \tag{1}$$

3.2 Link formation rules and network efficiency

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 types of preferences. Then,

[16] 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 to whom player i has a direct link. $\mu_i^d = |N_i^d(g)|$ is the *number* of players to whom player i has a direct link. We call this 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.

we simulate link formation processes where individuals follow the proposed rules and study what levels of session level efficiency these rules achieve. Our central message will be that simple myopic selfish best response in T1 is sufficient to achieve fully efficient networks (almost) always.

We follow much of the existing literature and assume myopic behaviour: the player with the turn chooses his link as if no more turns would follow, that is, as if the network which obtains after his link is added is the final network that determines players' payoffs. 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, new links affect the expected payoff of the player with the turn. Only one link is permitted. A new link to player j allows player i to access all the $\mu_{-ji}(g)$ individuals whom player j accesses. Thus, picking the partner j with the highest value of $\mu_{-ji}(g)$ maximises the expected payoff of player i.[17] The link formation rule of player i in T1 will hence be:

Rule 1. *Choose a link to the player j with the maximum value of $\mu_{ji}(g)$. In case of a tie, randomise.*

In T2, a purely selfish player would be indifferent between forming and not forming a link because $\pi_i(g + ji) = \pi_i(g)$. He would be equally indifferent about the consequences on the welfare of other players of the link he establishes. 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].[18] Following the literature on other-regarding preferences, we assume that players have a utility function that weights concerns for the player's own payoff and the payoff of all other players:

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

[17] In the appendix we show this formally.

[18] 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.

where $\pi_{-i}(g) = \{\pi_i, \pi_2, .., \pi_{i-1}, \pi_{i+1}, ..., \pi_n\}$. To advance further we have to make some assumptions about the shape of function f. We can explore two archetypal candidates. The first is:

$$u_i(g) = \pi_i(g) + \gamma \sum_{j \in N \setminus i} \pi_j(g) \tag{3}$$

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 player i creates a link g_{ji}, he increases the expected payoff of player j and of all the players who access player j. Intuitively thus, the effect on aggregate welfare of a new g_{ji} link is proportional to the number of individuals who access player j.[19] 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. *Choose a link to the player j with the maximum value of $\mu_{-ji}(g)$. In case of a tie, randomise.*

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:

$$u_i = \pi_i(g) + \gamma \min_{j \in N \setminus i} \pi_j(g) \tag{4}$$

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 j with the lowest value of $\mu_{ji}(g)$ in network g.[20] This is the player with the lowest chance of winning the prize : $\pi_j(g) = \frac{1+\mu_j(g)}{n}$. We predict that a fraction of players in in T2 have other-regarding preferences expressed by 4 and will thus play according to the following link formation rule:

[19] In the appendix we show this formally and explain one qualification that applies to g_{ji} links that create a small cycle.

[20] In T2, when i has the turn, $\mu_{ji}(g) = \mu_j(g)$ as there are no paths from j to i before i chooses his link.

Rule 3. *Choose a link to the player j with the minimum value of $\mu_{ji}(g)$. In case of a tie, randomise.*

Notice that, depending on the architecture of g, the two sets of players who satisfy rules 2 and 3 are disjoint or overlapping. Figure 15 in the appendix shows an example where the two sets are disjoint: F has the minimum value of $\mu_{ji}(g)$ while A,B,C, and D all have the maximum value of $\mu_{-ji}(g)$.

A third model of social preferences is that of inequity 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 choose links to partners with the maximum value of $\mu_{ji}(g)$. In T2 players will choose links either to partners with the maximum value of $\mu_{-ji}(g)$ or partners with the minimum value of $\mu_{ji}(g)$.*

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.

We simulate link formation games where players follow the link formation rules outlined above and we study the overall efficiency of the resulting architectures. We define efficiency as the ratio between the average number of players accessed by each individual in the final network and the maximum number of individuals that can be accessed: 5.

$$\text{Efficiency}_g = \frac{\frac{1}{n}\sum_{i=1}^{n}\mu_i(g)}{5} \tag{5}$$

The cycle architecture, where each player has a $\mu_i(g)$ value of 5 and all players win the prize, has efficiency 1 under this measure. All other possible architectures 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. Notice that the cycle architecture is also efficient in the Pareto sense.

Our first set of simulations shows that, in T1, link formation rule 1 delivers average efficiency of about 96 percent. Figure 4 gives an example of how play in accordance with rule 1 achieves the cycle architecture within 2 rounds. 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 configuration where the player who has to re-wire his link to form the cycle has already played his second turn. If we allow more rounds, the likelihood of this configuration occurring in every round becomes very small. For example, in three rounds rule 1 achieves 99 percent efficiency.

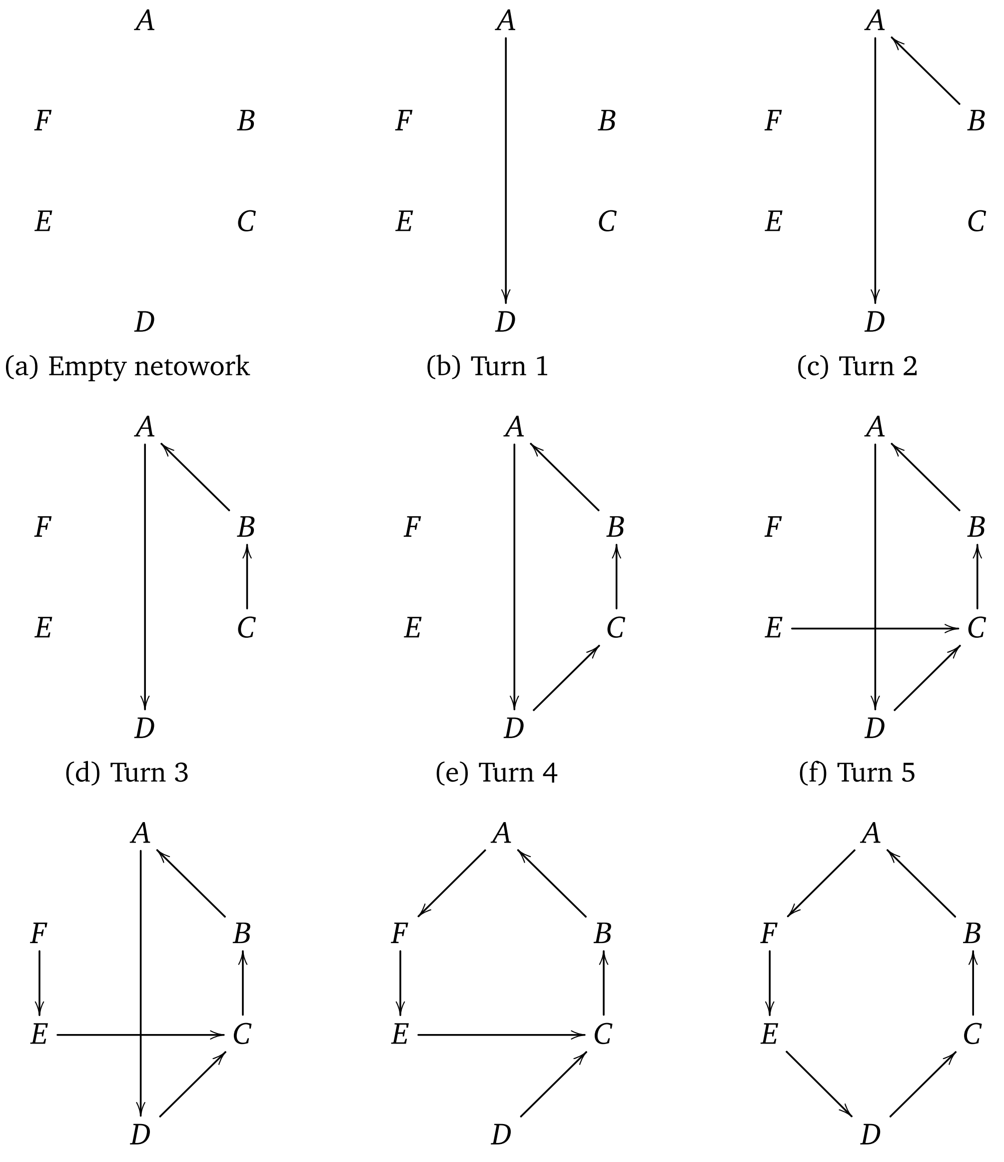
Our second set of simulations show that, in T2, link formation rule 2 also delivers average efficiency of about 96 percent.[21] Rule 3, on the other hand, delivers 67 percent efficiency. Figure 5 reports kernel density estimates and average efficiency for simulated sessions where play is in accordance, respectively, with rule 2, rule 3, and with a random link formation process. The random link formation process achieves average efficiency of about 52 percent.

We also study efficiency in sessions where a mix of rules is played. We simulate sessions where a fraction p of decisions follow rule 3, and a fraction 1-p of decisions follow rule 2. Results show that efficiency decreases monotonically with p in the interval between 96 and 67 percent. Figure 17 shows this graphically. We can thus formulate the following prediction on session level efficiency.

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

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

Figure 4: Network evolution under myopic selfish play in T1



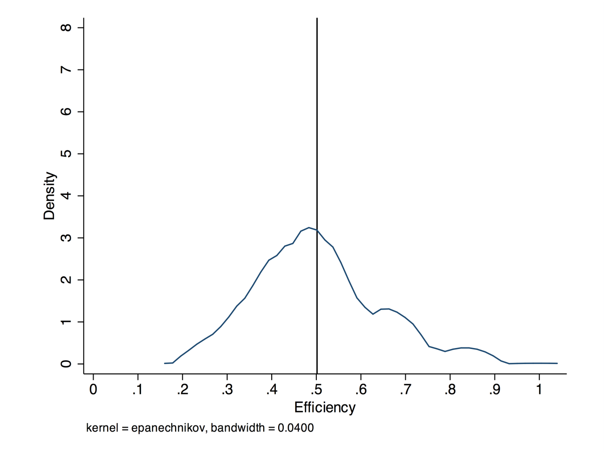
(a) Empty netowork (b) Turn 1 (c) Turn 2

(d) Turn 3 (e) Turn 4 (f) Turn 5

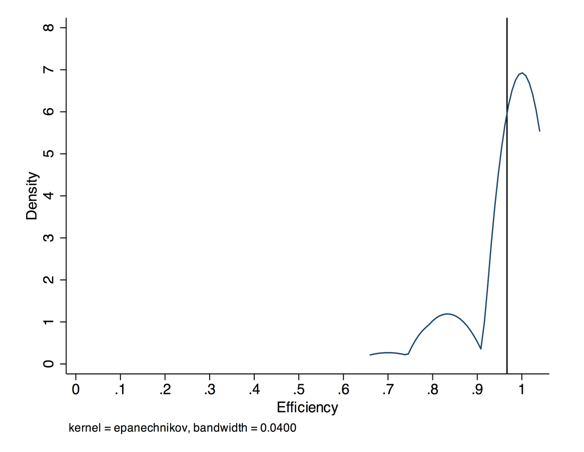
(g) Turn 6 (h) Turn 7 (i) Turn 11: **The cycle!**

Note. Order of play is assumed to be the order of the alphabet for ease of presentation. All players in this simulation play according the link formation 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.

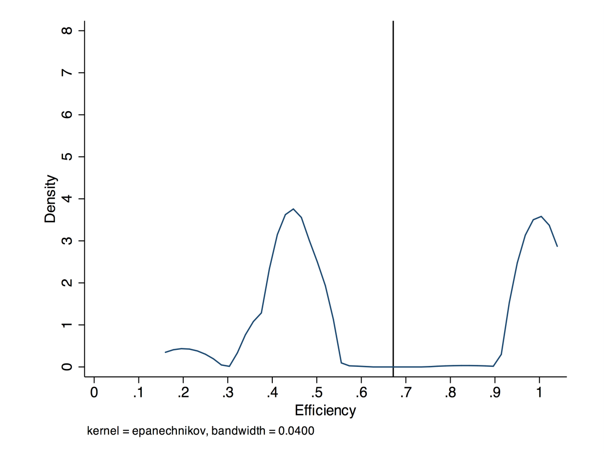
Figure 5: Average efficiency under different link formation rules in T2



(a) Random



(b) Rule 2: Maximum $\mu_{-ji}(g)$



(c) Rule 3: Minimum $\mu_{ji}(g)$

Note. Each panel reports kernel density estimates of the distribution of the average value of $\mu_i(g)$ after 12 turns of play for 500 simulated sessions. The vertical line indicates the average value for all simulations, for a given rule. The link formation rule used in each set of simulations is indicated below the panel.

Figure 16 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 social identity is common knowledge, we have to augment the utility function above. We follow the seminal paper of Akerlof and Kranton [2000] and introduce a positive effect on utility which comes from following a prescription P associated with the social group s_i to which player i is affiliated:

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

Suppose that individuals get positive utility whenever they create an in-group link: $P_i(s_i) = c > 0$ whenever i has linked to an in-group peer. 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.

When group identity becomes common knowledge, we expect two types of effects. First, for any positive value of c, whenever the set of individuals who satisfy a given link formation rule contains both in-group and out-group members, decision makers with preferences given by 6 would link to an in-group peer. The frequency of in-group links will increase. Efficiency, on the other hand, will not be affected as the frequency with which the various link formation rules are chosen does not change.

Second, if the benefit c from following the prescription is high enough, we will also observe more in-group links in cases when the best response set contains only out-group peers. Now the frequency of decisions consistent with the link formation rules is reduced. As a result, the efficiency of the network will decrease.

We can put a “price" on in-group links in terms of the extent to which players accept to be constrained in the satisfaction of their chosen link formation rule. For example, a player who otherwise plays according to rule 1 in T1 may be willing to link to an in-group partner j outside of his best response set as long the difference in $\mu_{ji}(g)$ between j and his otherwise preferred partner is at most one unit. Formally, $max_{k \in N \setminus i}(\mu_{ki}(g) - \mu_{ji}(g)) <= 1$. Figure 18 shows this analysis graphically for the case in which every player puts a price on identity of 2. Average efficiency drop to 53 percent when players play rule 2 and to 39 percent when players play rule 3.[22] We hence formulate the following, final prediction:

Prediction 3. *Common knowledge 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. This is a test of the null that the outcomes of the two treatments are drawn from same distribution against the alternative hypothesis that either outcome is stochastically greater than the other.[23] Further, we study individual decisions with dyadic regression analysis. In particular, we use models of the following form:

$$\text{link}_{ij,rs} = \alpha + \beta \textbf{Network Position}_{j,rs} + \gamma \mathbf{D}_{ij} + \delta \text{round}_r + e_{ij,rs} \tag{7}$$

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. $\text{link}_{ij,rs}$ 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

[22]As a limit case, notice that myopically selfish agents who put a price on in-group links of 6 or more would never link to an out-group peer. Such players will converge to two small cycles with 3 players each. The average value of $\mu_i(g)$ for this network architecture will be 2, corresponding to 40 percent efficiency.

[23]For two populations A and B, A is stochastically greater that 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.

link formation rules we propose above. For T2, these include a dummy for having the minimum value of $\mu_{ji}(g)$, and a dummy for having the maximum value $\mu_{-ji}(g)$. For T1, they include a dummy for having the maximum value of $\mu_{ji}(g)$. As a check, in T1, we also include a dummy for having the minimum value of $\mu_{-ji}(g)$. For robustness, we will also run specifications where we include the absolute values of $\mu_{ji}(g)$ and $\mu_{-ji}(g)$.

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.[24] 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, $E[e_{ij,rs}e_{ik,rs}] \neq 0$. This inference problem is typical in dyadic regression analysis [Fafchamps and Gubert, 2007]. We correct for intra-session correlation in error terms using cluster-robust standard errors for inference.

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

[24]For example, see figure 4. From the perspective of player A, while B's position in the network is evolving, B remains A's closest neighbour in the visual representation of the set of players. This may make 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.

4 Data

We run our field experiment in the Indian state of Maharashtra. We randomly sample from a census list of all villages in 4 “talukas" (sub-districts) of the Pune and Satara districts.[25] Villages in these subdstricts are situated approximately 1,30 to 3h hours away from Pune. This is a similar distance to the district capital as that of the villages selected in the study of Banerjee et al. [2013]. To reflect the large heterogoeneity in geographic conditions in this area of India, we choose two subdistricts which mostly comprise mountainous areas, and two subdistricts in the agricultural plains.

We select study participants through door-to-door random sampling. Before reaching the village, our team is shown a Google Earth map of the village. On alternating days, the teams start sampling from the periphery of the village or from the center of the village.[26] We invite all male adult farmers who are encountered in the door-to-door visit until we have enough farmers to fill in all planned sessions.

Data collection took place between September and October 2013. In total, we run 81 sessions with 486 subjects. We run 20 sessions of T1no, T1id and T2id, and 21 sessions of T2no. In three of the sessions one participant left before the beginning of the link formation stage. This leaves us with 483 subjects, which correspond to 4800 dyads.[27] Table 2 summarises the number of observations we have for each treatment.

Table 2: Number of observations by treatment

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

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

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

[27] Each person in sessions with 6 individuals creates 10 dyads (5 per round). Each person in sessions with 5 individuals creates 8 dyads (4 per round).

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

Table 3: Summary statistics: Individual Covariates

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

From session 9 onwards,[29] 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 all the other 5 participants. On average, farmers speak on 13.5 days in a month with each of the participant they know.

In tables 6 to 9 in the appendix, we present some regressions that test for covariates balance across treatments. We cannot find any statistically significant difference in average characteristics across treatments.

[28]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 3.

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

Table 4: Summary statistics: Session Networks

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

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

5 Results

We organise our discussion around four key results.

We first investigate overall efficiency. Table 10 summarises treatment level averages of μ_i[30] 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 benchmarks: the distribution of average session efficiency which would obtain if individuals chose their links at random, and the distribution under myopic selfish play (rule 1). We obtain the following result, which is represented graphically in figure 6 below:

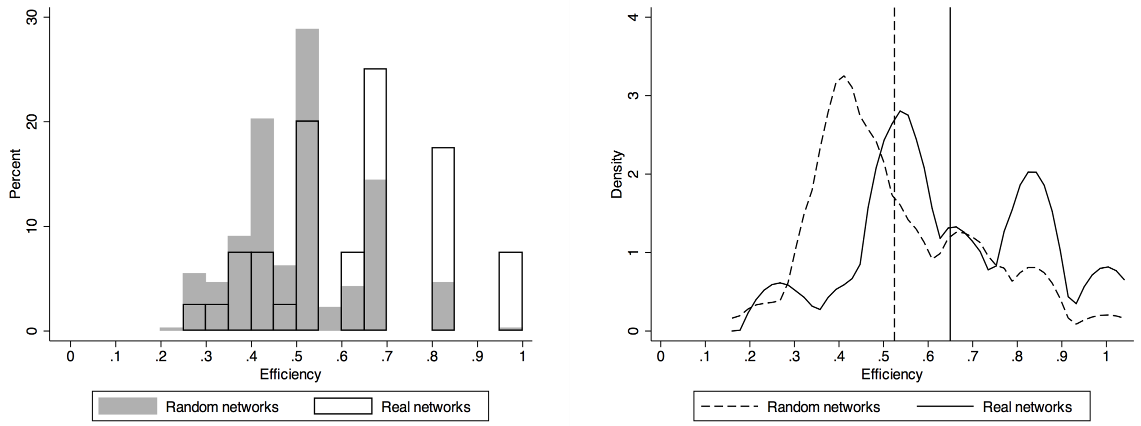
Result 1. *Architecture efficiency in T1no and T2no is 65 percent. This is 13 percentage points above average efficiency under random network formation, but 31 percentage points below average efficiency under "efficiency-minded" link formation (rule 1 in T1, rule 2 in T2). Both of these differences are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

The efficiency of the experimental networks is 31 percentage points below the average level achieved by the "efficiency-minded" link formation rules. A Wilcoxon rank sum test confirms that this difference 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

[30]For simplicity, from now on we will drop the reference to network g in the symbols $\mu_i(g)$, $\mu_{-i}(g)$, $\mu_{ji}(g)$, $\mu_{-ji}(g)$.

Figure 6: Efficiency in No-Identity Sessions and in Random Networks



(a) Histogram

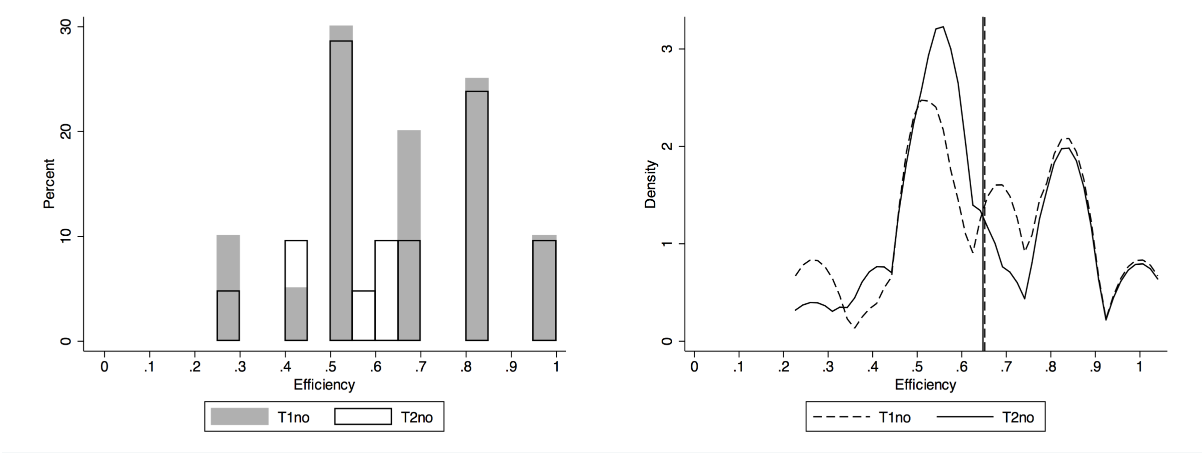
(b) Kernel density estimate and means

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 7 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 prediction for T1 is clearly rejected.

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

It is important to note that low efficiency is not an artefact of truncation at 12 turns: efficiency has no monotonic updward trend in either T1no or T2no, and efficiency at turn 12 is only a few percentage points higher than it was at turn 6. Figure 8 illustrates. Falk and Kosfeld [2012], on the other hand, document strong learning dynamics and positive efficiency trends in their experiment.

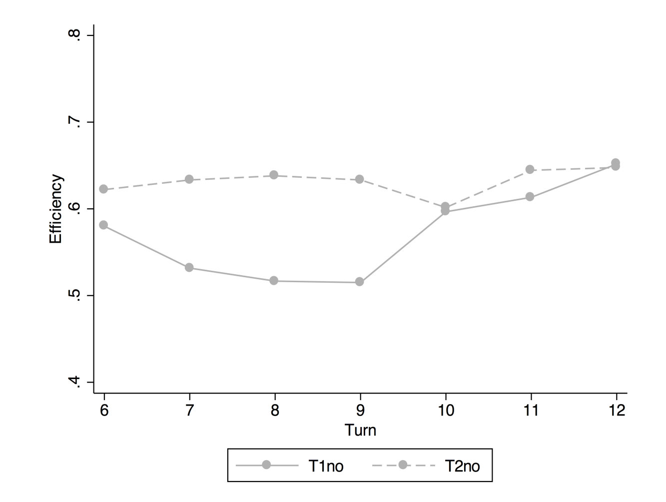
Figure 7: Efficiency in T1no and T2no sessions



(a) Histogram

(b) Kernel density estimate and means

Figure 8: Time series of effciency in T1no and T2no, turns 6-12



Efficiency is lower than what any of the rules we specified would have achieved. Further, it is not different across the two treatments. How do these results come about? We can answer this question estimating dyadic regression model 7. Results are reported in table 5.

As hypothesised, we find that in T1no ties are directed towards players with the maximum level of μ_{ji}. In T2no, on the other hand, ties are directed towards players with the minimum μ_{ji} and the maximum μ_{-ji}. 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 level of μ_{-ji}. As player who does not the have the maximum level of μ_{-ji} has a probability of being chosen of 32.3 percent, this amounts to a 40 percent increase. In T2no, player i is 11 percentage points more likely to choose a player with maximum μ_{-ji} and 7 percentage points more likely to pick a player with minimum μ_{ji}. A Wald test cannot reject the equality of these two coefficients.

Table 5: Dyadic Linear Probability Model 7

	(1)	(2)	(3)	(4)
Panel a				
$\max\left(\mu_{ji}\right)$	.132	.130		
	(.001)***	(.001)***		
$\min\left(\mu_{-ji}\right)$	.018	.016		
	(.235)	(.314)		
$\max\left(\mu_{-ji}\right)$			.111	.120
			(.004)***	(.002)***
$\min\left(\mu_{ji}\right)$			.073	.066
			(.04)**	(.072)*
Panel b				
$\max\left(\mu_{ji}\right) = \min\left(\mu_{-ji}\right)$	10.34	10.81		
	(.004)***	(.004)***		
$\max\left(\mu_{-ji}\right) = \min\left(\mu_{ji}\right)$			0.57	1.12
			(.459)	(.304)
Obs.	1200	910	1260	940
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: *** $\longleftrightarrow$ 99%, ** $\longleftrightarrow$ 95%, * $\longleftrightarrow$ 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 belonging to a non backward caste. Standard errors are corrected for clustering at session level. P-values obtained with wild bootstrap-t procedure reported in parentheses. Panel b reports the F statistics (and p value in parenthesis) for a Wald test of the equality of coefficients.

We confirm the robustness of these results by running a specification that substitutes the dummies with the absolute values of μ_{ji} and μ_{-ji}. This allows players to make mistakes, while requiring larger mistakes to be less likely than smaller mistakes. Table 12 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 value of μ_{ji} of 4 than a player with a value of μ_{ji} of 0. Given that a player with $\mu_{ji} = 0$ 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 constitutes evidence in favour of prediction 1:

Result 3. *In T1no, links to a partner who has the maximum level of μ_{ji} are significantly more likely to be formed than other links. In T2no, links to a partner who has the maximum μ_{-ji} or the minimum μ_{ji} are significantly more likely to be formed than other links.*

Table 12 shows a further significant effect: in T1no player i is more likely to establish a link with a player with a lower value of μ_{-ji}. A caveat is in order, as, in the previous specification, when we include a dummy for whether an individual has the minimum value of μ_{-ji} we report a positive, but small and insignificant coefficient. This suggests that the effect of μ_{-ji} 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-degree 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. We have however some qualitative evidence in support of it. In the post-play questionnaire farmers are asked the following question: "Do you think that choosing a farmer from your own group is a way of showing respect to him?". 51 percent of farmers answer yes to this question. This is consistent with the view that links carry social value, but represents by no means a full-fledged test.[31]

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

Rule 4. *Choose a link to the player j with the minimum value of $\mu_{-ji}(g)$. In case of a tie, randomise.*

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 frequencies of decisions consistent with the various rules. In T1no, 51 percent of

[31]An alternative explanation could appeal to invidious preferences. However, somebody who wants to myopically maximise the difference between his own payoff and that of the other players would prefer not to establish any link in the first place. So we do not think invidious preferences offer a convincing explanation for this result.

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.

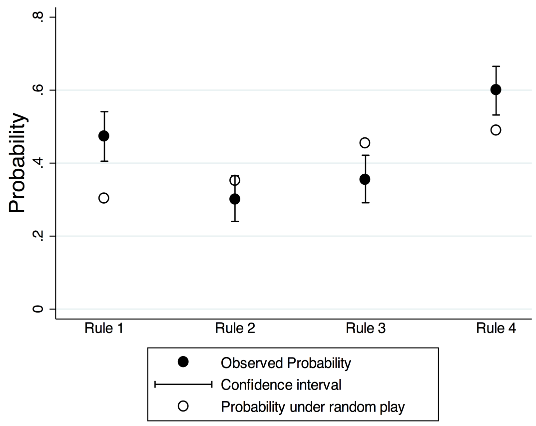
This exercise, however, poses two problems. First, there are often multiple individuals who satisfy a particular linking rule. Hence, rules with a larger number of candidates are selected more frequently when individuals choose randomly or make random mistakes. This makes it difficult to interpret frequency values Ï and to compare different rules. To address this, we calculate the probability of observing a decision consistent with a particular rule when links are uniformly randomised across all possible partners. We then calculate a confidence interval around the frequency with which we observe in the data decisions consistent with the same rule. 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 μ_{ji} and the minimum μ_{-ji}. The last individual, on the other hand, has the maximum μ_{-ji} and minimum μ_{ji}. This complicates comparison across rules. To address this problem we classify decisions on the basis of whether the sets of potential partners satisfying the efficiency-minded and Rawlsian rules (rule 1 and rule 4 for T1, rule 2 and rule 3 for T2) are disjoint, partially overlapping, or fully overlapping. Results are presented in panel b of figure 21 in the appendix. Overlaps are very frequent. We will hence repeat the analysis for the subsample of decisions where the player is presented with a network where the sets satisfying the two rules are disjoint or only partially overlapping.

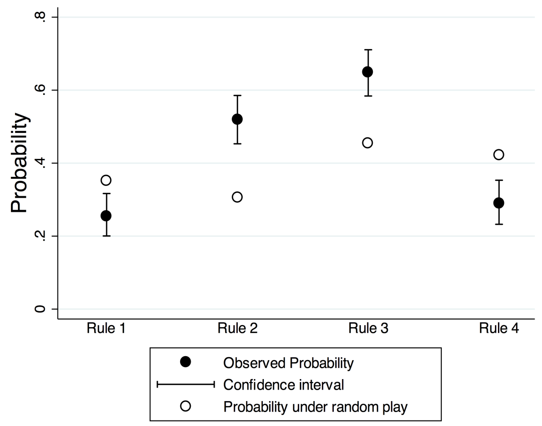
Figure 9 presents the analysis for the whole sample. In T1no, both decisions consistent with rule and decisions consistent with rule 4 are observed significantly more often than under random play. While decisions consistent with rule 4 decisions are more frequent, they also have a higher probability of occurring under random play. In T2no, rules 2 and 3 are also observed significantly more often than under random play.

The pattern above is basically unaffected if we restrict the analysis to decisions over network architectures where the sets of individuals satisfying the efficiency-minded and Ralwsian rules are not perfectly overlapping. In T1no, however, the difference between the observed frequency of decisions consistent with rule 4 and the frequency of these decisions under random play is significant only at the 10 percent level. Figure 24 in the appendix shows this.

Figure 9: Probability that a decision is consistent with a particular rule



(a) T1no



(b) T2no

The frequency of "efficiency-minded" and "Rawlsian" links is similar across the T1no and T2no treatments. This explains why these two treatments have close levels of session efficiency. In both T1no and T2no, about 70 percent of decisions are consistent at least one of the archetypal rules. Panel (a) of figure 21 shows this. The majority of these decisions are consistent with both rules, while about 16 percent of links in both treatments satisfy only the "Rawlsian" rule.

To explore the robustness of these results we also consider jointly the two decisions taken by each player. Table 25 shows the probability that an individual plays consistently with an archetypal rule in both rounds. The picture is qualitatively unchanged, with each archetypal rule being played more often than random, as hypothesised.

What about the remaining 30 percent of decisions 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*[32]

Rule 6. *Choose a link to the player who has chosen you in a previous round* [33]

[32]We refer to this player as the "most popular" player. In T1, this corresponds to the player with the maximum value of μ^d_{-j} when i has the turn. In T2 to the player with the maximum value of μ^d_j when i has the turn. Past links that have been rewired are not included in the computation.

[33]We refer to this as the "reciprocal" strategy.

Among decisions that are not consistent with archetypal rules 1-4, 65 percent of decisions target the most popular player in the network. On the other hand, only 25 percent of decisions that are consistent with the archetypal rules target the most popular player. Decisions consistent with rule 6 follow a similar pattern. Figures 22 and 23 illustrate. Despite being relatively frequent when the archetypal rules are not followed, regression analysis reported in table 13 shows that rules 5 and 6 are not significant predictors of overall link formation decisions in T1no and T2no.[34]

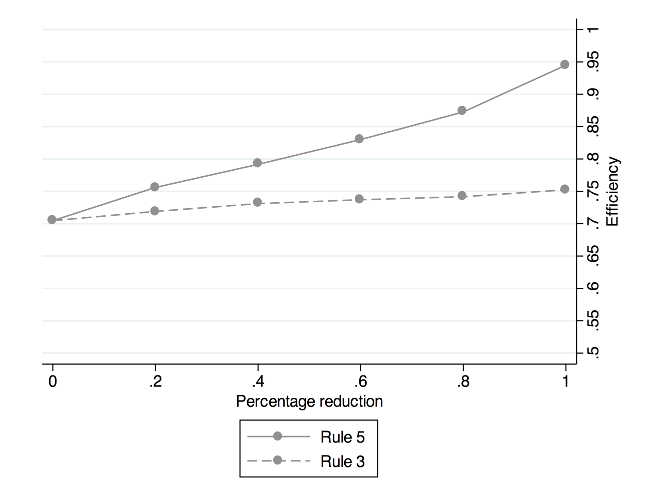
Simulation analysis shows that the largest efficiency gains can be achieved by reducing the proportion of links that are targeted to the most popular player, as opposed to reducing the proportion of Rawlsian links. We simulate a link formation process where 54 percent of decisions are consistent with rule 1, 16 percent with the rule 4 and the remaining 30 percent with the most popular player rule.[35] We then switch increasing proportions of decisions assigned to follow rule 4 to the 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.[36] Figure 10 illustrates. In figure 26 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. Rule 5 would be quite successful in games where the efficient architecture is a star network. In our game however, it performs quite badly.

[34]In future work, we plan to estimate a mixture structural model. This would allow us to allow a greater degree of heterogeneity in player strategies.

[35]This split reflects the data in our sample, 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.

[36]We also know from the simulations reported in figure 17 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 17 shows that network efficiency in such scenario would be above 90 percent, which corresponds to 20 percentage points gain.

Figure 10: Efficiency simulations



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

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

Figure 27 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.[37] 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 11 and figure 28 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.

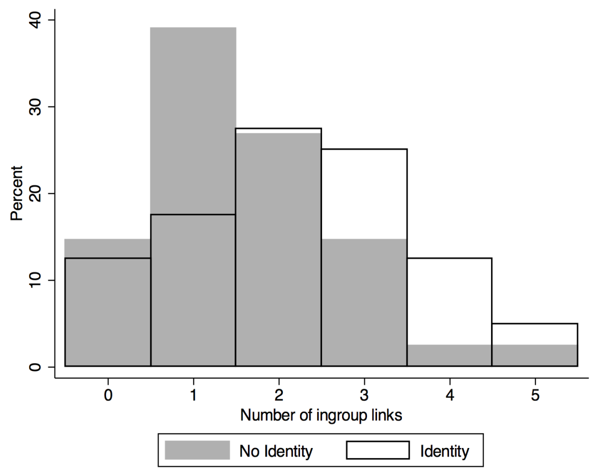
[37]Players in T1no and T2no also answer this question, albeit information about group affiliations was not disclosed in these treatments during the link formation game.

Our main result on the identity treatments is the following:

Result 4. *In treatments with common knowledge of group identity in-group links occur more frequently than in treatment with no knowledge of group identity, while architecture efficiency does not decrease.*

This result confirms the first part of prediction 3. First, in-group links increase. Figure 11 shows a histogram of the number of in-group links in the final architecture for identity and no identity sessions. 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 11: In-group Links in Identity and No-Identity treatments

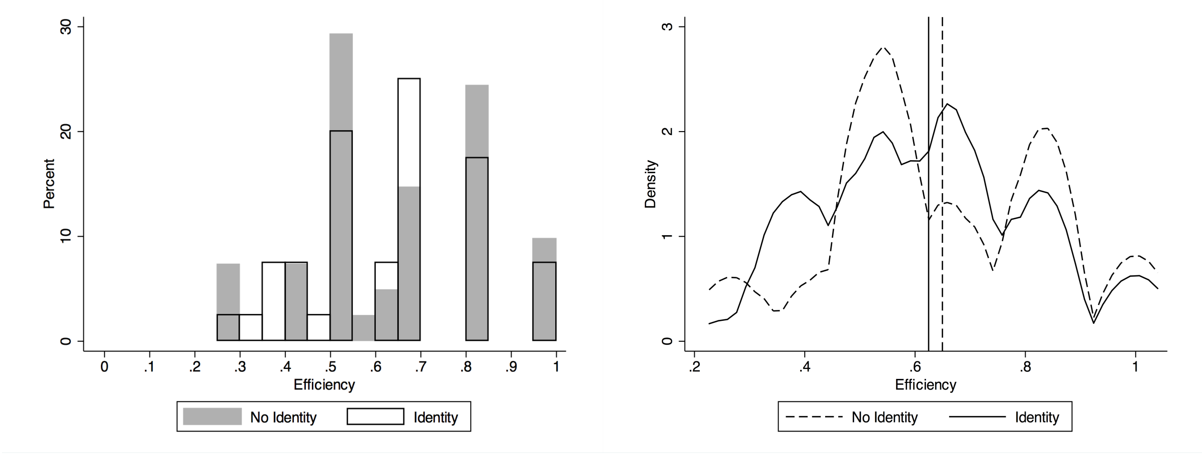


Note: Only links in final architectures are considered. No identity sessions include T1no and T2no. Identity sessions include T1id and T2id.

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 12.

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

Figure 12: Efficiency in identity and No-Identity Sessions



(a) Histogram

(b) Kernel density estimate and means

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

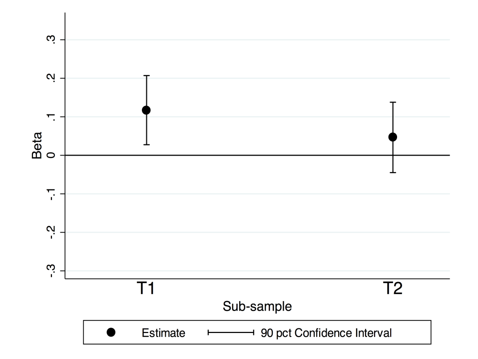
$$x_{dis} = \alpha + \beta_1 \text{Identity Session}_s + \beta_2 \text{z}_{is} + \beta_3 \left(\text{z}_{is} * \text{Identity Session}_s\right) + e_{dis} \tag{9}$$

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 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.[38] 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 13 shows graphically the coefficient estimates, while full regression tables are available in the appendix.

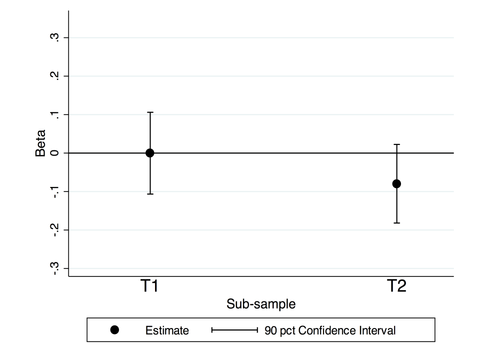
In-group links increase in both treatments. However, the effect is significant only for T1. In this condition, links to an in-group player are about 11 percentage points more likely once player identity is disclosed. This corresponds to a 40 percent increase in the probability of an in-group link. For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

[38]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 13: Linear probability Model 8: coefficient estimates

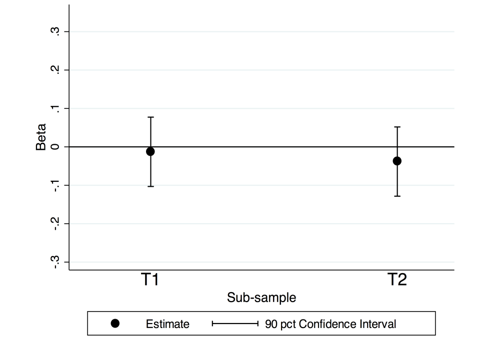


(a) x_{is}=: in-group link



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

Ï



(c) x_{is}= Rawlsian decision

Note. Coefficients estimates from linear probability model. The dependent variable is indicated below graph. Standard error are clustered at session level. Full regression results are reported in tables 14, 15, and 16.

The frequency of efficiency minded or Rawlsian links is unaffected by the disclosure of group identity in T1. Decisions consistent with the two archetypal rules are, on the other hand, observed less frequently in T2. Links towards players with the maximum level of μ_{-ji} 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 the minimum level of μ_{ji} also decrease by an insignificant 4 percentage points.

These results suggest that disclosure of group identity operates 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 being in an identity session on the likelihood of choosing an in-group link is not stronger for individuals who have showed 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.[39]

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

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

$understanding_{is}$ 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 identity is private. β_3 measures whether high understanding subjects are more likely to choose links consistent with strategy x in sessions where group identity is common knowledge. Tables 19 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 common knowledge, 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.

[39] As explained above, we plan to estimate a mixture structural model in future work. This will help us disentangle what kind of heterogeneity of motives could be driving this behaviour.

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 network architectures after repeated play [Bala and Goyal, 2000]. We offer the first experimental test of this prediction for a non-western population- a sample of farming communities in rural India. This is a policy-relevant setting: interest for new, cost-effective intervention designs that promote the diffusion of agricultural technology is currently high in India. We make a second contribution to the literature by exploring how a pervasive feature of the social world, group categorisation, affects the way networks are formed.

We find that the efficiency of the networks formed in our experiment is higher than that achieved by a purely random link formation process, but 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

G.A. Akerlof and R.E. Kranton. *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton University Press, 2010. ISBN 9781400834181. URL http://books.google.co.uk/books?id=1gKUpITuf0OC.

George A. Akerlof and Rachel E. Kranton. Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753, August 2000. URL http://ideas.repec.org/a/tpr/qjecon/v115y2000i3p715-753.html.

Siwan Anderson. Caste as an impediment to trade. *American Economic Journal: Applied Economics*, 3(1):239–63, September 2011. doi: 10.1257/app.3.1.239. URL http://www.aeaweb.org/articles.php?doi=10.1257/app.3.1.239.

James Andreoni and John Miller. Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2):pp. 737–753, 2002. ISSN 00129682. URL http://www.jstor.org/stable/2692289.

Venkatesh Bala and Sanjeev Goyal. Learning from Neighbours. *Review of Economic Studies*, 65(3):595–621, July 1998. URL http://ideas.repec.org/a/bla/restud/v65y1998i3p595-621.html.

Venkatesh Bala and Sanjeev Goyal. A noncooperative model of network formation. *Econometrica*, 68(5):1181–1230, September 2000. URL http://ideas.repec.org/a/ecm/emetrp/v68y2000i5p1181-1230.html.

Oriana Bandiera and Imran Rasul. Social networks and technology adoption in northern mozambique. *Economic Journal*, 116(514):869–902, October 2006. URL http://ideas.repec.org/a/ecj/econjl/v116y2006i514p869-902.html.

Abhijit Banerjee, Emily Breza, Arun Chandrasekhar, and Matthew O. Jackson. Come Play with Me: Experimental Evidence of Information Diffusion about Rival Goods. Working papers, 2012.

Abhijit Banerjee, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson. The diffusion of microfinance. *Science*, 341(6144), 2013. doi: 10.1126/science.1236498. URL http://www.sciencemag.org/content/341/6144/1236498.abstract.

Abigail Barr and Garance Genicot. Risk sharing, commitment, and information: An experimental analysis. *Journal of the European Economic Association*, 6(6):1151–1185, December 2008. URL http://ideas.repec.org/a/tpr/jeurec/v6y2008i6p1151-1185.html.

Michele Belot and Marcel Fafchamps. Good Samaritans and the Market: Experimental Evidence on Other-Regarding Preferences. Discussion Papers 2012001, University of Oxford, Nuffield College, August 2012. URL http://ideas.repec.org/p/cex/dpaper/2012001.html.

Ariel Ben Yishay and Mushfiq Mobarak. Communicating with farmers through social networks. *Working paper*, 2012.

Daniel J. Benjamin, James J. Choi, and A. Joshua Strickland. Social identity and preferences. *American Economic Review*, 100(4):1913–28, September 2010. URL http://ideas.repec.org/a/aea/aecrev/v100y2010i4p1913-28.html.

Erlend Berg, Maitreesh Ghatak, R Manjula, D Rajasekhar, and Sanchari Roy. 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, March 2013. URL http://ideas.repec.org/p/cep/stieop/42.html.

SiegfriedK. Berninghaus, Karl-Martin Ehrhart, and Marion Ott. A network experiment in continuous time: The influence of link costs. *Experimental Economics*, 9(3):237–251, 2006. ISSN 1386-4157. doi: 10.1007/s10683-006-9125-1. URL http://dx.doi.org/10.1007/s10683-006-9125-1.

Andre Beteille and Mysore Srinivas. 212. networks in indian social structure. *Man*, 64:pp. 165–168, 1964. ISSN 00251496. URL http://www.jstor.org/stable/2796572.

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

Erwin Bulte, Andreas Kontoleon, John List, Ty Turley, and Marteen Voors. When economics meets hierarchy: A field experiment on the workings of the invisible hand. Technical report, 2012.

Steven Callander and Charles R. Plott. Principles of network development and evolution: an experimental study. *Journal of Public Economics*, 89(8): 1469–1495, August 2005. URL http://ideas.repec.org/a/eee/pubeco/v89y2005i8p1469-1495.html.

A. Colin Cameron, Jonah B. Gelbach, and Douglas L. Miller. Bootstrap-Based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics*, 90(3):414–427, August 2008. URL http://ideas.repec.org/a/tpr/restat/v90y2008i3p414-427.html.

Lorenzo Casaburi, Michael Kremer, Sendhil Mullainathan, and Ravindra Ramrattan. Harnessing ict to increase agricultural production: Evidence from kenya. Working paper, 2014.

Damon Centola. The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–1197, 2010. doi: 10.1126/science.1185231. URL http://www.sciencemag.org/content/329/5996/1194.abstract.

Damon Centola. An experimental study of homophily in the adoption of health behavior. *Science*, 334(6060):1269–1272, 2011. doi: 10.1126/science.1207055. URL http://www.sciencemag.org/content/334/6060/1269.abstract.

Damon Centola and Michael Macy. Complex contagions and the weakness of long ties1. *American Journal of Sociology*, 113(3):702–734, 2007.

Gary Charness and Matthew Rabin. Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869, August 2002. URL http://ideas.repec.org/a/tpr/qjecon/v117y2002i3p817-869.html.

Gary Charness, Luca Rigotti, and Aldo Rustichini. Individual behavior and group membership. *American Economic Review*, 97(4):1340–1352, September 2007. URL http://ideas.repec.org/a/aea/aecrev/v97y2007i4p1340-1352.html.

Yan Chen and Sherry Xin Li. Group identity and social preferences. *The American Economic Review*, pages 431–457, 2009.

Shawn A. Cole and A. Nilesh Fernando. The Value of Advice: Evidence from Mobile Phone-Based Agricultural Extension. Harvard Business School Working Papers 13-047, Harvard Business School, November 2012. URL http://ideas.repec.org/p/hbs/wpaper/13-047.html.

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

Margherita Comola and Marcel Fafchamps. Testing unilateral and bilateral link formation. *The Economic Journal*, pages n/a–n/a, 2013. ISSN 1468-0297. doi: 10.1111/ecoj.12071. URL http://dx.doi.org/10.1111/ecoj.12071.

Timothy G. Conley and Christopher R. Udry. Learning about a new technology: Pineapple in ghana. *American Economic Review*, 100(1):35–69, March 2010. URL http://ideas.repec.org/a/aea/aecrev/v100y2010i1p35-69.html.

Anna Conte, Daniela Di Cagno, and Emanuela Sciubba. Strategies in social network formation. Birkbeck Working Papers in Economics and Finance 0905, Birkbeck, Department of Economics, Mathematics & Statistics, June 2009. URL http://ideas.repec.org/p/bbk/bbkefp/0905.html.

Sergio Currarini and Friederike Menge. Identity, Homophily and In-Group Bias. Working Papers 2012.37, Fondazione Eni Enrico Mattei, May 2012. URL http://ideas.repec.org/p/fem/femwpa/2012.37.html.

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

Esther Duflo, Michael Kremer, and Jonathan Robinson. Nudging farmers to use fertilizer: Theory and experimental evidence from kenya. *American Economic Review*, 101(6):2350–90, 2011. doi: 10.1257/aer.101.6.2350. URL http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.6.2350.

Thard Dunning and Janhavi Nilekani. Ethnic quotas and political mobilization: Caste, parties, and distribution in indian village councils. *American Political Science Review*, 107:35–56, 1 2013. ISSN 1537-5943. doi: 10.1017/S0003055412000573. URL http://journals.cambridge.org/article_S0003055412000573.

Catherine C. Eckel and Philip J. Grossman. Managing diversity by creating team identity. *Journal of Economic Behavior & Organization*, 58(3): 371–392, November 2005. URL http://ideas.repec.org/a/eee/jeborg/v58y2005i3p371-392.html.

Marcel Fafchamps and Flore Gubert. The formation of risk sharing networks. *Journal of Development Economics*, 83(2):326–350, July 2007. URL http://ideas.repec.org/a/eee/deveco/v83y2007i2p326-350.html.

Falk and Kosfeld. It's all about connections: Evidence on network formation. *Review of Network Economics*, 11(3):1–36, September 2012. URL http://ideas.repec.org/a/bpj/rneart/v11y2012i3n2.html.

Ernst Fehr and Klaus M. Schmidt. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868, August 1999. URL http://ideas.repec.org/a/tpr/qjecon/v114y1999i3p817-868.html.

Andrew D Foster and Mark R Rosenzweig. Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy*, 103(6):1176–1209, December 1995. URL http://ideas.repec.org/a/ucp/jpolec/v103y1995i6p1176-1209.html.

Jacob K. Goeree, Arno Riedl, and Aljaz Ule. In search of stars: Network formation among heterogeneous agents. *Games and Economic Behavior*, 67(2): 445–466, November 2009. URL http://ideas.repec.org/a/eee/gamebe/v67y2009i2p445-466.html.

Benjamin Golub and Matthew O. Jackson. How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, 127(3):1287–1338, 2012. URL http://ideas.repec.org/a/oup/qjecon/v127y2012i3p1287-1338.html.

S. Goyal. *Connections: An Introduction to the Economics of Networks*. Princeton University Press, 2012. ISBN 9781400829163. URL http://books.google.co.uk/books?id=h4V250eVEKgC.

R. Guha. *India After Gandhi: The History of the World's Largest Democracy*. Ecco book. HarperCollins, 2008. ISBN 9780060958589. URL http://books.google.co.uk/books?id=2fvd-CaFdqYC.

Glenn W. Harrison and John A. List. Field Experiments. *Journal of Economic Literature*, 42(4):1009–1055, December 2004. URL http://ideas.repec.org/a/aea/jeclit/v42y2004i4p1009-1055.html.

Karla Hoff and Priyanka Pandey. Discrimination, social identity, and durable inequalities. *American Economic Review*, 96(2):206–211, May 2006. URL http://ideas.repec.org/a/aea/aecrev/v96y2006i2p206-211.html.

Kelsey B. Jack. Constraints on the adoption of agricultural technologies in developing countries. Literature review. agricultural technology adoption initiative, j-pal (mit) and cega (uc berkeley), 2013.

Matthew O. Jackson and Asher Wolinsky. A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44–74, October 1996. URL http://ideas.repec.org/a/eee/jetheo/v71y1996i1p44-74.html.

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

Dean Karlan, Markus Mobius, Tanya Rosenblat, and Adam Szeidl. Trust and social collateral. *The Quarterly Journal of Economics*, 124(3):1307–1361, August 2009. URL http://ideas.repec.org/a/tpr/qjecon/v124y2009i3p1307-1361.html.

Georg Kirchsteiger, Marco Mantovani, Ana Mauleon, and Vincent Vannetelbosch. Myopic or Farsighted? An Experiment on Network Formation. CEPR Discussion Papers 8263, C.E.P.R. Discussion Papers, February 2011. URL http://ideas.repec.org/p/cpr/ceprdp/8263.html.

Rachel Kranton, Matthew Pease, Seth Sanders, and Scott Huetter. Identity, group conflict, and social preferences. Working paper, 2012.

Pramila Krishnan and Manasa Patnam. Neighbors and extension agents in ethiopia: who matters more for technology dffusion. Working papers, International Growth Center, 2012.

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

John A. List. Testing Neoclassical Competitive Theory in Multilateral Decentralized Markets. *Journal of Political Economy*, 112(5):1131–1156, October 2004. URL http://ideas.repec.org/a/ucp/jpolec/v112y2004i5p1131-1156.html.

Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001. doi: 10.1146/annurev.soc.27.1.415. URL http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.soc.27.1.415.

Kaivan Munshi. Social learning in a heterogeneous population: technology diffusion in the indian green revolution. *Journal of Development Economics*, 73 (1):185–213, February 2004. URL http://ideas.repec.org/a/eee/deveco/v73y2004i1p185-213.html.

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

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

Bassel Tarbush and Alexander Teytelboym. Friending. 2014.

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

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

Angelino Viceisza. 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), 2012. URL http://ideas.repec.org/p/fpr/fsprac/7.html.

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

Toshio Yamagishi and Toko Kiyonari. The group as the container of generalized reciprocity. *Social Psychology Quarterly*, 63(2):pp. 116–132, 2000. ISSN 01902725. URL http://www.jstor.org/stable/2695887.

Daniel Zizzo. Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1):75–98, March 2010. URL http://ideas.repec.org/a/kap/expeco/v13y2010i1p75-98.html.

7 Appendix

7.1 Formal derivation of the rules

7.1.1 Rule 1

We assume that in T1 player i maximises expected payoff:

$$\max_{j} \ \pi_i(g + ij) \tag{11}$$

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 strategy.

Proposition 1. *Player i maximises $\pi_i(g + ij)$ by choosing a player j with the maximum value of $\mu_{ji}(g)$ in the network*

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 $\mu_{ji}(g)$. □

7.1.2 Rule 2

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

$$\begin{aligned} u_i(g + ji) - u_i(g) &= \pi_i(g + ji) - \pi_i(g) + \gamma \sum_{k \in N \setminus i} \pi_k(g + ji) - \pi_k(g) \\ &= \gamma \sum_{k \in N \setminus i} \pi_k(g + ji) - \pi_k(g) \\ &= \gamma f(g, g_{ji}) \end{aligned} \tag{12}$$

where $f(g, g_{ji}) = \sum_{k \in N \setminus i} \pi_k(g + ji) - \pi_k(g)$ We can express $f(g, g_{ji})$ in the following way:

$$f(g, g_{ji}) = \begin{cases} (\mu_{-ji} + 1)(\mu_{-ji} + 2)/2 & \text{if } j \in N_i(g) \\ (\mu_i + 1)(\mu_{-ji} + 1) & \text{if } j \notin N_i(g) \end{cases} \tag{13}$$

equation 13 can be derived in three steps.

Step 1. We define a component as a connected set of nodes such that no node has a link with an individual outside of the set. Also, when we refer to an individual k, we call $g_{kj} = 1$ an “outgoing" link and $g_{jk} = 1$ an “incoming" link. In the first step, we establish the following lemma:

Lemma 1. *In T2, when i has the turn, $i \cap N_i(g)$ is a component of the network*

Proof of lemma. In T2, when player i has the turn, player i has no incoming links and can have multiple outgoing links. Individuals accessed by paths starting from outgoing links are part of the set $N_i(g)$. Thus player i has no (outgoing or incoming) links with individuals outside of $i \cap N_i(g)$.

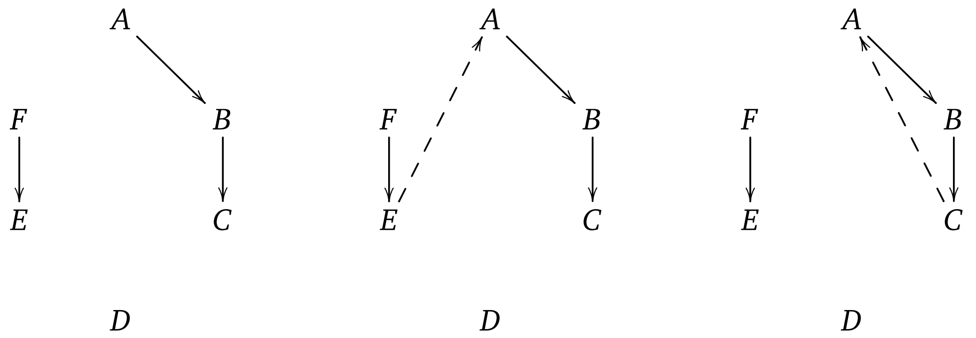
We will now further prove that no individual k in $N_i(g)$ has any outgoing or incoming link with an individual that is not in $i \cap N_i(g)$. First, in T2, an individual k in $N_i(g)$ has at most one incoming link g_{jk}. As k is in the set $N_i(g)$ this single incoming link can be either a direct link from player i to player k, or a link from a player z such that $i \to^g z$. Player z is hence part of the set $i \cap N_i(g)$. Second, in T2, an individual k in $N_i(g)$ may have multiple outgoing links. If a node k is in $N_i(g)$ and has a positive number of outgoing links, player i has a path to each of the players to whom k has a path. Each of these players is thus also in $i \cap N_i(g)$. So no player in $N_i(g)$ has either an incoming, or an outgoing link with a player who is not in $i \cap N_i(g)$. □

Lemma 1 is useful because it establishes that if j is not in $N_i(g)$, j has no path to any of the μ_i players in $N_i(g)$. Furthermore players who have a path to j also do not have a path to any of the players in $N_i(g)$. This fact will help us in the second step.

Step 2. If $j \notin N_i(g)$, a g_{ji} link increases the value of μ_j by 1 plus μ_i, as player j is now linked to player i and to all players to whom player i has a path. Lemma 1 tells us that none of the μ_i players in $N_i(g)$ was previously accessed by player $j \notin N_i(g)$. So a g_{ji} link increases player j's expected payoff by $\mu_i + 1$. Furthermore, all μ_{-ji} individuals in j's network benefit by the same amount. Hence the effect of a g_{ji} link to a player $j \notin N_i(g)$ on $f(g, g_{ji})$ is given by:

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

Figure 14: Effect of a g_{ji} link



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

Step 3. If, on the other hand, $j \in N_i(g)$, then a g_{ji} link creates a cycle where before the was a line from i to j. The effect on $f(g, g_{ji})$ is now reduced, because some of the information that j will pass on to the players who access j is redundant. Figure 14c shows an example. In the example, A has a path to C through player B. If player A links to player C, player C now has a path to A and B, and hence increases his value of μ_i by 2. Player B, however, increases his value of μ_i by 1 only, as he already has a link to player C. More generally, in T2, when i creates a link to a player j to whom i has a path, the first player in the path experiences an increase in μ_i of 1, the second player an increase of 2, and so on.. until all the μ_{-ji} players in the path from i to j are exhausted. Finally, player j's expected payoff increases by $\mu_{-ji} + 1$. Hence the effect of a g_{ji} link to a player $j \in N_i(g)$ on $f(g, g_{ji})$ is given by:

$$\begin{aligned} f(g, g_{ji}) &= \sum_{n=1}^{\mu_{-ji}+1} n \\ &= \frac{(\mu_{-ji}+1)(\mu_{-ji}+2)}{2} \end{aligned} \tag{15}$$

This completes the derivation. □

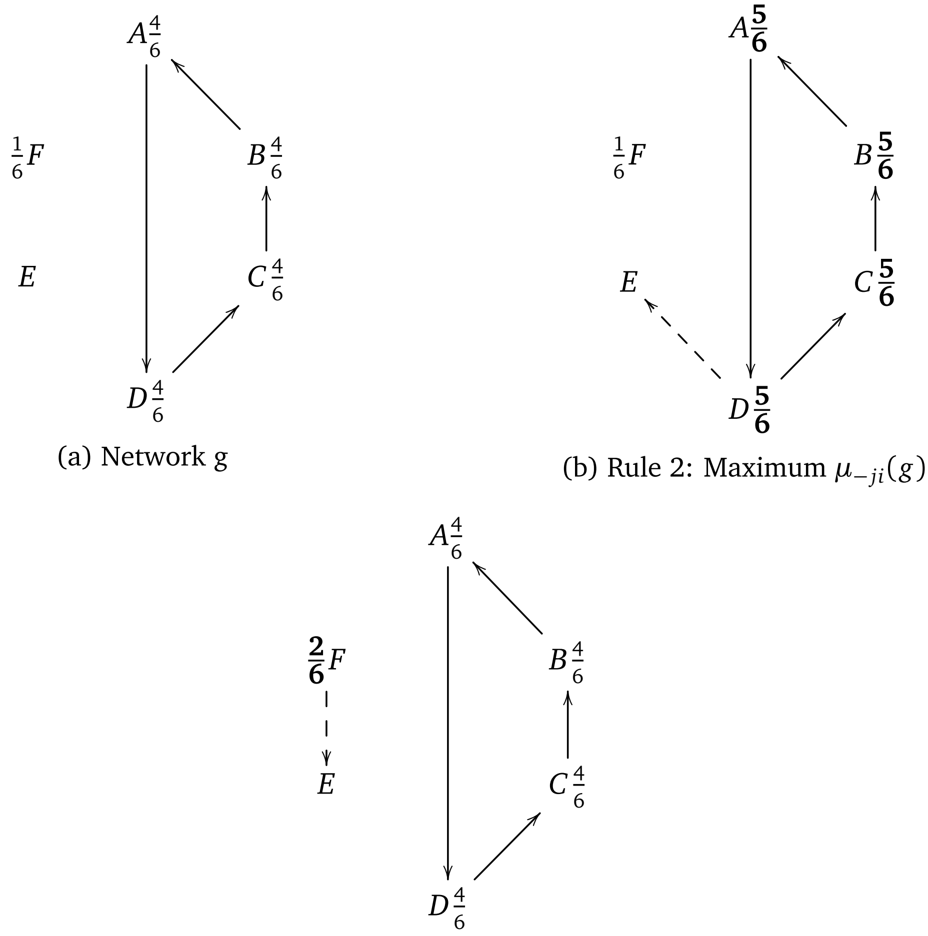
The simple heuristic of choosing the player with the maximum value of μ_{-ji} approximates well the more complicated rule which would always maximise 13 above.

1. When $N_i(g) = \emptyset$, $f(g, g_{ji})$ monotonically increases in μ_{-ji}. Link formation rule 2 always identifies the g_{ji} link that maximises $f(g, g_{ji})$.
2. When $N_i(g) \neq \emptyset$ and no player in the set of individuals with the maximum level of μ_{-ji} is also in the $N_i(g)$ set, link formation rule 2 maximises $f(g, g_{ji})$ for the

same reason as in point 1 above.

3. When $N_i(g) \neq \emptyset$ and at least some of the players in the set of individuals with the maximum level of μ_{-ji} are in $N_i(g)$, link formation rule 2 sometimes fails to maximise $f(g, g_{ji})$.

Figure 15: Example of a network



(a) Network g

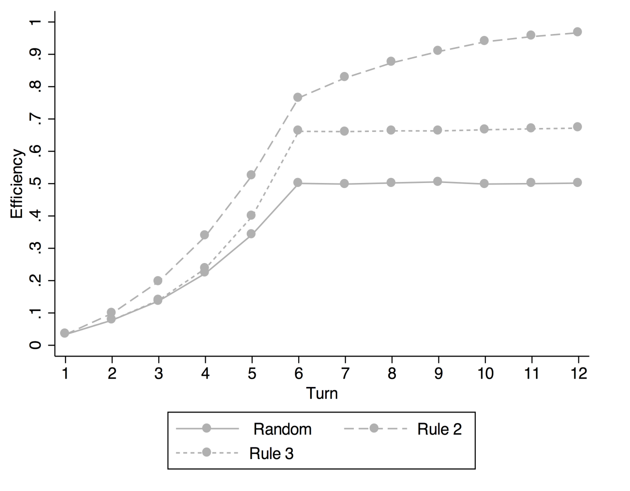
(b) Rule 2: Maximum $\mu_{-ji}(g)$

(c) Rule 23: Minimum $\mu_{ji}(g)$

Note. E has the turn. Probability of winning the prize reported next to each player. Panel (b) and (c) show network $g + ji$, where the new link is in accordance with either link formation rule 2 or link formation rule 3.

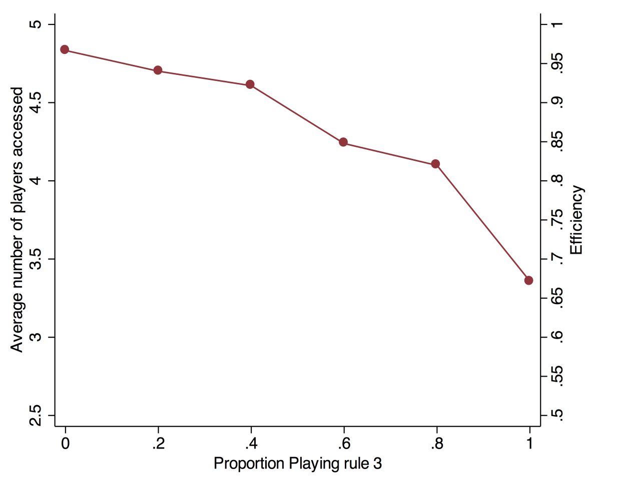
7.2 Figures

Figure 16: Simulated time series of average out-degree



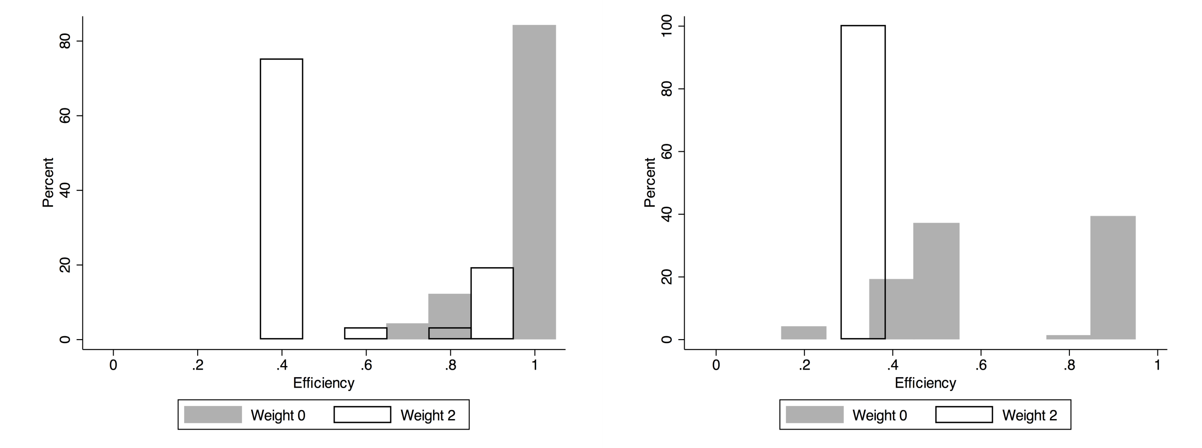
Note. Each rule is simulated 500 times.

Figure 17: Link formation process with mixed rules



Note. Simulation where min out-degree is played with probability p and max in-degree with probability 1-p. 500 simulation for each level of p.

Figure 18: Simulated effect of group identity concerns on network structure

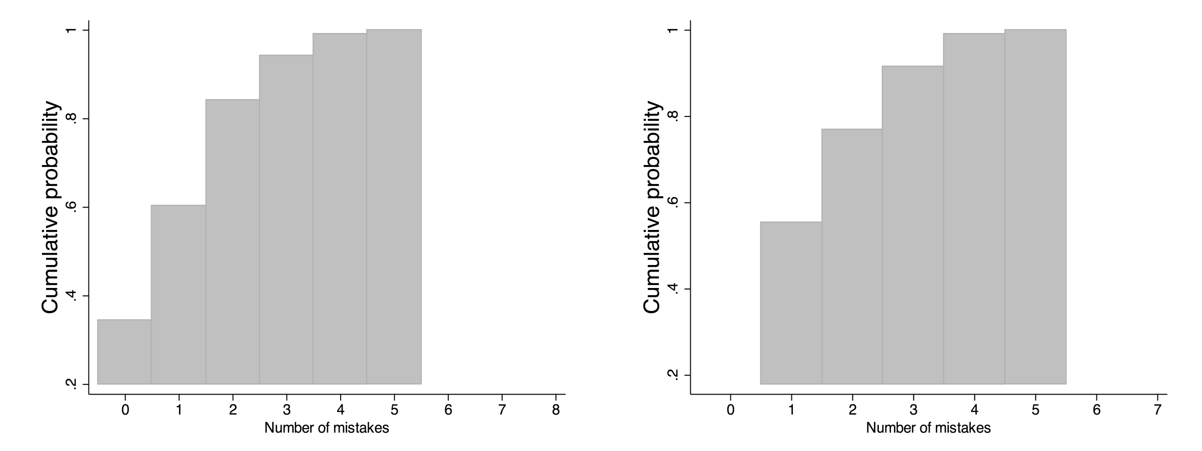


(a) Rule 2

(b) Rule 3

Note. 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 μ_{-ji} (panel a) or two units of μ_{ji} (panel b). We report results summarising 100 simulation for each of the 4 rules.

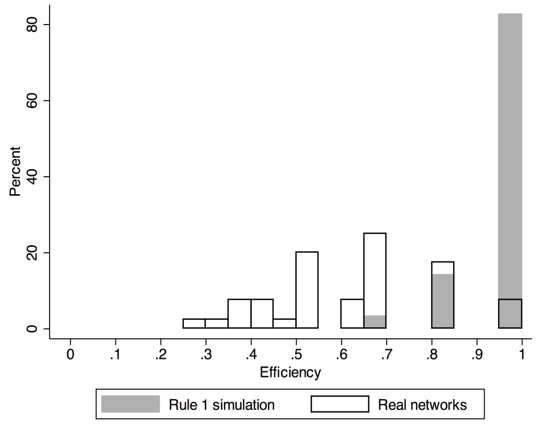
Figure 19: Cumulative distribution of mistakes in understanding questions



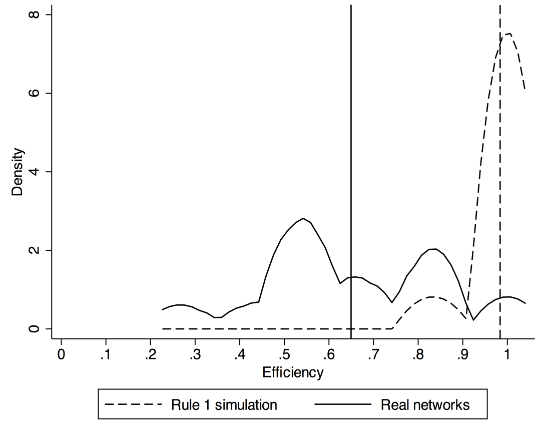
(a) T1

(b) T2

Figure 20: Efficiency in No-Identity Sessions and under Rule 1

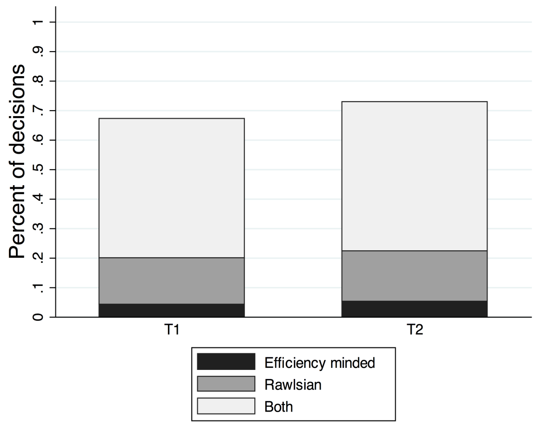


(a) Histogram

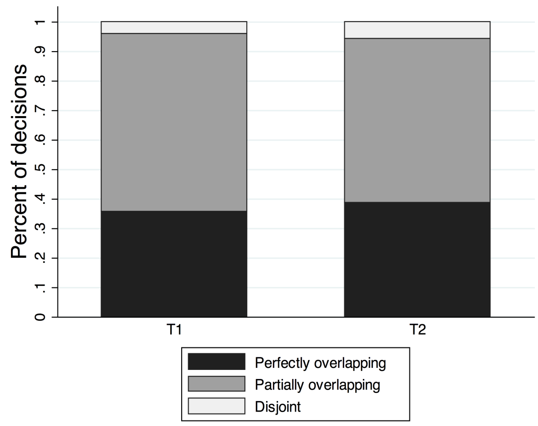


(b) Kernel density estimate and means

Figure 21: Overlap in decisions and in choice sets

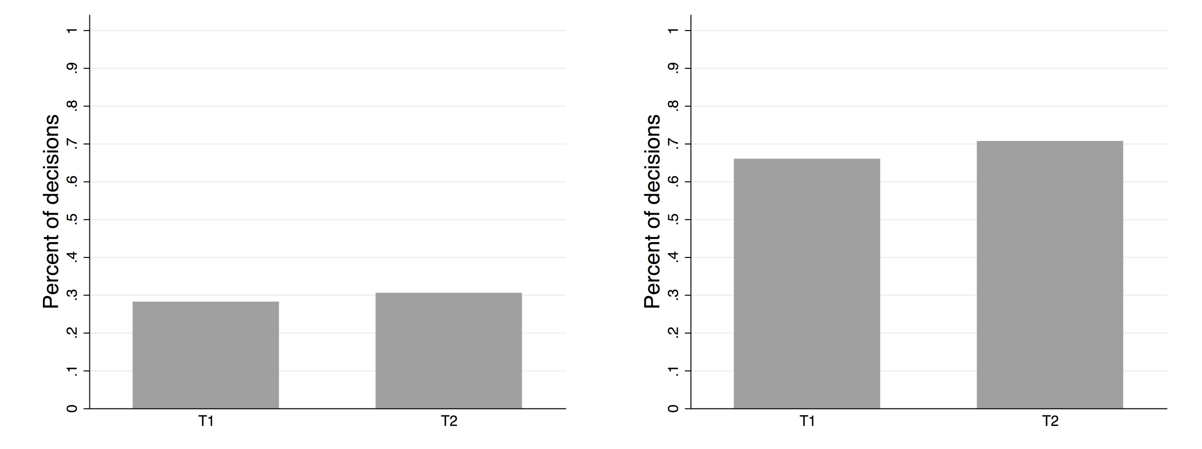


(a) Decisions



(b) Best response sets

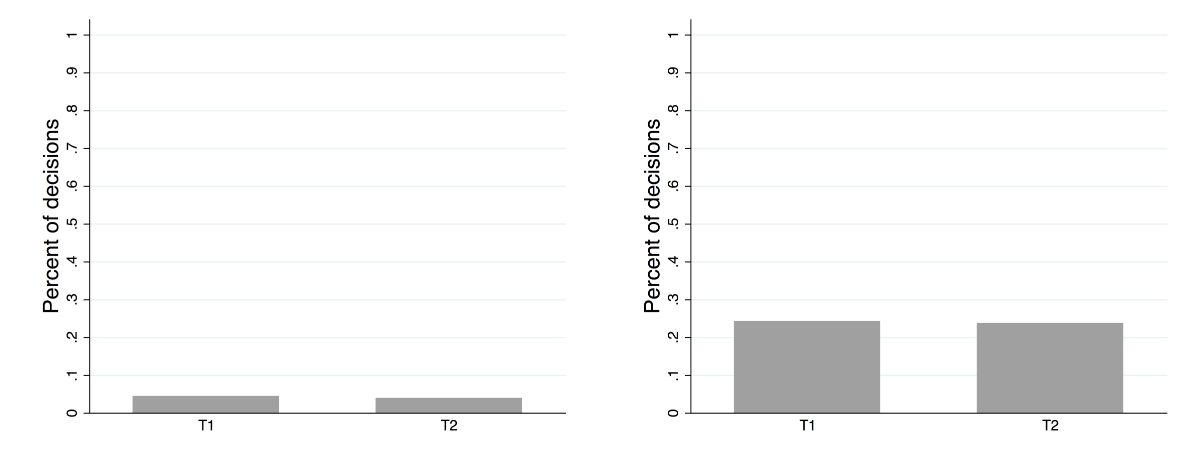
Figure 22: Links to the most popular player



(a) Decision consistent with archetypal rules (b) Decision not consistent with archetypal rule

Percentage of decisions consistent with rule 5. Only data for sessions with no knowledge of group identity is shown.

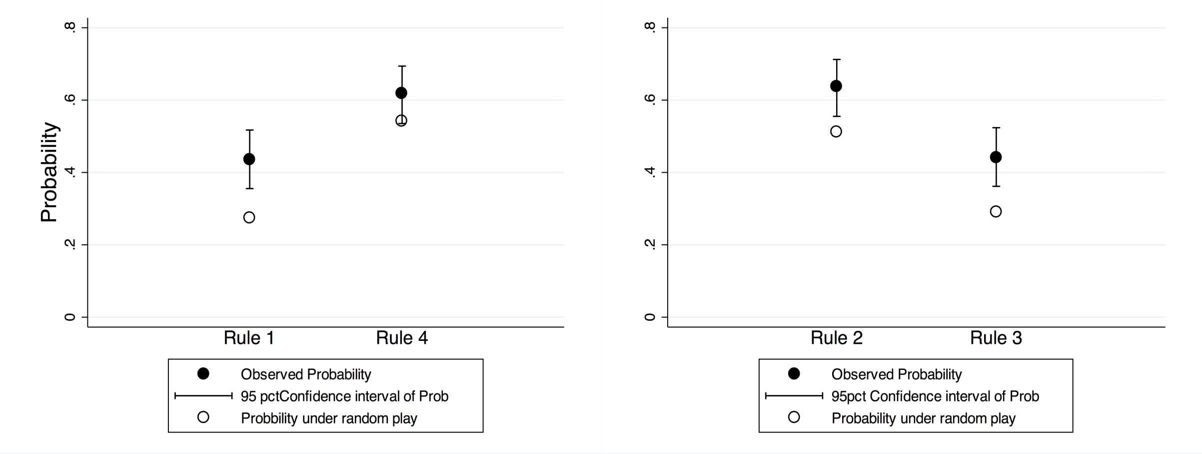
Figure 23: Reciprocal links



(a) Decision consistent with archetypal rules (b) Decision not consistent with archetypal rule

Percentage of decisions consistent with the rule 6. Only data for sessions with no knowledge of group identity is shown.

Figure 24: Probability that a decision is consistent with a particular strategy. Predicted Best Response Sets not fully Overlapping

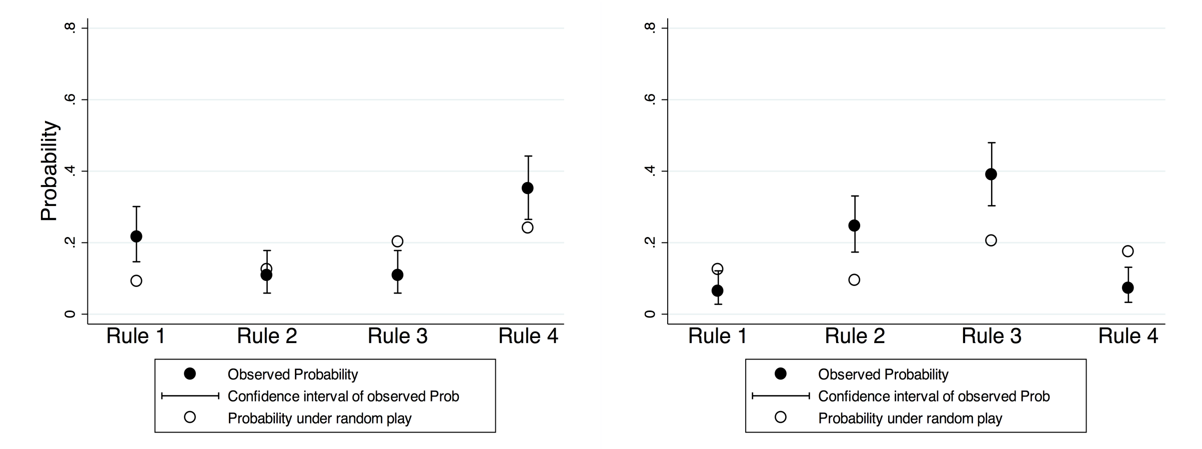


(a) T1

(b) T2

Note. T1 sessions: rounds in which the set of individuals with maximum μ_{ji} is not perfectly overlapping with the set of individuals with minimum μ_{-ji}. T2 sessions: rounds in which the set of individuals with maximum μ_{-ji} is not perfectly overlapping with the set of individuals with minimum μ_{ji}.

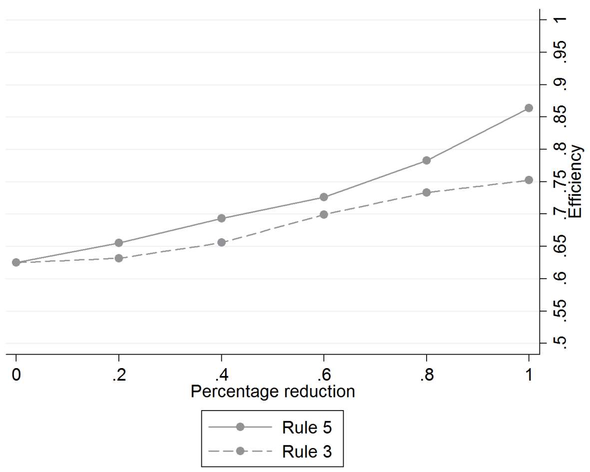
Figure 25: Probability that an individual plays twice consistently with a particular strategy



(a) T1

(b) T2

Figure 26: Efficiency simulation



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

Figure 27: Distribution of coin allocations to in-group partner

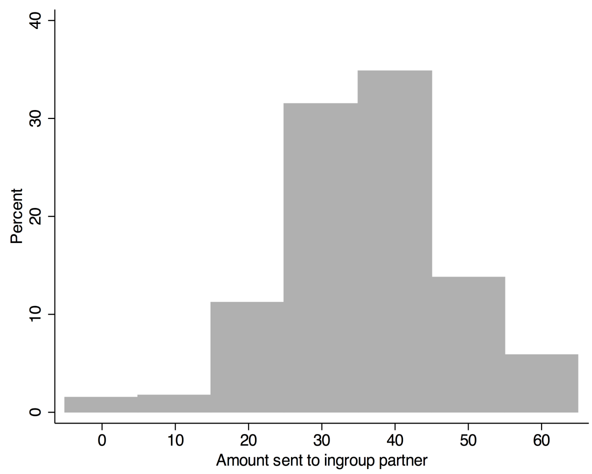
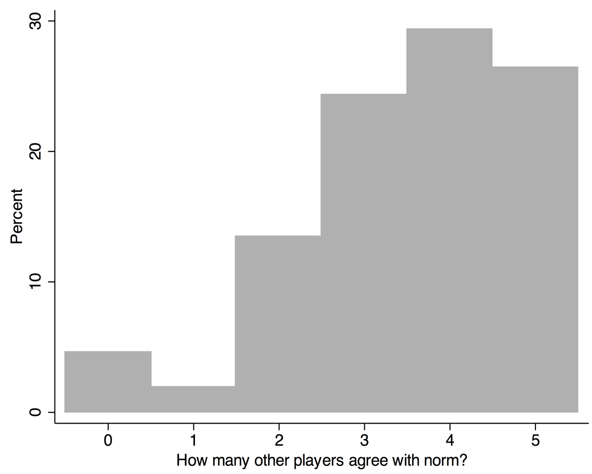


Figure 28: “How many of the other 5 players in the session do you think answered YES to the previous question?" Distribution of expectations



7.3 Tables

Table 6: 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 7: 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 8: Balance test: Identity in T1 sessions

	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 9: Balance test: Identity in T2 sessions

	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: *** $\longleftrightarrow$ 99%, ** $\longleftrightarrow$ 95%, * $\longleftrightarrow$ 90%. Standard errors clustered at the session level reported in parentheses.

Table 10: Out-degree and efficiency in final architectures

Treatment	**Average μ_i**	**Percent efficiency**
T1no	3.258	.652
T1id	2.895	.582
T2no	3.238	.648
T2id	3.333	.666
Total	3.167	.637

Table 12: Dyadic Linear Probability Model 7

	(1)	(2)	(3)	(4)
$\mu_{ji}(g)$	.054	.055	-.042	-.039
	(.001)***	(.001)***	(.001)***	(.006)***
$\mu_{-ji}(g)$	-.033	-.031	.052	.051
	(.005)***	(.046)**	(.001)***	(.002)***
Const.	.373	.395	.264	.279
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 belonging to a non backward caste. Confidence: *** $\longleftrightarrow$ 99%, ** $\longleftrightarrow$ 95%, * $\longleftrightarrow$ 90%. Standard errors corrected for clustering at session level. P-values obtained with wild bootstrap-t procedure reported in parentheses.

Table 11: Summary statistics of allocation, expectations and norms

Variable	Obs	Mean	Std. Dev.	Min	Max
Allocated to in-group	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

"in-group bias in allocation" is a dummy equal to 1 if the player has allocated more than half of the endowment to an in-group partner. "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 13: Dyadic Linear Probability Model 7

	(1)	(2)
$\max\left(\mu_{ji}\right)$	.121	
	(.006)***	
$\min\left(\mu_{-ji}\right)$	-.0004	
	(.535)	
$\max\left(\mu_{-ji}\right)$		.122
		(.012)**
$\min\left(\mu_{ji}\right)$		.059
		(.102)
Reciprocal_{ij}	-.093	-.010
	(.002)***	(.2)
Most chosen_j	-.031	-.011
	(.153)	(.342)
Obs.	910	940
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 14: Linear Probability Model 8: in-group links

	(T1)	(T2)
Identity session$_s$	.117	.046
	(.055)**	(.056)
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 15: Linear Probability Model 8: Efficiency-minded links

	(T1)	(T2)
Identity session$_s$	-.0003	-.080
	(.065)	(.062)
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 16: Linear Probability Model 8: Rawlsian links

	(T1)	(T2)
Identity session$_s$	-.013	-.038
	(.055)	(.055)
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 17: 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 18: 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: *** $\longleftrightarrow$ 99%, ** $\longleftrightarrow$ 95%, * $\longleftrightarrow$ 90%. Standard errors clustered at the session level reported in parentheses.

Table 19: Linear Probability Model: Understanding in T1

	max (μ_{ji})	min (μ_{-ji})	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 strategy indicated in the heading. Confidence: *** $\longleftrightarrow$ 99%, ** $\longleftrightarrow$ 95%, * $\longleftrightarrow$ 90%. Standard errors clustered at the session level reported in parentheses.

Table 20: Linear Probability Model: Understanding in T2

	$\max\left(\mu_{-ji}\right)$	$\min\left(\mu_{ji}\right)$	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
Obs.	41	41	41	41	41

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