Can People Form Links to Efficiently Access Information?

A. Stefano Caria* and Marcel Fafchamps† ‡

May 2019

Abstract

We investigate how individuals form social connections to access information. In our link formation game of one-way-flows, the myopic best response is to link to the node with the highest informational reach, which depends on indirect links. This strategy leads to the efficient circle network. We find that myopic best response predicts the links that subjects form. However, some subjects target high-degree nodes instead, neglecting indirect connections. This reduces network efficiency. We obtain similar findings when subjects link to transfer information to others. Using a minimal group treatment, we also find evidence of in-group homophily in link formation.

Keywords: Network formation, myopic best response, homophily, India.

JEL codes: D85; D90; O12.

*University of Bristol. Email: stefano.caria@bristol.ac.uk. Address: Department of Economics, University of Bristol, 12 Priory Rd, BS8 1TU, Bristol, UK

†Stanford University. Email: fafchamp@stanford.edu. Address: Freeman Spogli Institute for International Studies, Encina Hall Room E113, 616 Serra St, Stanford, CA 94305-6044.

‡We would like to thank Johannes Abeler, Antonio Cabrales, Francesco Cecchi, Giacomo De Giorgi, Alain De Janvry, David Huffman, Edoardo Gallo, Jessica Goldberg, Sanjeev Goyal, Kelsey Jack, Rachel Kranton, Julien Labonne, Karen Macours, Jeremy Magruder, Kaivan Munshi, Esteban Ortiz Ospina, Simon Quinn, Elisabeth Sadoulet, Bassel Tarbush, Mateusz Ujma and seminar participants at NEUDC, PACDEV, CSAE, CESifo, Barcelona GSE Summer Forum and FERDI workshop on learning for adopting for useful comments and suggestions. We are indebted to Gaurav Puntambekar for his outstanding management of the project. All mistakes remain naturally ours. This study is an output from research funding by the UK Department for International Development (DFID) as part of the iiG, a research programme to study how to improve institutions for pro-poor growth in Africa and South-Asia. The views expressed are not necessarily those of DFID. Stefano Caria further acknowledges financial support from the Economic and Social Research Council, UK.

A core insight from network theory is that the diffusion of information depends on the structure of social networks (Goyal, 2011; Banerjee et al., 2013). For example, we know that the circulation of information is facilitated when networks are densely connected, and hindered when they are split into distinct communities (Bala and Goyal, 2001; Golub and Jackson, 2012). The mechanisms that create different network structures, however, are poorly understood. Part of the difficulty arises from the fact that, even in network formation problems of moderate size, the decision tree faced by each decision maker is so large that deriving an optimal forward-looking strategy is quite complex.

An important class of theoretical models has sought to address this difficulty by proposing sequential link formation models in which each agent plays a myopic best response strategy (MBR), defined as forming the links that maximise the subject's utility in the current network (Bala and Goyal, 2000), without considering how these links may affect the future play of others (Jackson and Watts, 2002). Little evidence however exists on whether and when people actually follow such strategies.

To fill this knowledge gap, we experimentally investigate a simple network formation game in which MBR is intuitive and naturally leads to an efficient outcome. MBR is not the only plausible strategy to follow, however. People may rely on heuristics instead, that is, on strategies that follow a simple rule of thumb. In games of information sharing, such a heuristic may be to link to the highest degree node: if everyone follows this heuristic and information flows two-way, the resulting network has a star shape which is known to be efficient for many games. Since we wish to establish whether our subjects can play MBR, we select an experimental game in which linking to the highest degree node is not a good heuristic. This enables us to test whether subjects adopt MBR when their heuristic is ill-suited.

Our experimental design also serves another important purpose: to test whether human subjects take into account the *indirect* connections of their information partners when they form new links. While many formal and informal networks have a star shape[1], where a single node aggregates all information, highly centralised network structures are not always possible. For example, using data from Banerjee et al. (2013)

[1]For example, formal online exchange networks such as Google, Facebook, or Amazon. Further, Galeotti and Goyal (2010) report evidence suggesting that a number of important informal learning networks (online and offline) have a core-periphery structure, which is closely related to the star network.

we show in the next section that in villages in rural India – the setting of our study – there are no global information aggregators. A likely reason is that serving as the information aggregator at the centre of a star imposes large costs that few are willing to bear (Caria and Fafchamps, 2017) and requires information processing capabilities that few people have. If the social network cannot be expected to take a perfect star shape, individuals should condition their link formation decision on the *indirect* connections of the nodes to which they link. The question is whether they are capable of doing so.

Our experimental design is chosen to investigate these two research questions simultaneously: do people follow MBR; and do people take indirect links into account when forming a link.[2] Since we do not know *a priori* what preferences subjects have, we perform the analysis with two games that have a well-defined MBR for two different types of generic preferences: selfish and altruistic.[3]

We also investigate the distortion that social identity can bring to the formation of efficient information sharing networks (Akerlof and Kranton, 2000). Empirical studies have documented that people have a strong tendency to interact with socially similar individuals (McPherson et al., 2001; Currarini et al., 2009; Golub and Jackson, 2012). Less is known on whether people are prepared to compromise private or aggregate welfare in order to restrict social interaction to their in-group. To investigate this possibility, we introduce a minimal group treatment to ascertain whether group identity distorts link formation and reduces information sharing.

To achieve our objectives we conduct a laboratory experiment in which links enable subjects to observe the information others have. In the first treatment, which has a selfish MBR, subjects are invited to sequentially create one social link to observe others' information. This information is non-rival: all the subjects who learn the information receive the same payoff. The MBR is to link to the node with the highest reach, that

[2]Our experiment, on the other hand, is not well suited to study network formation in contexts where some subjects are willing and capable to serve as information aggregators at the centre of the star, or where subjects interact only in small tight-knit groups. Our experiment is also not designed to study the formation of networks that emerge to enforce informal transactions (Greif, 1993; Chandrasekhar et al., 2018). While in the real world these two types of networks often overlap, an advantage of our laboratory setting is that we are able to observe link formation decisions that are only motivated by information acquisition and information transmission.

[3]In the recent literature, altruistic preferences are also referred to as 'efficiency-minded' (Charness and Rabin, 2002).

is, the node who *directly or indirectly* accesses the largest amount of information. Information flows one way, without decay, and each subject can only form one link. These design features prevent subjects from forming a direct link to all other players: subjects must prioritise to whom they link. They also ensure that players can easily identify the player with the highest reach.[4] If everyone plays according to MBR, the resulting network is a circle and information reaches all players. In the second treatment, the link that a subject forms serves only to give their information to another person, not to obtain information from them. The efficient network is also a circle, but in order to reach it using an MBR, subjects must behave altruistically.

To study the effect of social identity, we randomly divide players into two arbitrary groups, and we vary whether players are informed of other subjects' group affiliation. This social identity treatment arm is crossed with the other two treatments. Experimental designs based on minimal arbitrary groups have the key advantage that the effect of social identity is not confounded by other factors that may be correlated with group membership. Previous research suggests that artificially created minimal groups are sufficient to activate group identity and lead subjects to discriminate against out-group individuals as much as they would do in real groups (Tajfel, 1981; Brewer, 1999; Yamagishi and Kiyonari, 2000; Charness et al., 2007; Akerlof and Kranton, 2010; Goette et al., 2012; Lane, 2016).

We implement our design with a population of male farmers in rural India. This population has a large need for information about agricultural technology, but has limited access to media or government advice services. As a result, farmers often search for relevant information by asking for advice from better informed neighbours (Comola and Fafchamps, 2014). Indirect connections play a key role in this process of information diffusion (Banerjee et al., 2013).[5] Since we restrict the sample to men (as they are most likely to have decision power over agricultural technologies — see for example Majumder and Shah (2017)), an important caveat is that our results do not necessarily apply to women, who may face different constraints and incentives when forming

[4]When information flows two-way with decay and multiple links are allowed, calculating the optimal strategy becomes substantially more complex, even in small networks.

[5] Using experimental data from a micro-finance awareness campaign in Indian villages, Banerjee et al. (2013) show that targeting the campaign on farmers with more *direct* connections to other villagers does not improve overall diffusion; what does is targeting farmers who have more *direct and indirect* connections.

social networks (Beaman et al., 2018).

The main results of our experiment are as follows. We find that, in all treatments, MBR is highly predictive of individual link-formation decisions. Half of the observed linking choices are consistent with an MBR either to obtain information for oneself (in the first treatment) or to disseminate information to others (in the second treatment). However a sufficiently large number of subjects deviate from these strategies so that only about 10 % of the experimental sessions converge to a circle. On average, the networks formed by participants generate expected payoffs that are 35 % lower than those generated by the circle network.

To identify what drives subjects away from efficiency, we look at deviations from MBR. The biggest efficiency loss comes from subjects who link to the node with the largest number of direct links. Since it is hard to rationalise such behaviour as part of an MBR or more sophisticated forward-looking strategy, we suspect that it is a manifestation of a common social heuristic: some subjects pick an action that seems appropriate in a general information-sharing context, instead of working out a best response for the game at play (Gabaix and Laibson, 2005; DellaVigna, 2009). We validate this conjecture in several ways. We first note that, if the conjecture is correct, linking to the highest degree node should become more common over time as the complexity of the network increases and mental resources are depleted. By the same reasoning, this action should also become more frequent when identifying the optimal link is harder. This is indeed what we find in the data. We also provide some evidence that subjects who link to popular individuals outside of the experiment – and can thus be expected to have this heuristic – are also those who link to high degree nodes in the experiment.

Regarding our social identity treatment, we are able to verify that randomly dividing people in minimal groups triggers in-group norms in the link formation game. When we disclose information about group membership, subjects form more links with their in-group. However, this does not generate an additional loss of efficiency. The reason appears to be that subjects exhibit an in-group preference only when choosing between in-group and out-group nodes that are in an equivalent network position.

Our findings complement the literature in several ways. We contribute to the study of efficiency in network formation (Jackson and Wolinsky, 1996; Bala and Goyal, 2000) by showing that, in a simple link formation game, (i) MBR predicts subjects' decisions, but (ii) deviations from MBR are sufficiently frequent to prevent the formation of the efficient network and to generate a loss in expected payoffs of about 35 %. Some subjects follow an inappropriate heuristic instead of relying on a simple myopic best response. This indicates a need for theoretical work on behavioural network formation. In particular, our evidence suggests to 'stress-test' MBR network models to ascertain the extent to which their theoretical predictions are robust to behavioural deviations from simple MBR.

Our findings also cast some doubt on the ability of human subjects to form efficient networks when an information aggregator – e.g., a for-profit company such as Google or a government agency – does not exist. Our evidence shows that, even in the simplest information sharing environments, some people make linking decisions that decrease welfare for themselves and others. As a result, efficiency in information diffusion depends on the extent to which common heuristics fit strategic realities. As the evidence provided by Banerjee et al. (2013) demonstrates, they often do not.

We also contribute to the lab experiment literature on network formation. Experiments published to date have explored the role of inequity aversion (Goeree et al., 2009), coordination (Berninghaus et al., 2006), and far-sightedness (Callander and Plott, 2005; Conte et al., 2009; Kirchsteiger et al., 2016). We add to this literature by studying an experiment where (i) the MBR is based on indirect connections, and (ii) playing MBR naturally leads to the efficient network. A separate strand of the experimental literature has documented that degree is a strong predictor of play in games of strategic complements (Gallo and Yan, 2015) and strategic substitutes (Rosenkranz and Weitzel, 2012; Charness et al., 2014). We show that degree is also an important predictor of link-formation decisions, even when ill-suited to the context. Falk and Kosfeld (2012) – who study a game of unilateral, one-way-flow link formation based on Bala and Goyal (2000) – is the experiment most closely related to ours. Our experiment differs from theirs along several important dimensions. First, we focus on a population that relies heavily on informal social networks for accessing relevant information (Comola and Fafchamps, 2014). Second, we include several features that minimise coordination issues and eliminate the need for computation, thereby reducing the risk that inefficient behaviour is due to strategic or computational complexity.[6] We find that, in our experiment, an efficient network is reach in about 10 % of the sessions, compared to half

[6] In our design, links are added to the network one at a time and players are allowed only one link, so that the only cost of a link is the opportunity cost of not forming another link.

the time in Falk and Kosfeld (2012). This illustrates how tight bounds on efficiency can emerge when small groups of players rely on ill-suited heuristics. Thirdly, we include a social identity treatment, something that is absent from their work.

A final contribution is to the literature on homophily and identity economics more generally (Akerlof and Kranton, 2000). Empirical studies of social networks document that individuals interact more with people from the same group (McPherson et al., 2001; Rogers, 2003; Berg et al., 2017). Most of these studies, however, cannot convincingly distinguish the desire to link to similar individuals – homophily means 'love of the same' – from linking opportunities, which are typically lower with out-group members. We improve on this literature by implementing an alternative test of homophily based on variation in the information available about the group affiliation of other individuals. Our results confirm that individuals in our experiment display a desire to link to in-group members, in spite of the fact that group membership is randomly assigned within the experiment.

1 Design

Our experiment is designed to focus on our two main questions of interest: do subjects follow a myopic best response; and do subjects take indirect links into account. To this effect, we opt for a network formation game in which information flows one way without decay. One key advantage of this design is that it produces simple predictions about myopic best response for different types of preferences that can be put to the test. Studying myopic best response would be more challenging in a two-way-flow model: if information travels through the network without decay, any link to the giant component is equally beneficial (e.g., Bala and Goyal (2000)); and if information travels with decay, calculating the optimal link is beyond the calculating capacity of most subjects (e.g., Jackson (2010)).

In the game design we consider one node receives a valuable non-rival signal, but which node receives the signal is not known a priori. The objective of the game is to be connected, directly or indirectly, to as many nodes as possible, to maximise the probability of being connected to the player who receives the signal and hence to observe their valuable information. We thus obviate issues related to the aggregation of multiple signals and biased inference (Enke and Zimmermann, 2017; Chandrasekhar et al., 2015). By allowing players to form a single, one-way connection, we rule out the emergence of a node serving as information aggregator and make indirect connections salient. This serves to focus attention on our second question – do people take indirect links into account.

Given our experimental design, the informational *reach* of a player i can naturally be defined as the number of players whose signal i observes, either directly or indirectly (i.e., through other nodes). A player with a reach of three thus 'observes' three other players. We also define the *in-reach* of a player i as the number of individuals who directly or indirectly observe i's signal. A player with an in-reach of two is thus 'observed' by two other players. If the number of nodes in the network is n and we normalise the value of the signal to 1, the expected payoff π_i of a player i is given by a linear function of his reach:[7]

$$\pi_i = (reach_i + 1)/n \ . \tag{1}$$

In practice, to implement the idea of a valuable signal, we give a monetary prize to the player that receives the signal and to the other subjects that reach this player in the network. The prize is worth 100 Indian Rupees (5.2 USD at purchasing power parity), which is about 85 % of the average daily wage paid by the NREGA public work program in that area in 2012-2013. In the instructions, we explain that the prize can be thought of as a valuable piece of information about a new agricultural technology that farmers share across the social network.[8]

1.1 The link-formation game

We play the link-formation game with groups of six players. Each player can form one link and one link only. The game starts with no links yet formed. In the first round, each player takes a turn in which he can link to one other player. Links are formed unilaterally – without requiring the consent of the other player. They are recorded on a network map displayed on a white board visible to all players and updated after every turn. This means that, when a player's turn comes, he can see all the links already formed before choosing who to link to. A number of design features ensure sequential updating takes place without breaking anonymity.[9] In the second round, each player

[7]We present additional notation and definitions in the Appendix.

[8]All instructions are available in the online Appendix.

[9]Participants record their decisions on a personal game sheet. Cardboard screens ensure participants cannot see what other players write on their game sheet. Further, all participants have to update

is given a chance to rewire their link, i.e., to drop the link they formed in round one and link to another node instead. The order of players in each round is randomly assigned.[10]

Information flows one way over the link that each player adds to the network: either *to* or *from* the player forming the link.[11] In treatment T1, a player who links to a node reaches all the information that node accesses. This means that if, for instance, player A links to player B, then player A receives the monetary prize if B wins, but not vice versa. In the left panel of Figure 1 this is represented with an arrow pointing from A to B, meaning that A 'observes' B. In Figure 1, by connecting to player B, A also reaches players C and D and hence receives the prize if C or D are declared winners. Player A's reach is given by the number of players in set $\{B, C, D\}$. In this example player A reaches three players in the network and thus has four chances out of six of receiving the prize – one by himself, and three more chances through B, C and D.

In treatment T2, information flows in the reverse direction: instead of accessing information, the player who forms the link transfers all their information to the other player. For instance, if A links to B, then B receives the prize whenever A wins it, but not vice versa. This is illustrated in the right panel of Figure 1 where an arrow pointing from B to A means that B observes A's information. Player A's in-reach is given by number of players in the set $\{B\}$, that is, the set of nodes that observe the information available to A.

Since the objective of our experiment is not to test whether subjects are able to read network graphs (Dessi et al., 2016), information about reach (in T1) or in-reach (in T2) is provided to subjects; they do not have to calculate it themselves.[12] After two rounds the

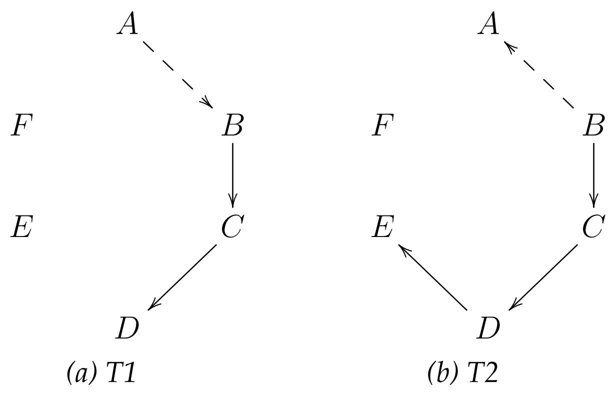
their game sheet at every turn of the game, making it impossible to infer which player has the turn. A set of regressions, available upon request, shows that the connections between participants outside of the experiment are not a significant predictor of the links chosen during the game. This confirms that anonymity was maintained in the experiment.

[10] Participants are informed that the order of play is chosen randomly, but do not know the particular order of play which has been drawn for their session. In both rounds, players have the option not to form any link. This option is used extremely rarely.

[11] This does not rule out the possibility that two players directly observe each other. However, for this to be the case, each player in the pair has to create a link to the other person.

[12] 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, in T1 (T2) the software produces a table with the reach (in-reach) of each player in the current network. This number is written next to the respective

Figure 1: Links in T1 and T2



game ends, at which point a winner is drawn at random and the network configuration at the end of round two determines who will receive the prize.

1.2 Group identity

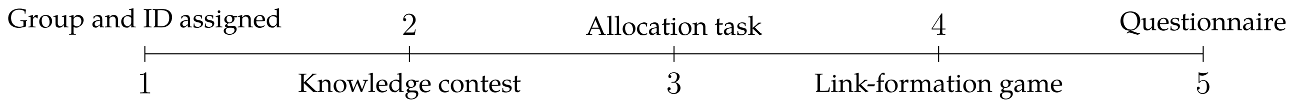
The experiment includes a minimal group identity treatment as follows. At the beginning of the experiment we randomly assign subjects to one of two groups. We then disclose information about group membership in some experimental sessions and not in others. The group identity treatment is cross-cut with T1 and T2, resulting in four treatment cells. In the first two cells, T1no and T2no, individuals have no information about the group affiliation of the other players. In the last two cells, T1id and T2id, players' group identity is displayed publicly on the network map by identifying players with the symbol of their group.[13]

Group affiliation is assigned by inviting each subject to pick a card from an urn. We do not use art or sport preferences to assign subjects into groups because these preferences may correlate with other individual characteristics. To strengthen group identity, the groups compete in an agricultural knowledge quiz. While there are no monetary prizes associated with this activity, we expect competition to strengthen group identification: contests of this kind have been used in previous experimental studies to make group affiliation more salient (Eckel and Grossman, 2005).

player ID on the white board, immediately after the network map has been updated with a new link.

[13] In the experimental instructions the groups are called the mango and pineapple groups. On the network map, these groups are identified, respectively, with a circle and a triangle.

Figure 2: Order of activities in the experiment



In our quiz, players answer three questions related to local weeds, pesticides, etc. If all players in a group answer all questions correctly, the group obtains one "point" (see Table C3 in the Appendix for summary statistics). Every player can thus have a strong marginal impact on the group's outcome: if the player makes one or more mistakes, the whole group fails to gain the point. Players answer the agricultural knowledge questions before the link formation game, but the results are disclosed publicly only after all parts of the experiment are completed. After results are disclosed, each group that has won a point is congratulated with an applause.

In addition, the points won by a group in a session are added to those accumulated by subjects assigned to the same group in all previous experimental sessions, and participants are informed of which group is leading the knowledge contest across sessions. We do not disclose any information about group performance before the experiment so that players cannot make any inference about the personal identity or characteristics of other group members.[14] However, participants are informed at the start of the session that group performance in the knowledge contest will be disclosed after all other activities have been completed.

We check the impact of the group identity manipulation in two ways. First, we ask players to take part in an incentivised allocation task. In this task, players divide 30 Rupees between an in-group and an out-group recipient. Recipients are randomly drawn from the participants in the *following session* of the experiment. This task is played right after the questions on agricultural knowledge. Second, we ask a number of questions about norms and expectations at the end of the experiment. These questions give us a self-reported, unincentivised measure of whether the manipulation has activated social norms related to group identity. The sequence of activities is summarised in Figure 2.

[14]For example, failure to get the point in the agricultural contest may be interpreted by a player as a negative signal on the ability of the other people in his group.

Table 1: Summary of treatments

	No identity	Identity
T1	T1no	T1id
T2	T2no	T2id

1.3 Discussion

Comprehension. We take several steps to ensure that participants understand the rules of the game: we develop simple standardised instructions that are read out to participants; we double translate all written materials; and we rely on physical randomisation devices (Barr and Genicot, 2008; Viceisza, 2012). We also test players' comprehension before the game starts. In particular we check whether players: (i) understand the direction of links and the implication this has for winning the prize; (ii) are aware of the possibility and implications of indirect connections; and (iii) are able to identify, in a simple network map, either (in T1) the link that maximises their reach, or (in T2) the link that benefits the largest number of subjects as well as the worst-off subject. After players answer the comprehension test, our enumerators check their answers and give further explanations to correct mistakes. To further increase comprehension, we run a trial round of the link-formation game before the main game is played. Answers to the comprehension test thus give a lower bound for the final level of comprehension. Overall, we find that, in both T1 and T2, more than 50 % of players make at most 1 mistake and about 80 % of players make at most 2 mistakes.[15] In the results section, we present evidence suggesting that our findings are not driven by poor comprehension of the rules of the game.

Side payments. In order to make side payments unlikely, personal identity is never disclosed during the experiment and payments are disbursed privately.

Wealth effects. Both the allocation task and the link-formation game are incentivised with monetary payments. In the allocation task individuals choose how to split a sum of money between two subjects in a future session of the experiment. Thus, the

[15] We ask 8 questions in T1 and 7 questions in T2. In Table C1 in the Appendix we report the share of correct answers for each question. The exact wording of the questions can be found in the experimental materials.

allocation decision does not affect the payoff of the decision maker or any other player in the same session. This rules out unintended influences across the two tasks created by endogenous shocks to players' wealth.

Experimenter demand effects. These arise when subjects respond to implicit cues embedded in the experimental design in an attempt to please the experimenter (Zizzo, 2010; de Quidt et al., 2018). To minimise these concerns, we rely on a between-subjects design because it is less vulnerable to demand effects (Zizzo, 2010). We also refrain from revealing players' experimental group identity in the trial round to avoid making unintended suggestions about how we expect players to use this information. A final source of experimenter demands may come from the visual reminder of the network reach of each player and from the explanations given before the game. These may encourage subjects to use efficiency-enhancing strategies. In the light of this, our finding that efficient networks are rarely obtained becomes even more compelling.

Information aggregator. Our design intentionally focuses on networks that lack an information aggregator in order to study whether subjects take informational reach into account when forming a link. While there is evidence that many social networks include high degree individuals who recirculate information, in our specific context information networks typically do not have a global information aggregator, i.e., a star node to whom all others are linked. To illustrate this point, we rely on data on social networks in Indian villages collected by Banerjee et al. (2013).[16] In none of these villages do the authors find an individual who exchanges information directly with all other nodes in their village: the most central individual in the average village shares information directly with only 5.6 % of other villagers, as shown in Figure B3. Furthermore, in about 80 % of villages the individual with the highest degree is not the individual with the highest reach (see Table C4 in the Appendix). These results also

[16]This data is well suited for our purposes. It is collected in rural India, in villages that are within a three-hours drive from the city of Bangalore with an average size of 200 households. Our sample includes villages at a similar distance from the city of Pune and with a similar size. Further, the data has unusually rich information about social interaction: the researchers have collected census data on all households in the village and have administered an in-detail survey to about 50 % of these households. They use this data to compile an adjacency matrix of farmers' information networks in each village. We plot the distribution of degree and reach in these networks in Figures B1 and B2 in the Appendix. We calculate reach by counting the number of connections up to five links away in the network. Finally, in Figure B4, we plot the structure of the largest component of the network in a number of selected villages.

hold at the level of sub-village communities.[17]

2 Predictions

In this section we formulate our predictions about individual decisions and about the overall structure of the networks that are formed in the experiment. In the online Appendix we present formal notation and proofs.

2.1 Myopic best response

In T1, the MBR for a selfish player is to connect to the partner who has the highest reach in the network. This allows him to be indirectly connected to the largest possible number of players at that point in the game. Consequently, a player who follows an MBR in treatment T1 will act according to this rule:[18]

Rule 1. *Connect to the player with the highest reach.*

In T2 sessions, players can only give the information they have to others. There is therefore no myopic best response for purely selfish players. An MBR nonetheless exists for players who have other-regarding preferences. A large body of evidence from controlled experiments shows that individuals care about the payoffs of others in systematic, heterogeneous ways (Charness and Rabin, 2002; Andreoni and Miller, 2002). Let us consider two types of social objectives that are particularly relevant in our setting. The first is the altruistic concern for total welfare discussed, for instance, in Charness and Rabin (2002):

$$u_i = \pi_i + \gamma \sum_{j \in N \setminus i} \pi_j. \tag{2}$$

N is the set of players in the game. For a player with such altruistic preferences, the MBR is to connect to the node with the maximum in-reach because this ensures that

[17] We identify communities using a popular community-detection algorithm due to Pons and Latapy (2006). Communities are composed of twelve nodes on average. The most central node in a community exchanges advice with an average of 38pct of the other nodes of the community.

[18] Links to the subject who has the turn are excluded from the count of the reach of the other players. This is because these links give no additional information to the subject who has the turn. In the online Appendix, we explain this point in more detail.

the information that the player has reaches the largest possible number of individuals.[19] Thus, for players who have preferences given by (2), the MBR in T2 is to form a link according to the following rule:

Rule 2. *Connect to the player with the maximum in-reach.*

The second social preference we consider is:

$$u_i = \pi_i + \gamma \min_{j \in N \setminus i} \pi_j. \tag{3}$$

This utility function captures a 'Rawlsian' concern for the player with the lowest welfare (Yaari and Bar-Hillel, 1984). The MBR for a Rawlsian player in T2 is to link to the least connected individual in the network to increase their payoff, i.e.:

Rule 3. *Connect to the player with the minimum reach.*

We can summarise the above observations into the following prediction:

Prediction 1. *If players follow an MBR: in T1 selfish players will link to the node with the highest reach; in T2 altruistic players will link to the node with the highest in-reach and Rawlsian players will link to the node with the minimum reach.*

For ease of exposition we will sometimes refer to rules 1 and 2 as the 'max' rules since both arise from a desire to maximise payoffs – one's own or that of the group. We refer to rule 3 as the 'Rawlsian rule'.[20]

2.2 Network efficiency

In this section we report the results of a set of simulations where we study the efficiency of the networks that are formed when individuals play according to the rules specified above. We define welfare as the sum of expected payoffs across players. We also define

[19] In the online Appendix, we show this formally and explain one qualification that applies to links that create a small circle.

[20] A third model of social preferences is that of inequality aversion (Fehr and Schmidt, 1999). Under inequality aversion, a player feels guilt towards players with a lower expected payoff and envy towards players with a higher expected payoff. An inequality averse player would not form a link in the first turn of a T2 session, because when the network has no links all players earn the same expected payoff. This prediction is strongly falsified by the data and we thus we do not consider this type of preferences in the discussion.

the efficient network as the network that maximises welfare. In both T1 and T2 this is the circle network, in which each player reaches the other five players. To normalize our efficiency measure, we redefine it as the average reach of the n players in network g divided by the average reach in the circle network:

$$\text{Efficiency}_g = \left(1/n * \sum_{i=1}^{n} \text{reach}_i\right) \Big/ 5 \,. \tag{4}$$

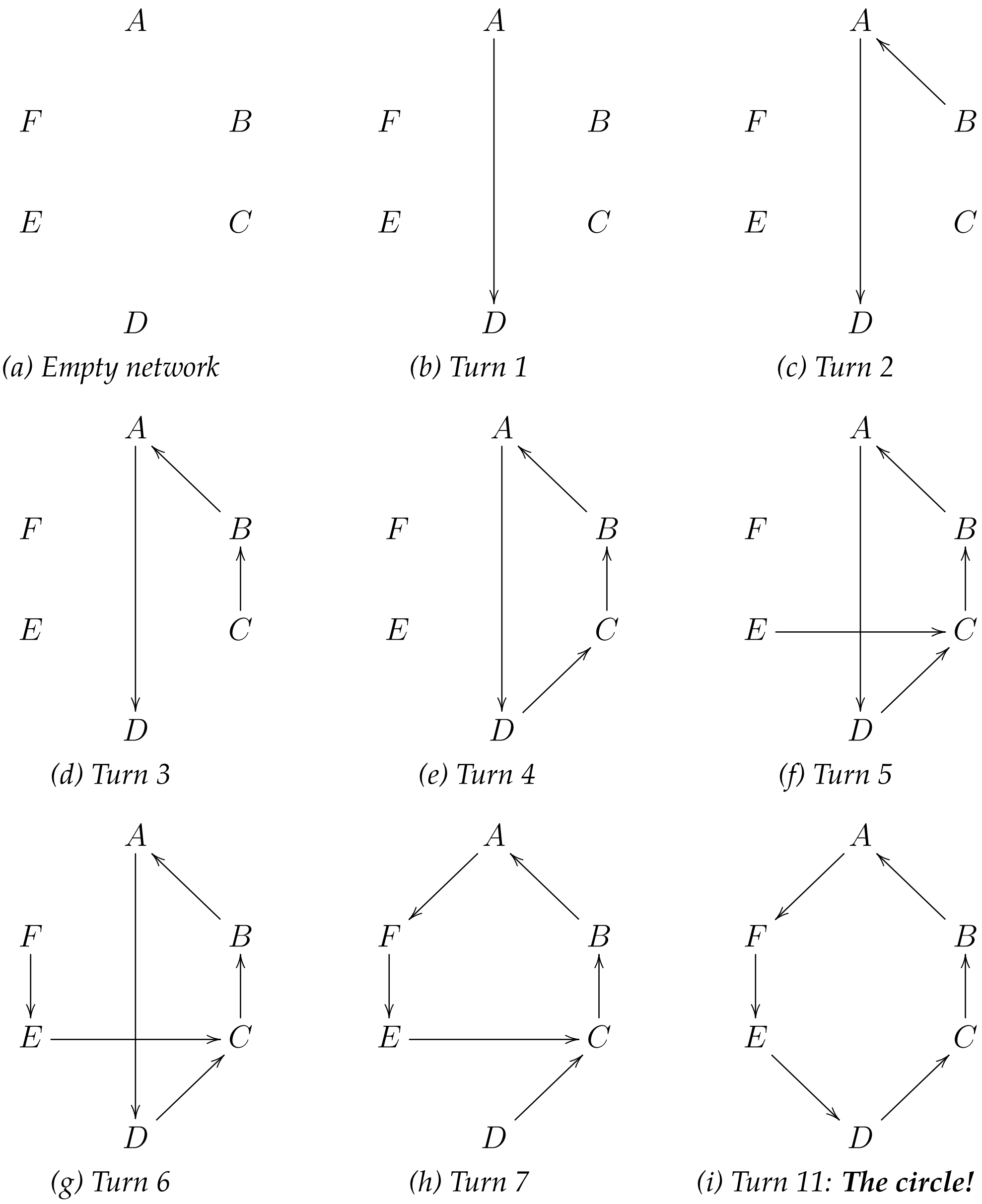
Our first set of simulations shows that when all players follow MBR rule 1 in T1, average efficiency is about 96 %. Figure 3 gives an example of how players can achieve the circle network within two rounds by playing according to rule 1. Under this rule, once the circle network is reached, no player wants to rewire his link.[21] This rule remains quite effective even when a small proportion of players link randomly. For example, average efficiency is about 80 % when 20 % of players choose links randomly and the remaining players follow rule 1.

Our second set of simulations studies efficiency in T2. We find that when all players play according to MBR rule 2 in T2, average efficiency is also about 96 %. This is not surprising since rule 2 generates a link-formation sequence that is symmetrical to that generated by rule 1 in T1. When all players play according to rule 3, on the other hand, network efficiency falls to 67 %, and when players link at random, average efficiency is about 52 %. Figure 4 depicts these results graphically.

Finally, we examine efficiency in sessions where a fraction p of players follow MBR rule 3 (the Rawlsian rule), and a fraction $1 - p$ of players follow MBR rule 2 (the max rule). Results show that efficiency decreases monotonically with p from a maximum of 96 % when $p = 0$ to a minimum of 67 % when $p = 1$. Figure B6 in the Appendix shows this graphically.

[21] In a very small number of cases the link formation process does not converge to the circle. This happens when two nodes have the same reach, in which case the MBR is to pick one of them at random. This sometimes leads to a situation where there is only one player who can form the circle network by re-wiring his link, but this subject has already played his second turn (by definition the MBR does not take future play into account). If we allow more rounds, the likelihood of this occurring becomes vanishingly small. For example, with three rounds rule 1 achieves 99 % efficiency.

Figure 3: Network evolution under rule 1 in T1



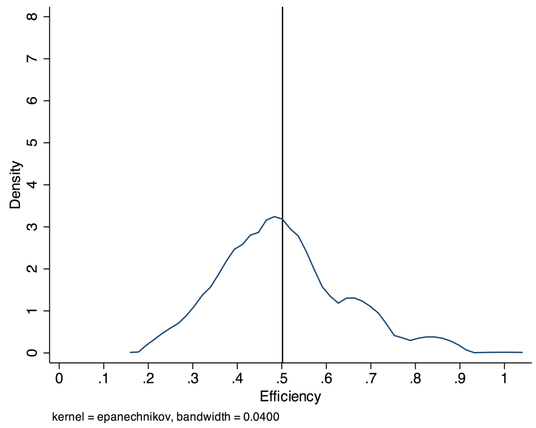
(a) Empty network *(b) Turn 1* *(c) Turn 2*

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

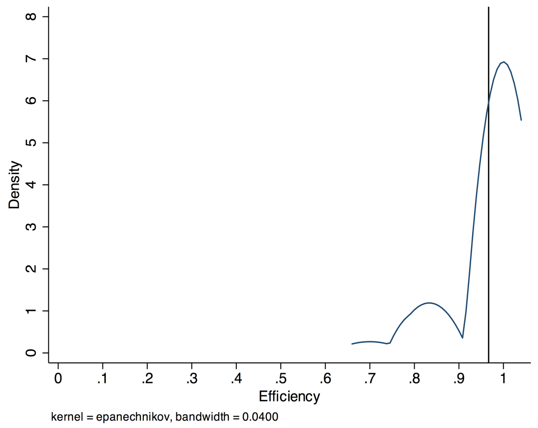
(g) Turn 6 *(h) Turn 7* *(i) Turn 11:* ***The circle!***

For ease of presentation, play is assumed to be in alphabetical order. All players in this simulation play according to 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.

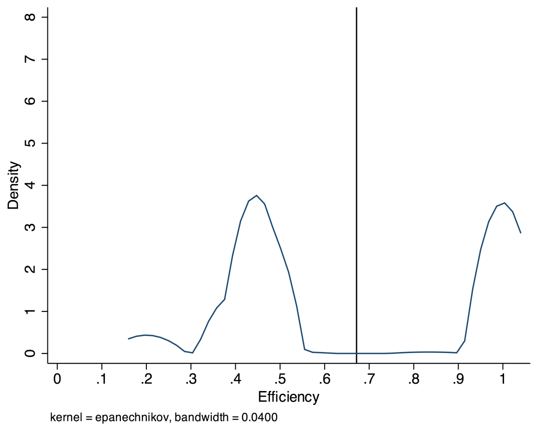
Figure 4: Efficiency in T2



(a) Random



(b) Rule 2: Maximum in-reach



(c) Rule 3: Minimum reach

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

2.3 Disclosure of group identity

We now discuss the effect of disclosing group identity in our game. Following Akerlof and Kranton (2000), we assume that players experience positive utility when they follow norms prescribed for members of their social group. Given the cultural context of India and the strong observed homophily in social interaction and marriage, we expect that making group identity salient should trigger a norm encouraging linking to an in-group node. If the utility obtained from following the in-group norm is high, players may sacrifice payoff maximisation in order to follow this norm. In this case, we would expect the disclosure of group identity to increase the *frequency* of in-group links, while limiting the use of MBR and consequently reducing the *efficiency* of the network.

To illustrate, imagine that when group identity is disclosed, players in T2 prefer to link with an in-group node rather than an out-group node whose in-reach is up to two units higher. Simulations show that in this case average network efficiency would fall to 53 %, which is only marginally above the average efficiency achieved through random links. Alternatively, suppose that players in T1 never link to an out-group node and use rule 1 to choose an in-group link. In this example, the network would converge to two circles of 3 players and the average reach would be 2, which corresponds to 40 % efficiency – much less than random linking. Figure B7 in the Appendix illustrates.

On the other hand, if the strength of the in-group norm is sufficiently low, the positive utility from following the norm will not outweigh the desire to maximise payoffs. In this case, the disclosure of group identity would not change the frequency with which players choose 'max' (or 'Rawlsian') links and hence would not affect the *efficiency* of the network. However, the *frequency* of in-group links would still increase. This is because when asked to choose between two nodes with the same position – one in their group and one not – players would prefer to link to the in-group node. These considerations yield the following prediction:

Prediction 2. *Disclosure of group identity generates networks characterised by (i) more in-group links and, if the in-group norm is sufficiently strong, (ii) lower efficiency.*

3 Data

We run the experiment in the Indian state of Maharashtra. We randomly sample from a census list of all villages in 4 'talukas' (sub-districts) of the Pune and Satara districts.[22] Villages in these sub-districts 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). We select study participants through door-to-door random sampling. On alternating days, we start sampling from the periphery of the village or from the centre of the village. We invite all male adult farmers who are encountered in the door-to-door visit until we have enough farmers to conduct all planned sessions.

Data collection took place between September and October 2013. In total, we completed 81 sessions with 486 subjects. We ran 20 sessions of T1no, T1id and T2id, and 21 sessions of T2no. In three of the sessions one participant left before the beginning of the link formation stage. This leaves us with 483 subjects, and a corresponding dyadic dataset with 4,800 dyads.[23] Table 2 summarises the number of observations in 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

At the end of the game, participants compile a short questionnaire. We hence have a small set of variables that describe the study sample. Average age is 43 years. 95 %

[22]We exclude from the sample large towns on the main highway of the district.

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

of participants are Hindu, 72 % do not belong to a scheduled caste, tribe or another backward caste (OBC), 28 % 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 regularly sharing information about agriculture with 11 other farmers.[24] From session 9 onwards, we ask each farmer whether he knows each of the other 5 participants and how often he has spoken with them in the previous 30 days. The density of links among participants in a session is very high: 87 % of participants know every other farmer in their session. On average a farmer speaks with each known participant on 13.5 of the previous 30 days. In table C5 in the Appendix we present a set of regressions that test whether covariates are balanced across treatments. We cannot find any statistically significant imbalance.

Table 3: Summary statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
Age	478	43.36	12.92	22	85
Hindu	456	0.95	0.22	0	1
Completed High School	465	0.28	0.45	0	1
Upper Caste	432	0.72	0.45	0	1
Land Owned (Ha)	474	4.08	4.68	0.1	50
Land Cultivated (Ha)	469	3.6	4.18	0.1	45
Information network size	427	10.91	8.94	1	60

4 Results

4.1 Network efficiency

We first investigate overall efficiency. Table 4 summarises average reach across treatments and the related measure of efficiency. On average, players in T1no and T2no are connected to 3.2 other individuals in the network. This determines a loss in expected payoffs, compared to the circle network, of about 35 %. We then look at the distribution

[24]When participants fail to answer a question or report an illegible answer, we code a missing value. This explains the changing number of observations in Table 3.

of network efficiency, pooling together all T1no and T2no sessions, and find that the efficient network is achieved in 4 sessions out of 41 (that is, in 9.76 % of the sessions).

Table 4: Reach and efficiency in final networks

Treatment	Average reach	Efficiency
T1no	3.258	0.652
T1id	2.895	0.582
T2no	3.238	0.648
T2id	3.273	0.666
All	3.167	0.637

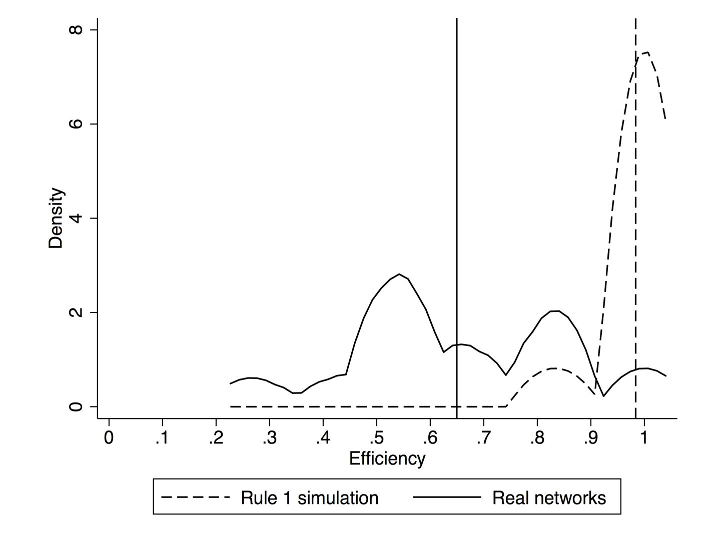
We compare the efficiency of the networks in our experiment to two benchmarks: the distribution of efficiency that we would observe if players follow a 'max' rule (rule 1 in T1 and rule 2 in T2) and the distribution of efficiency that we would observe if individuals choose their links at random. Figure 5 below and Figure B5 in the Appendix present this analysis. We find that the efficiency of the experimental networks is 31 percentage points below the average level achieved when all players follow a 'max' rule. A Wilcoxon rank-sum test using session-level data confirms that the difference between the distribution of network efficiency in our data and the simulated distribution is statistically significant at the 1 % level (Z = 12.08, p <.001). On the other hand, the efficiency of the experimental networks is higher than the average efficiency which random play would have achieved, by a significant 13 percentage points (Z = 4.62, p <.001).

We summarise our findings in this result:

Result 1. *The circle network is formed in about 10 % of the sessions. Expected payoffs in T1no and T2no are 35 % lower than in the circle network. Expected payoffs are consequently significantly lower than the payoffs that would be generated if everyone played according to the 'max' link-formation rule.*

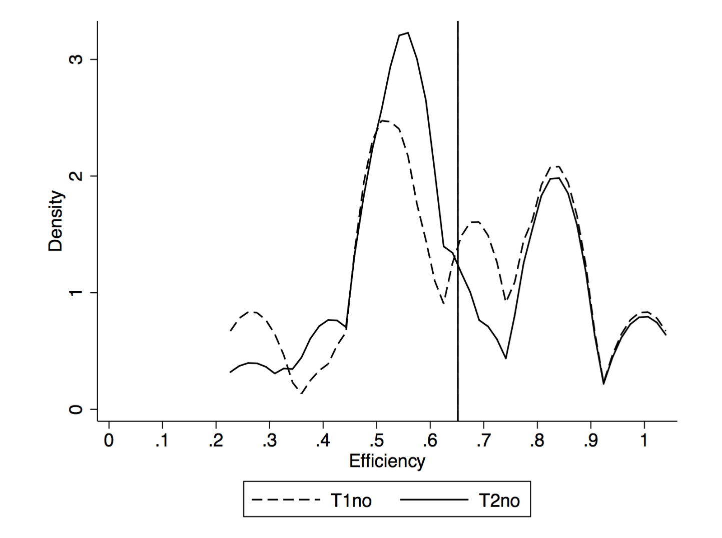
Further, we find that the direction of the flow of benefits associated with the links does not affect network efficiency. Average efficiency in T1no and T2no is very similar. A Wilcoxon rank sum test cannot reject the equality of the two distributions (Z = -0.11, p = 0.91). Figure 6 presents this result graphically.

Figure 5: Efficiency in real and simulated networks



Note. 'Real networks' include all sessions in T1no and T2no. 'Rule 1 simulation' includes networks simulated assuming all players play according to rule 1.

Figure 6: Efficiency in T1no and T2no sessions



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

We study the evolution of efficiency during the course of the experiment and find that low efficiency is not an artefact of truncation at 12 turns. Efficiency has no monotonic upward trend in either T1no or T2no, and efficiency at turn 12 is only a few percentage points higher than it was at turn 6.[25] Figure B8 in the Appendix illustrates. Falk and Kosfeld (2012), on the other hand, document strong learning dynamics and positive efficiency trends in their experiment.

4.2 Individual decisions

In this section we study individual decisions. We present evidence showing that players are significantly more likely to connect to partners who satisfy one of the link-formation rules compared to partners who do not. The majority of decisions are actually consistent with at least one of the rules. Further, we show that when players deviate from the rules, their links often target the 'most popular' player in the network. Simulation analysis reveals that this strategy is particularly detrimental to efficiency. A small number of links targeting the 'most popular' player substantially affects overall network efficiency.

To investigate whether the link-formation rules predict subjects' decisions we estimate the following dyadic regression model:

$$\text{link}_{ijrs} = \alpha + \textbf{Network Position}_{ijrs}\,\beta + \delta\,\text{round}_r + e_{ijrs}. \tag{5}$$

Link_{ijrs} is a dummy which takes a value of one if player i connects to player j in round r of session s. The vector of variables 'Network Position' describes the network position of player j when it is player i's turn to play in round r. We include all the variables suggested by the rules described above. For T2, we include a dummy for having the minimum reach and a dummy for having the maximum in-reach. For T1,

[25]Similarly, the length of the longest line in the network does not increase in the last 6 turns in T2no, and experiences only a very modest increase in T1no. Throughout the second round the average longest line in the network is composed of about four links. This means that, to reach efficiency, two links (33% of the total) will need to be rewired.

we include a dummy for having the maximum reach and, for completeness, we also include a dummy for having the minimum value of in-reach. Furthermore, we control for round specific effects. We estimate model (5) using OLS, correcting standard errors for arbitrary correlation at the session level. This technique requires at least 40 clusters (Cameron et al., 2008). When we run regressions with less than 40 clusters, we apply the wild bootstrap correction to p-values proposed by Cameron et al. (2008).

We report results in Table 5. As hypothesised, we find that in T1no connections are directed towards players with the maximum reach. Further, in T2no ties are directed towards players with the maximum in-reach and the minimum reach. The effects are statistically significant and economically meaningful. Players in T1no are 13 percentage points more likely to connect to a partner who has the maximum reach in the network. Players in T2no are 12 percentage points more likely to choose a player with maximum in-reach and 9 percentage points more likely to pick a player with minimum reach. A Wald test cannot reject the equality of these two coefficients. We summarise this analysis in the following result, which supports prediction 1:

Result 3. *In T1no players are more likely to form links with partners who have the maximum reach in the network. In T2no players are more likely to form links with partners who have the maximum in-reach and with partners who have the minimum reach in the network.*

We confirm the robustness of these results by running a specification that substitutes the dummies with the values of reach and in-reach. This allows players to make mistakes, while requiring larger mistakes to be less likely than smaller mistakes. Table C6 in the Appendix reports the estimates. Our findings from Table 5 are confirmed by this analysis. In T1no, for example, players are about 19 percentage points more likely to choose a partner with a reach of 4 than a partner with a reach of 0. Table C6 shows a further significant effect: in T1no players are more likely to establish a link with a partner with a lower in-reach.[26] To describe this behaviour, we define a fourth rule:

[26] A possible explanation for this result is that links carry some form of 'social value' for the person who receives the link. Individuals who choose partners with a low in-reach in T1no could thus be targeting the players who have accumulated the minimum social value in the game so far. We cannot provide a direct test for this interpretation. We have however some qualitative evidence in support of it. In the questionnaire administered at the end of the experiment, subjects 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 % of subjects answer yes to this question.

Table 5: *Links and partner network position*

	Link$_{ijr}$			
	(1)	(2)	(3)	(4)
max reach	.126	.124		
	(.001)***	(.005)***		
min in-reach	.020	.021		
	(.490)	(.564)		
max in-reach			.123	.128
			(.001)***	(.011)**
min reach			.091	.092
			(.010)**	(.026)**
Const.	.107	.099	.045	.033
	(.001)***	(.096)*	(.190)	(.426)
max reach = min in-reach	6.80	4.99		
	(.012)**	(.055)*		
max in-reach = min reach			0.38	0.35
			(.542)	(.551)
Obs.	1200	910	1260	940
Sample	T1no	T1no	T2no	T2no
Cluster N	20	20	21	21
Controls		✔		✔

The dependent variable is a dummy which takes a value of one if player i connects to player j in round r of the game. Each regression contains round fixed effects. 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 comprehension questions and dummies for having completed secondary education, for being Hindu, and for belonging to an upper caste. Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%. Standard errors are corrected for clustering at session level. P-values obtained with wild bootstrap-t procedure reported in parentheses. The second panel reports the F statistics (and p-value in parenthesis) for a Wald test of the equality of coefficients.

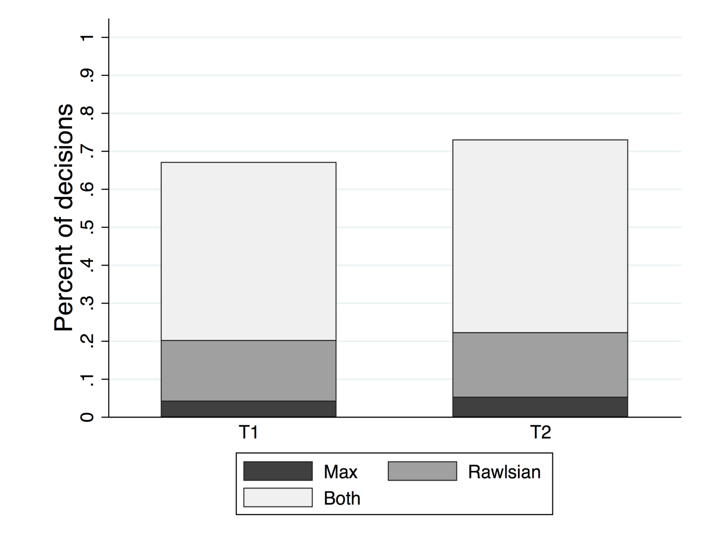
Rule 4. *Connect to the player with the minimum in-reach.*

We will refer to rule 3 in T2 and rule 4 in T1 as the 'Rawlsian' rules. Rule 4, however, does not have significant predictive power in the original specification (see the first two columns of Table 5).

Overall, about 70 % of decisions in each treatment are consistent with at least one of the rules (Figure 7). In particular, 51 % of decisions in T1no are consistent with rule 1 and 63 % with rule 4. In T2no, 56 % of decisions are consistent with rule 2 and 68 % with rule 3. The percentage of players following 'max' and 'Ralwsian' rules is very similar across treatments. This explains why players achieve similar levels of efficiency in T1 and T2.

The two percentages we report for each treatment do not add up to 70 % because the sets of possible partners satisfying different rules often overlap. We show the extent of

Figure 7: Decisions and rules



this overlap in Figure B9 in the Appendix.[27] Intuitively, overlaps are frequent because in a line network (i) the last individual has both the minimum reach and the maximum in-reach, and (ii) the first individual has both the maximum reach as well as the minimum in-reach. The line network is formed frequently in the game. Further, as shown in Figure 7, players frequently choose a partner who satisfies both of the rules.

4.2.1 Decisions that are not consistent with link-formation rules 1-4

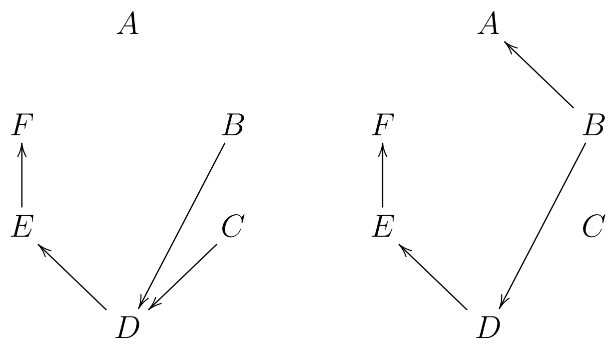
What about the remaining 30 % of links that are not consistent with the link-formation rules we have specified? We explore two possible additional rules:

Rule 5. *Connect to the player who has been chosen by most other players in the current network.*

We refer to the individual who has been chosen by the largest number of players as the 'most popular' player in the network. In T1, this is the player that is *directly observed* by the largest number of individuals (using network terminology, this is the player with the maximum *in-degree*). In the left panel of Figure 8, for example, player

[27] Consider turn t when player i has to play. Let BR_t^1, BR_t^2, BR_t^3 and BR_t^4 be the sets of players who, from the point of view of player i, satisfy link-formation rules 1, 2, 3 and 4, respectively. For T1, we focus on BR_t^1 and BR_t^4 and define three mutually exclusive cases: fully overlapping ($BR_t^1 \cap BR_t^4 = BR_t^1 = BR_t^4$), disjoint ($BR_t^1 \cap BR_t^4 = \emptyset$), and partially overlapping (not disjoint and not fully overlapping). For T2, we focus on BR_t^2 and BR_t^3 and similarly define the three cases.

Figure 8: The most popular player in T1 and T2



(a) T1 *(b) T2*

D is the most popular player in the network, while players B and C have the highest reach.[28] In T2 the most popular player is the player that *directly observes* the largest number of individuals (the player with maximum degree). In panel (b) of Figure 8 this is player B. Rule 5 is consistent with models of *non-strategic* link formation where new links preferentially target well-connected nodes (Barabási and Albert, 1999).

We also specify a rule that reflects a simple norm of procedural reciprocity:

Rule 6. *Connect to the player who has chosen you in a previous round.*

We find that about 65 % of the links that are not consistent with the link-formation rules are directed to the 'most popular' player in the network. Reciprocal links are not as common: they account for 18 % of those decisions. Figure B10 shows these results graphically.[29]

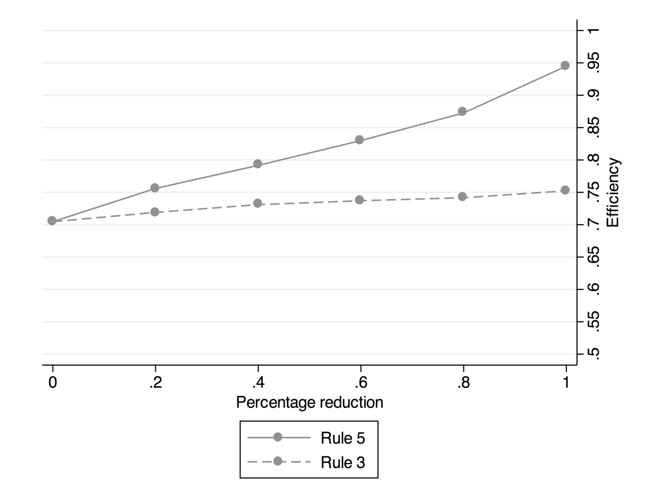
We use simulation analysis to study how to improve the efficiency of the network formation process. We find that reducing the proportion of links that are targeted to the 'most popular' player would generate large efficiency gains. On the other hand, reducing the proportion of 'Rawlsian' links would only have a limited impact on efficiency. In particular, we simulate a link formation process where 54 % of decisions

[28] Player D is observed by two players (B and C). All other players are observed by either zero or one player. Thus, player D is the most popular player in the network. Players B and C, on the other hand, have the highest reach in the network (they reach players D, E and F).

[29] While rule 5 is consistent with most decisions that do not follow the link-formation rules 1-4, it does not significantly predict links in the *full* sample. We show this using regression analysis reported in Table C7 in the Appendix.

are consistent with rule 1, 16 % with the rule 4 and the remaining 30 % with the 'most popular' player rule.[30] We then switch to rule 1 increasing proportions of decisions originally assigned to follow rule 4, keeping the proportion of rule 5 decisions fixed. We repeat the same exercise for rule 5: we switch to rule 1 increasing proportions of decisions assigned to follow rule 5, keeping the proportion of rule 4 decisions fixed. The results are stark: switching all rule 5 decisions to rule 1 delivers an efficiency gain of 25 percentage points, while an equivalent reduction of 'Rawlsian' decisions results only in a 5 percentage points gain in efficiency.[31] Figure 9 illustrates. In Figure B11 in the Appendix we repeat this exercise with different assumptions about the baseline proportions of decisions following the different rules. Results are not qualitatively affected.

Figure 9: Efficiency simulations



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

[30]These figures reflect the decisions in our data, with two simplifying assumptions: (i) all decisions that are consistent with both rule 1 and rule 4 are assumed to follow the rule 1, (ii) all decisions that are not consistent with link-formation rules 1-4 are assumed to follow rule 5.

[31]Figure B6 in the online Appendix shows what would happen if switch all rule 5 decisions to rule 3 (that is, if we run a simulation where 46 % of decisions follow rule 3 and 54 % follow rule 1): network efficiency in such scenario would be above 90 %, which corresponds to a 20 percentage points gain.

4.2.2 Why do players target the most popular individual?

We hypothesise that linking to the most popular individual in the network is a heuristic rule that individuals use to economise the cognitive cost of making an optimal decision. We use insights from recent models of thinking in complex environments to develop a test of this hypothesis (Gabaix and Laibson, 2005; DellaVigna, 2009). In these models, attention is scarce and individuals allocate more attention to a problem when information is salient and when the cost of doing a mistake is large. Thus, if our hypothesis is correct, links to the most popular players should become *less* likely when the myopic best response (the player with the maximum reach in T1 and the player with maximum in-reach in T2) is in a visually salient position of the network map. For example, when he is placed next to the decision maker.[32] Further, individuals may decide not to pay attention to the reach of other players when the difference between maximum and minimum reach is small, and hence the cost of picking the wrong player is limited. Lastly, links to the most popular player should be more frequent in the second round of the game, when the network map is more complex and mental resources have been depleted (Baumeister et al., 1998).[33]

We find qualified support for these predictions. In Table C8 in the Appendix we report three sets of regressions. In the first three columns we study whether the most popular player is less likely to be chosen when the myopic best response is in a salient position of the network map. The salient position is that adjacent to the person forming a link at that point in the game. We find that, when the myopic best response is in a salient position, players are about 10 (3) percentage points *less* likely to link to the most popular player in T2no (T1no). This difference is on the margin of significance in T2no (p-value of 0.108) but is not significant in T1no. In the next three columns of Table C8, we test whether players are less likely to connect to the most popular node when the difference between nodes with maximum and minimum reach increases. Coefficients have the hypothesised negative sign, but they are small in magnitude and insignificant. Finally, in the last three columns of Table C8, we show that in T1no players are almost 8 percentage points more likely to choose the most popular partner in round 2, when the network structure is more complicated and mental resources more depleted. The

[32] See Dessi et al. (2016) for a study of network cognition.

[33] In the second round of the game, the network map always has 6 links. In all turns of the first round, on the other hand, the map has less than 6 links. This should reduce complexity in the first round.

effect is marginally significant (p-value of 0.093).

We also investigate whether the people who link to the most popular player in the experiment tend to be connected to popular players in their networks outside of the experiment. To perform this analysis, we exploit the data about the real networks of our subjects. We ask each person (from session 9 onwards) to report the total number of individuals with whom they regularly exchange information about agriculture. Let us call this the player's 'degree'. Further, we ask each participant to indicate whether they interact with each of the other five people in the session. These two pieces of information enable us to calculate the average degree among the social ties of each participant. We regress a variable capturing the number of times a subject played according to the most popular player rule on this network statistics and report the results of the analysis in Table C9 in the Appendix. We find a quantitatively meaningful correlation between these two variables. A one standard deviation increase in the average degree of a player's social ties outside of the experiment is associated with a 0.2 standard deviations increase in the number of times a subject plays according to the most popular player rule. Using robust standard errors, this is significant at the 10 % level.[34] On the other hand, the correlation between the degree of a player's social ties and the use of other link formation rules is weaker and insignificant. These results give further qualified support to the hypothesis that choosing the most popular player is a rule that some players apply across different network formation games.

4.2.3 Forward-looking strategies

In this section, we provide suggestive evidence on whether some of the deviations from myopic best response documented above are the product of forward-looking strategies. In particular, we are interested in the hypothesis that some subjects correctly expect others to make mistakes and, on the basis of these expectations, they find it optimal to deviate from MBR.[35] We will provide two pieces of evidence to shed light on this hypothesis, focusing on T1. It is important to note, however, that our experiment was ultimately not designed to study forward-looking strategies and thus this evidence should not be considered as conclusive.

The first piece of evidence looks at the correlation between deviations from MBR

[34]With clustered standard errors, the p-value is 0.16.

[35]We thank an anonymous referee for this suggestion.

and players' final reach in game. If deviations from MBR are optimal forward-looking responses to other players' mistakes, then we would expect players that deviate from MBR to have a higher reach at the end of the game. We explore this hypothesis in Table C10 in the Appendix. What we find is that players that deviate from MBR in both rounds have a final reach that is a significant 0.36 standard deviations lower than the reach of players who play MBR. Players that deviate from both MBR *and* the most-popular-player heuristic also have a final reach that is significantly lower than that of subjects who play MBR (by 0.34 standard deviations).

Second, we study whether the proportion of players who play MBR changes in the last turn of the game. Our rationale for this test is that even a forward-looking player should choose a myopic best response in the last turn of the game since the network will not evolve further after their decision is made. Thus, if some players are forward looking and this is what induces them to select an action different from myopic best response, we should see them switching to a myopic best response in the last turn of the game. As a result, the fraction of decisions consistent with myopic best response should increase in the last turn of game. Figure B12, however, shows that the fraction of players choosing MBR remains essentially unchanged in the last period of the game.[36]

Overall, these two pieces of evidence *suggest* that deviations from MBR may not be due to forward-looking strategies. However, as explained above, this conclusion is necessarily preliminary and we thus flag this as an area for future research.

4.3 Group identity

We now study the effect of group identity on network formation. We start by reporting the results of the two manipulations checks included in our design. First, we look at decisions in the initial allocation task where each player has to divide a sum of money between an individual who belongs to the same group and an individual who belongs to the other group. The most frequent decision, as shown in Figure B13 in the Appendix, is to allocate *more* money to the recipient affiliated to the same group. Overall, 54 % of individuals show such in-group bias, while 30 % choose equal allocations. The amount of in-group favouritism is large: the standardised discrimination coefficient is 0.92, which is in the upper range of the estimates reported in the literature (and about

[36] We are also unable to find significant evidence that players target their links on the basis of the round one decisions of the other players (see Table C11 in the Appendix).

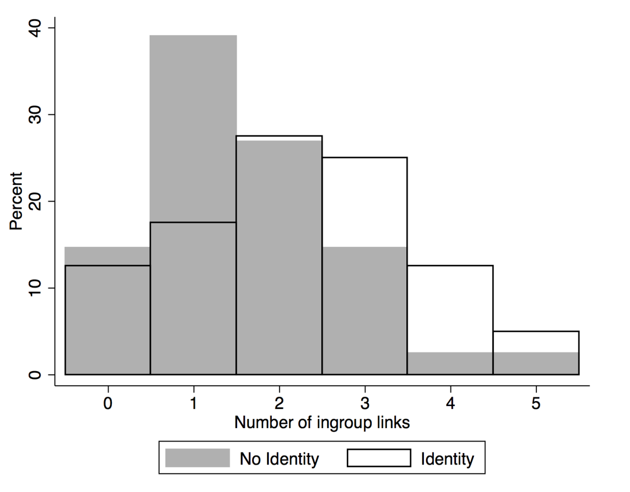
average for experiments that use minimal groups) (Lane, 2016). Second, we investigate whether players perceive that linking to individuals outside one's group violates a social norm. For this purpose, at the end of the experiment, we ask participants whether they think that in the link-formation game a player 'should' only connect to players in the same group. Fifty-seven % of participants answer yes to this question. Further, about 70 % of players expect the majority of the other 5 players in the session to also answer affirmatively. Second-order beliefs of this type are powerful motivators (Bursztyn et al., 2018) and a key ingredient in the formation of social norms (Bicchieri, 2005). We report these results in detail in Table C2 and Figure B13 in the Appendix.

Our main result on the identity treatments is the following one:

Result 4. *In-group links occur more frequently in sessions where group identity is disclosed than in sessions with no knowledge of group identity. Disclosure of group identity does not affect network efficiency.*

This result confirms the first part of prediction 2. First, in-group links increase. Figure 10 shows a histogram of the number of in-group links in the final network for sessions where group identity is disclosed and sessions where it is not disclosed. The distribution clearly shifts to the right. A Wilcoxon rank-sum test confirms this difference is significant at the 5 % level (Z= 2.23, p= .02).

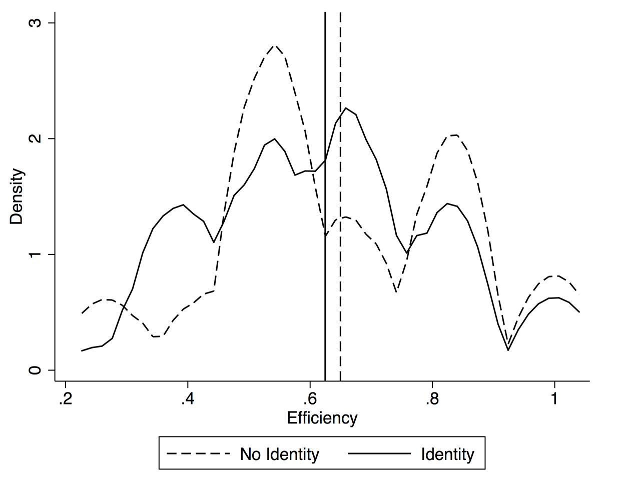
Figure 10: In-group links in identity and no-identity sessions



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

Second, we cannot detect a systematic effect of disclosing players' group identity on session level efficiency. Mean efficiency decreases from 65 to 58 % in T1 and is basically unchanged in T2. A Wilcoxon rank-sum test cannot reject the equality of the distributions (Z= -0.51, p= 0.61). This is documented graphically in Figure 11.

Figure 11: Efficiency in identity and no-identity sessions



To understand these aggregate results we study how disclosure of group identity affects individual decisions. We use linear probability models of the following form:

$$x_{dis} = \alpha + \beta_1 \text{Identity Session}_s + e_{dis}, \tag{6}$$

$$x_{dis} = \alpha + \beta_1 \text{Identity Session}_s + \beta_2 \text{z}_{is} + \beta_3 \left(\text{z}_{is} * \text{Identity Session}_s\right) + e_{dis}. \tag{7}$$

In these models x_{dis} is an indicator variable which takes a value of one if decision d by player i in session s has characteristic x. In some cases, we interact the identity treatment with an individual covariate z (for example, whether the individual discriminated in favour of the in-group in the initial allocation task). We perform the analysis with three definitions of x_{dis}: whether decision d is a link to a player in the same group, whether decision d is consistent with a 'max' rule, and whether decision d is consistent with a 'Rawlsian' rule. In model 7, we also study whether the effect of being in an identity session is stronger for certain types of players, for example, players who have allocated more money to the in-group partner in the allocation task. Standard errors are clustered at the session level. We report the results in tables C12, C13, C14, and C15 in the Appendix.

We find that there is a higher proportion of in-group links in both treatments. In T1, links to an in-group player are about 12 percentage points more likely once we disclose player group identity. This corresponds to a statistically significant 40 % increase in the probability of an in-group link. For T2 treatments the effect drops to 5 percentage points and is not significant.

We also find that the frequency of 'max' and 'Rawlsian' links is not significantly affected by disclosure of group identity. This is consistent with our discussion in section 2: when the strength of the in-group norm is moderate, players link to in-group partners only when these have the desired network position. If this is what is driving the results, we should expect to see an increase in the proportion of 'max' and 'Rawlsian' links that are directed towards an in-group partner. To study this, we restrict the sample to 'max' links in T1 and T2 and run model 6 with the in-group link dummy as dependent variable. Consistent with our hypothesis, we find that when we disclose group identity in T1 'max' links are about 14 percentage points more likely to be directed to an in-group player. This effect is significant at the 5 % level. For T2, we find a smaller effect of 5 percentage point, which is in the hypothesised direction but not significant.

We are unable to shed more light on these mechanisms through estimation of model 7: the effect of group identity disclosure on the likelihood of choosing an in-group link is not stronger for individuals who show in-group bias in the allocation task, who agree with the norm of homophily, or who expect more peers to agree with the norm of homophily.

4.4 Additional tests

In this section, we test whether the likelihood of choosing a link consistent with a given link-formation rule is correlated with the answers given in the comprehension test. We estimate the following regression models:

$$x_{dis} = \alpha + \beta_1 \text{comprehension}_{is} + e_{dis}, \tag{8}$$

$$
\begin{aligned}
x_{dis} =& \alpha + \gamma_1 \left(\text{comprehension}_{is} * \text{Identity Session=0}_s\right) \\
&+ \gamma_2 \left(\text{comprehension}_{is} * \text{Identity Session=1}_s\right) \\
&+ \gamma_3 \text{Identity Session}_s + e_{dis}.
\end{aligned} \tag{9}
$$

As before, x_{dis} is an indicator variable which takes a value of one if decision d by player i in session s has characteristic x. 'comprehension$_{is}$' is the z-score of the number of correct answers players give in the initial comprehension questions. In the first model, β_1 captures whether subjects with a higher comprehension score are more likely to choose links consistent with rule x. The second model captures the effect of higher comprehension in sessions where group identity is not disclosed (coefficient γ_1) and sessions where group identity is disclosed (coefficient γ_2). Tables C18 to C21 report estimation results for these two models. We start by pooling all sessions together and then look separately at T1 and T2 sessions. When we estimate model 8 over the pooled sample, we are unable to find evidence that people with a higher comprehension score use different rules. We are similarly unable to find evidence of a comprehension effect in all but two of the remaining thirty tests and, in both of these cases, the result is not robust to a standard correction for multiple comparisons (Benjamini et al., 2006).

Further, we test whether the answers that individuals give to specific questions correlate with the most-popular-player heuristic. We report this analysis in Figures B14 and B15 in the Appendix. We are generally unable to find significant correlations. We report both p-values and q-values for each test, to address the problem of multiple comparisons (Benjamini et al., 2006). q-values are always above 0.5, indicating that to reject any of the null hypotheses we need to tolerate a probability of making a false discovery of at least 50 %.

Overall, these results suggest that the strategies observed in the experiment are generally not associated with players' answers in the comprehension test. As we discussed earlier in the paper, participants receive further instructions after giving these answers and then have an opportunity to play a trial round of the game. Hence, our preferred interpretation for the results in this section is that subjects' imperfect comprehension was corrected before the beginning of the main experiment.

5 Discussion and Conclusion

In this paper we study network efficiency in a game of sequential, one-way-flow link formation (Bala and Goyal, 2000). The efficient solution is for players – a sample of male Indian farmers – to form an information circle, something they can easily do by following a myopic best response. We find that efficiency is achieved in only 10 % of the sessions and that subjects lose about 35 % of expected payoffs. While many individuals play according to an MBR, these large welfare losses arise because some subjects connect to the 'most popular' individual in the network. We also find that disclosing group identity leads to networks with significantly more homophily, but not less efficiency.

Researchers and policy makers are often puzzled by the fact that individuals hold biased beliefs in several domains, including the returns to new technologies, relative ability, and the behaviour of others (Malmendier and Tate, 2008; Bryan et al., 2014; Caria and Falco, 2016). A possible explanation for the persistence of these beliefs is that individuals fail to correctly incorporate available information (Hanna et al., 2014; Falk and Zimmermann, 2017; Epley and Gilovich, 2016). Our results offer a complementary explanation: social networks may not always be structured in a way that makes relevant information available. This evidence echoes some of the theoretical claims of Bala and Goyal (2001) and lends support to interventions that change the structure of social networks (Feigenberg et al., 2013; Vasilaky and Leonard, 2018; Fafchamps and Quinn, 2016; Cai and Szeidl, 2017). These interventions have often focused on creating *more* connections. Our results suggest that creating incentives for individuals to form *different* links – even if these links are to less popular or socially distant nodes – may be effective at improving information diffusion when nobody is willing or capable to serve as the information aggregator of the network.

A second key concern in the literature on beliefs is that homophily in information networks can lead to the creation of 'echo chambers' where extreme views prevail because individuals are not exposed to information that contradicts their prior beliefs (Sunstein, 2009; Gentzkow and Shapiro, 2011). We see that, in our experiment, subjects are not prepared to reduce their payoff in order to restrict links to the in-group. This raises the question of whether in field settings individuals underestimate the cost of echo chambers. In future work, we plan to explore whether interventions that alert individuals to these costs can improve information diffusion and welfare.

Authors’ affiliations

- A. Stefano Caria: Department of Economics, University of Bristol.
- Marcel Fafchamps: Freeman Spogli Institute for International Studies, Stanford University.

References

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

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

Andreoni, J. and J. Miller (2002). Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica 70*(2), 737–753.

Bala, V. and S. Goyal (2000). A Noncooperative Model of Network Formation. *Econometrica 68*(5), 1181–1230.

Bala, V. and S. Goyal (2001). Conformism and Diversity under Social Learning. *Economic theory 17*(1), 101–120.

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

Barabási, A.-L. and R. Albert (1999). Emergence of Scaling in Random Networks. *Science 286*(5439), 509–512.

Barr, A. and G. Genicot (2008). Risk Sharing, Commitment, and Information: An Experimental Analysis. *Journal of the European Economic Association 6*(6), 1151–1185.

Baumeister, R. F., E. Bratslavsky, M. Muraven, and D. M. Tice (1998). Ego Depletion: Is the Active Self a Limited Resource? *Journal of Personality and Social Psychology 74*(5), 1252.

Beaman, L., N. Keleher, and J. Magruder (2018). Do Job Networks Disadvantage Women? Evidence from a Recruitment Experiment in Malawi. *Journal of Labor Economics 36*(1), 121–157.

Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006). Adaptive Linear Step-Up Procedures that Control the False Discovery Rate. *Biometrika 93*(3), 491–507.

Berg, E., M. Ghatak, R. Manjula, D. Rajasekhar, and S. Roy (2017). Motivating Knowledge Agents: Can Incentive Pay Overcome Social Distance? *The Economic Journal 129*(617), 110–142.

Berninghaus, S., K.-M. Ehrhart, and M. Ott (2006). A Network Experiment in Continuous Time: The Influence of Link Costs. *Experimental Economics 9*(3), 237–251.

Bicchieri, C. (2005). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press.

Brewer, M. B. (1999). The Psychology of Prejudice: Ingroup Love and Outgroup Hate? *Journal of Social Issues 55*(3), 429–444.

Bryan, G., S. Chowdhury, and A. M. Mobarak (2014). Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh. *Econometrica 82*(5), 1671–1748.

Bursztyn, L., A. L. González, and D. Yanagizawa-Drott (2018). Misperceived Social Norms: Female Labor Force Participation in Saudi Arabia. *Working Paper*.

Cai, J. and A. Szeidl (2017). Interfirm Relationships and Business Performance. *The Quarterly Journal of Economics 133*(3), 1229–1282.

Callander, S. and C. R. Plott (2005). Principles of Network Development and Evolution: an Experimental Study. *Journal of Public Economics 89*(8), 1469–1495.

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

Caria, S. and M. Fafchamps (2017). Expectations, Network Centrality and Public Good Contributions: Experimental Evidence from India. *Working Paper*.

Caria, S. and P. Falco (2016). Do Employers Trust Workers too Little? An Experimental Study of Trust in the Labour Market. *Working Paper*.

Chandrasekhar, A. G., C. Kinnan, and H. Larreguy (2018). Social Networks as Contract Enforcement: Evidence from a Lab Experiment in the Field. *American Economic Journal: Applied Economics 10*(4), 43–78.

Chandrasekhar, A. G., H. Larreguy, and J. P. Xandri (2015). Testing Models of Social Learning on Networks: Evidence From a Lab Experiment in the Field. *NBER Working Paper No. 21468*.

Charness, G., F. Feri, M. A. Meléndez-Jiménez, and M. Sutter (2014). Experimental Games on Networks: Underpinnings of Behavior and Equilibrium Selection. *Econometrica 82*(5), 1615–1670.

Charness, G. and M. Rabin (2002). Understanding Social Preferences With Simple Tests. *Quarterly Journal of Economics 117*(3), 817–869.

Charness, G., L. Rigotti, and A. Rustichini (2007). Individual Behavior and Group Membership. *American Economic Review 97*(4), 1340–1352.

Comola, M. and M. Fafchamps (2014). Testing Unilateral and Bilateral Link Formation. *The Economic Journal 124*, 954–975.

Conte, A., D. D. Cagno, and E. Sciubba (2009). Strategies in Social Network Formation. *Birkbeck Working Papers in Economics and Finance* (0905).

Currarini, S., M. O. Jackson, and P. Pin (2009). An Economic Model of Friendship: Homophily, Minorities, and Segregation. *Econometrica 77*(4), 1003–1045.

de Quidt, J., J. Haushofer, and C. Roth (2018). Measuring and Bounding Experimenter Demand. *American Economic Review 108*(11), 3266–3302.

DellaVigna, S. (2009). Psychology and Economics: Evidence from the Field. *Journal of Economic Literature 47*(2), 315–372.

Dessi, R., E. Gallo, and S. Goyal (2016). Network Cognition. *Journal of Economic Behavior & Organization 123*, 78–96.

Eckel, C. C. and P. J. Grossman (2005). Managing Diversity by Creating Team Identity. *Journal of Economic Behavior & Organization 58*(3), 371–392.

Enke, B. and F. Zimmermann (2017). Correlation Neglect in Belief Formation. *The Review of Economic Studies 86*(1), 313–332.

Epley, N. and T. Gilovich (2016). The Mechanics of Motivated Reasoning. *The Journal of Economic Perspectives 30*(3), 133–140.

Fafchamps, M. and S. Quinn (2016). Networks and Manufacturing Firms in Africa: Results from a Randomized Field Experiment. *The World Bank Economic Review 32*(3), 656–675.

Falk, A. and M. Kosfeld (2012). It's all about Connections: Evidence on Network Formation. *Review of Network Economics 11*(3), 1–36.

Falk, A. and F. Zimmermann (2017). Information Processing and Commitment. *The Economic Journal 128*(613), 1983–2002.

Fehr, E. and K. M. Schmidt (1999). A Theory Of Fairness, Competition, And Cooperation. *Quarterly Journal of Economics 114*(3), 817–868.

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

Gabaix, X. and D. Laibson (2005). Bounded Rationality and Directed Cognition. *Working Paper*.

Galeotti, A. and S. Goyal (2010). The Law of the Few. *American Economic Review 100*, 1468–1492.

Gallo, E. and C. Yan (2015). Efficiency and Equilibrium in Network Games: An Experiment. *Working Paper*.

Gentzkow, M. and J. M. Shapiro (2011). Ideological Segregation Online and Offline. *The Quarterly Journal of Economics 126*(4), 1799–1839.

Goeree, J. K., A. Riedl, and A. Ule (2009). In Search of Stars: Network Formation among Heterogeneous Agents. *Games and Economic Behavior 67*(2), 445–466.

Goette, L., D. Huffman, and S. Meier (2012). The Impact of Social Ties on Group Interactions: Evidence from Minimal Groups and Randomly Assigned Real Groups. *American Economic Journal: Microeconomics 4*(1), 101–115.

Golub, B. and M. O. Jackson (2012). How Homophily Affects the Speed of Learning and Best-Response Dynamics. *Quarterly Journal of Economics 127*(3), 1287–1338.

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

Goyal, S. (2011). Learning in Networks. *Handbook of Social Economics*, 679–727.

Greif, A. (1993). Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition. *The American Economic Review*, 525–548.

Hanna, R., S. Mullainathan, and J. Schwartzstein (2014). Learning through Noticing: Theory and Evidence from a Field Experiment. *The Quarterly Journal of Economics 129*(3), 1311–1353.

Jackson, M. O. (2010). *Social and Economic Networks*. Princeton university press.

Jackson, M. O. and A. Watts (2002). The Evolution of Social and Economic Networks. *Journal of Economic Theory 106*(2), 265–295.

Jackson, M. O. and A. Wolinsky (1996). A Strategic Model of Social and Economic Networks. *Journal of Economic Theory 71*(1), 44–74.

Kirchsteiger, G., M. Mantovani, A. Mauleon, and V. Vannetelbosch (2016). Limited Farsightedness in Network Formation. *Journal of Economic Behavior & Organization 128*, 97–120.

Lane, T. (2016). Discrimination in the Laboratory: A Meta-Analysis of Economics Experiments. *European Economic Review 90*, 375–402.

Majumder, J. and P. Shah (2017). Mapping the Role of Women in Indian Agriculture. *Annals of Anthropological Practice 41*(2), 46–54.

Malmendier, U. and G. Tate (2008). Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction. *Journal of Financial Economics 89*(1), 20–43.

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

Pons, P. and M. Latapy (2006). Computing Communities in Large Networks Using Random Walks. *Journal of Graph Algorithms and Applications, 10*(2), 191–218.

Rogers, E. M. (2003). *Diffusion of Innovations*. Free Press.

Rosenkranz, S. and U. Weitzel (2012). Network structure and strategic investments: An experimental analysis. *Games and Economic Behavior 75*(2), 898–920.

Sunstein, C. R. (2009). *Republic.com 2.0*. Princeton University Press.

Tajfel, H. (1981). *Human Groups and Social Categories: Studies in Social Psychology*. CUP Archive.

Vasilaky, K. N. and K. L. Leonard (2018). As Good as the Networks they Keep? Improving Outcomes through Weak Ties in Rural Uganda. *Economic Development and Cultural Change 66*(4), 755–792.

Viceisza, A. (2012). Treating the Field as a Lab: A Basic Guide to Conducting Economics Experiments for Policymaking.

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

Yamagishi, T. and T. Kiyonari (2000). The Group as the Container of Generalized Reciprocity. *Social Psychology Quarterly 63*(2), pp. 116–132.

Zizzo, D. (2010). Experimenter Demand Effects in Economic Experiments. *Experimental Economics 13*(1), 75–98.

For Online Publication

A Formal derivation of the rules

A.1 Notation

We define some basic notation following Goyal (2007) and Bala and Goyal (2000). Let N=(1,2,..,n) be the set of players. In T1, each player i chooses a (pure) strategy $s_i = (s_{i1}, s_{i2}, ..., s_{ii-1}, s_{ii+1}, ..., s_{in})$[37]. $s_{ij} \in \{0, 1\}$ and $s_{ij} = 1$ if there is a link from i to j. In T2, on the other hand, every player chooses a strategy $s_i = (s_{1i}, s_{2i}, ..., s_{i-1i}, s_{i+1i}, ..., s_{ni})$. As before, $s_{ji} \in \{0, 1\}$ and $s_{ji} = 1$ if there is a link from j to i. Let $\mathcal{S}_i$ be the set of possible values of s_i. $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 \times ... \times \mathcal{S}_n$ is the set of all possible combinations of player strategies. A vector of player strategies $s = (s_1, s_2, ..., s_n)$, drawn from $\mathcal{S}$, can be represented as a directed network g. $g_{ij} \in \{0, 1\}$ captures whether there is a link from i to j in this network ($g_{ij} = 1$ indicates that there is a link, and that player i observes player j's information thanks to this link). $g \oplus ij$ is the network obtained from adding the ij link to network g.

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

We need to introduce two crucial concepts for our analysis. First, let $N_{\underrightarrow{i}}(g) = \{k \in N | i \xrightarrow{g} k\}$ and $\nu_{\underrightarrow{i}}(g) = |N_{\underrightarrow{i}}(g)|$. $N_{\underrightarrow{i}}(g)$ is the set of players whom player i reaches in network g and $\nu_{\underrightarrow{i}}(g)$ represents the number of players whom player i reaches in network g. We call $\nu_{\underrightarrow{i}}(g)$ the **reach** of player i in network g. Sometimes we want to exclude from the count the path from player i to player j. Let $N_{\underrightarrow{i \sim j}}(g) = \{k \in N \setminus j | i \xrightarrow{g} k\}$ and $\nu_{\underrightarrow{i \sim j}}(g) = |N_{\underrightarrow{i \sim j}}(g)|$. These are, respectively, the set of players and the number of players whom player i reaches in network g, *excluding player* j. Intuitively, $\nu_{\underrightarrow{i \sim j}}(g)$ captures the reach of player i among all players other than j.

[37]We rule out links from player i to player i.

Second, let $\underleftarrow{N_i}(g) = \{k \in N | i \xleftarrow{g} k\}$ and $\underleftarrow{\nu_i}(g) = |\underleftarrow{N_i}(g)|$. $\underleftarrow{N_i}(g)$ and $\underleftarrow{\nu_i}(g)$ represent, respectively, the set of players and the number of players who reach player i in network g. We call $\underleftarrow{\nu_i}(g)$ the **in-reach** of player i in network g. Again, we sometimes need to exclude the path from player j to player i. Let $\underleftarrow{N_{i \sim j}}(g) = \{k \in N \setminus j | k \xrightarrow{g} i\}$ and $\underleftarrow{\nu_{i \sim j}}(g) = |\underleftarrow{N_{i \sim j}}(g)|$. $\underleftarrow{\nu_{i \sim j}}(g)$ is the number of players who reach player i in network g, *excluding player* j.

Reach and in-reach ($\underrightarrow{\nu_i}$ and $\underleftarrow{\nu_i}$) should not be confused with the notions of out-degree and in-degree, which represent the number of direct links of a player in the network.[38]

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:

$$\pi_i(g) = \Big(1 + \underrightarrow{\nu_i}(g)\Big) \Big/ n \; . \tag{A.1}$$

A.2 Rule 1

We assume that in T1 player i chooses which player j to link to in order to myopically maximise his expected payoff:

$$\max_j \quad \pi_i(g \oplus ij). \tag{A.2}$$

Note that whenever i has the turn, $\underrightarrow{N_i}(g) = \{\varnothing\}$ and $\underrightarrow{\nu_i}(g) = 0$. This is because (i) in the first round of the game, player i has not yet established a link, and (ii) in the second round of the game, player i's first-round link will be removed when it is his turn to choose his new link.

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

Proposition 1. *Player i maximises $\pi_i(g \oplus ij)$ by choosing the partner j with the maximum value of $\nu_{\underrightarrow{j \sim i}}$ in network g.*

Proof. Rewrite $\pi_i(g \oplus ij)$ as: $\Big(1 + \nu_{\underrightarrow{i}}(g \oplus ij)\Big) \Big/ n$. Notice that, as $\nu_{\underrightarrow{i}}(g) = 0$, $\nu_{\underrightarrow{i}}(g \oplus ij) = 1 + \nu_{\underrightarrow{j \sim i}}(g)$. Thus $\pi_i(g \oplus ij) = \Bigg(2 + \nu_{\underrightarrow{j \sim i}}(g)\Bigg) \Big/ n$, which is monotonically increasing in $\nu_{\underrightarrow{j \sim i}}(g)$. $\square$

The proof above captures a simple intuition. Player i wants to maximise his reach in the network. To do that, he has to link to the player who enables him to access the largest number of subjects. This is the player with the highest reach among all subjects that are not i: $\nu_{\underrightarrow{j \sim i}}$ (as connections to i are redundant from i's perspective). This proposition motivates rule 1.

A.3 Rule 2

We now consider a case where player i cares both about his own expected payoff and about the sum of the expected payoffs of the other players. This case is particularly relevant in T2 and motivates rule 2. Formally, in this case, player i's utility is given by:

$$u_i = \pi_i + \gamma \sum_{j \in N \setminus i} \pi_j. \tag{A.3}$$

The utility that player i gains by creating a ji link in T2 can be then expressed as:

$$\begin{aligned} u_i(g \oplus ji) - u_i(g) &= \pi_i(g \oplus ji) - \pi_i(g) + \gamma \sum_{k \in N \setminus i} \pi_k(g \oplus ji) - \pi_k(g) \\ &= \gamma \sum_{k \in N \setminus i} \pi_k(g \oplus ji) - \pi_k(g) \\ &= \gamma f(g \oplus ji). \end{aligned} \tag{A.4}$$

where $f(g \oplus ji) = \sum_{k \in N \setminus i} \pi_k(g \oplus ji) - \pi_k(g)$. The first two terms on the right hand side drop because in T2 player i's link does not affect player i's reach or his expected payoff. Self-regarding considerations thus become irrelevant (at least for myopic players). In T2, player i will choose his link on the basis of $f(g \oplus ji)$, which captures the increase in the expected payoff of the other players. In what follows, we study how $f(g \oplus ji)$ is related to player j's position in the network. We need to consider two separate cases, depending on whether player i reaches player j or not (and thus whether the information of player i is entirely new to player j or not).

Case 1: $j \notin \underrightarrow{N_i}(g)$ (player i does not observe the information of player j). In this case, we can derive a simple analytic expression for $f(g \oplus ji)$:

$$f(g \oplus ji) = \left(1 + \underrightarrow{\nu_i}(g)\right)\left(1 + \underleftarrow{\nu_j}\right). \tag{A.5}$$

Equation (A.5) shows that $f(g \oplus ji)$ increases monotonically with player j's in-reach. To derive equation (A.5) we first need to show the following property of networks in T2: if player i does not observe the information of player j, then it follows that player j (or any player reached by player j) does not observe the information of player i. If this property did not hold, player i would need to consider whether some of his information is redundant for player j (something that he has to do in case 2 below). Formally, we describe this property with the following lemma:

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

Proof of lemma. We refer to a ij link as an 'ingoing' link for player i (player i obtains new information from player j thanks to this link) and to a ji link as an 'outgoing' link (player i shares his information with player j through this link). In T2, players can have multiple ingoing links and at most one outgoing link. When it is player i's turn to play, player i has no outgoing links. He may have one or more ingoing links, in which case $\underrightarrow{N_i}(g)$ is nonempty. Every individual k in $\underrightarrow{N_i}(g)$ has exactly one outgoing link. This outgoing link is either a link that shares information with i, or a link that shares information with a third player z in $\underrightarrow{N_i}(g)$ (so that there is a path that lets i observe k's information, which has to be the case since $k \in \underrightarrow{N_i}(g)$). Thus, no player in the set

$i \cup \underset{\longrightarrow}{N_i}(g)$ has an outgoing link with a player who is not in the set $i \cup \underset{\longrightarrow}{N_i}(g)$. This means that player j, who is not in $\underset{\longrightarrow}{N_i}(g)$ by assumption, cannot reach any player in $i \cup \underset{\longrightarrow}{N_i}(g)$. Further, this implies that no player in $\underset{\longleftarrow}{N_j}(g)$ can reach any player in $i \cup \underset{\longrightarrow}{N_i}(g)$.[39] $\square$

Having established Lemma 1, we can now show how equation (A.5) is obtained. Note that if player i creates link ji, he will increase the expect payoff of player j and of all players who reach player j. The number of players who benefit from this link is thus $1 + \underset{\longleftarrow}{\nu_j}$. By how much does each of these players benefit? Lemma 1 tells us that none of the players in $i \cup \underset{\longrightarrow}{N_i}(g)$ was previously reached by player j or by the players in $\underset{\longleftarrow}{N_j}(g)$. As a result, each of these players experiences an increase in reach that is equal to all of the information that player i observes: $1 + \underset{\longrightarrow}{\nu_i}(g)$. Thus, the total increase in reach is given by:

$$
\begin{aligned}
f(g \oplus ji) &= \left(1 + \underset{\longrightarrow}{\nu_i}(g)\right)\left(1 + \underset{\longleftarrow}{\nu_j}\right) \\
&= \left(1 + \underset{\longrightarrow}{\nu_i}(g)\right)\left(1 + \underset{\longleftarrow}{\nu_{j \sim i}}\right).
\end{aligned}
$$

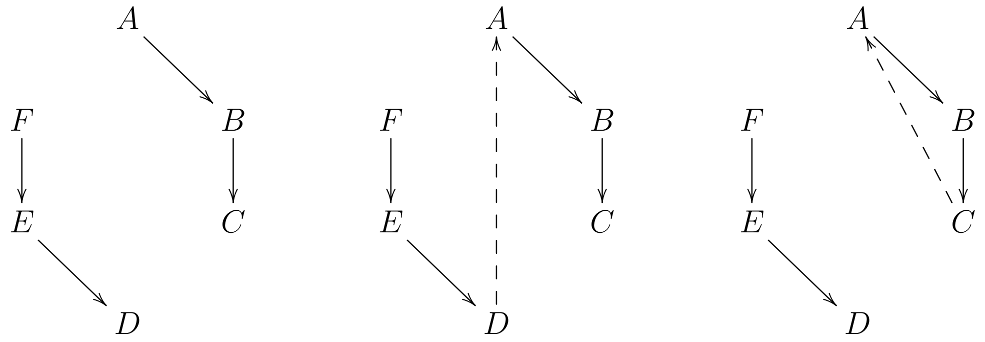
where the second line follows since $j \notin \underset{\longrightarrow}{N_i}(g)$.

Case 2: $j \in \underset{\longrightarrow}{N_i}(g)$ (player i observes the information of player j). Now a link ji creates a circle and some of the information shared with player j is redundant. The effect on $f(g \oplus ji)$ of an outgoing link to j is thus smaller compared to that of an outgoing link to a player who is not in $\underset{\rightarrow}{N_i}(g)$ and has the same in-reach as j. Figure A.1 illustrates this point.[40] As a result, player i cannot simply compare the in-reach of his potential

[39] Suppose a player z in $\underset{\longleftarrow}{N_j}(g)$ reached a player x in $\underset{\longrightarrow}{N_i}(g)$. x has only one outgoing link. So for x to be part of $\underset{\longrightarrow}{N_i}(g)$, player z would need to have an outgoing link to another player whom player i reaches. However, through this link i would observe the information of player z. Further, through z, they will also observe the information of player j, which violates the assumption that $j \notin \underset{\longrightarrow}{N_i}(g)$

[40] In the example, player A has the turn. Both player C and player D have an in-reach of two. However, a link to player D is more effective at increasing aggregate payoffs compared to a link to C. If player A links with D, there are three players (D, E and F) who experience an increase in reach of 3. On the other hand, If player A links to player C, the reach of player C increases by 2, the reach of player B increases

Figure A.1: Effect of a ji link



(a) *(b)* *(c)*

partners in order to figure out which link maximises $f(g \oplus ji)$. A good heuristic in this case is to pick the player with the maximum value of $\underleftarrow{\nu_{j\sim i}}(g)$ (as opposed to $\underleftarrow{\nu_{j}}(g)$), as $\underleftarrow{\nu_{j\sim i}}(g)$ does not count the redundant path to i and thus it is less likely that linking to the person who has maximum $\underleftarrow{\nu_{j\sim i}}(g)$ would create a cycle.[41]

The analysis of case 1 and case 2 shows why linking to the player with the maximum value of $\underleftarrow{\nu_{j\sim i}}(g)$ is an effective strategy in T2:

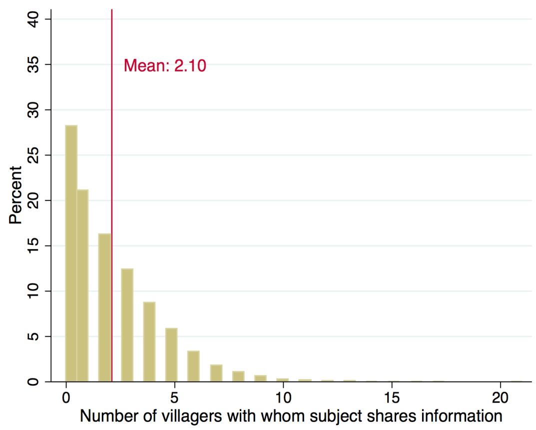
1. When $\underrightarrow{N_i}(g) = \emptyset$, $f(g \oplus ji)$ monotonically increases in $\underleftarrow{\nu_{j\sim i}}(g)$.
2. When $\underrightarrow{N_i}(g) \neq \emptyset$ and no player with the maximum value of $\underleftarrow{\nu_{j\sim i}}(g)$ is part of set $\underrightarrow{N_i}(g)$, $f(g \oplus ji)$ monotonically increases in $\underleftarrow{\nu_{j\sim i}}(g)$.
3. When $\underrightarrow{N_i}(g) \neq \emptyset$ and at least some of the players with the maximum value of $\underleftarrow{\nu_{j\sim i}}(g)$ are part of set $\underrightarrow{N_i}(g)$, linking to a player with the highest value of $\underleftarrow{\nu_{j\sim i}}(g)$ is likely to lead to the largest possible increase in $f(g \oplus ji)$ (although this is not always guaranteed).

by 1 and the reach of player A is unaffected. The difference lies in the fact that a link to player C creates a cycle among a subset of players where some information is redundant.

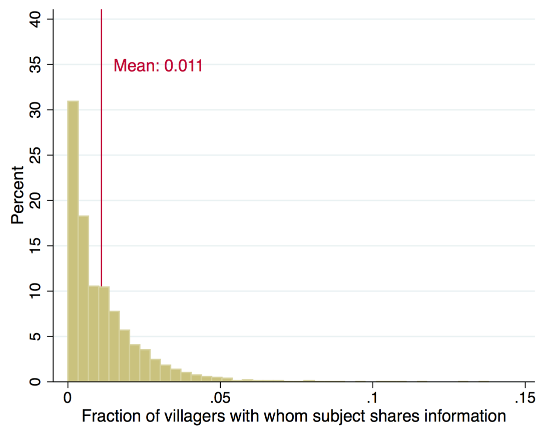
[41] In Figure A.1, for example, $\underleftarrow{\nu_{D\sim A}}(g) = 2$ and $\underleftarrow{\nu_{C\sim A}}(g) = 1$. Player D thus has the maximum value of $\underleftarrow{\nu_{j\sim i}}(g)$ in this network.

B Figures

Figure B1: Degree distribution in farmers' information networks



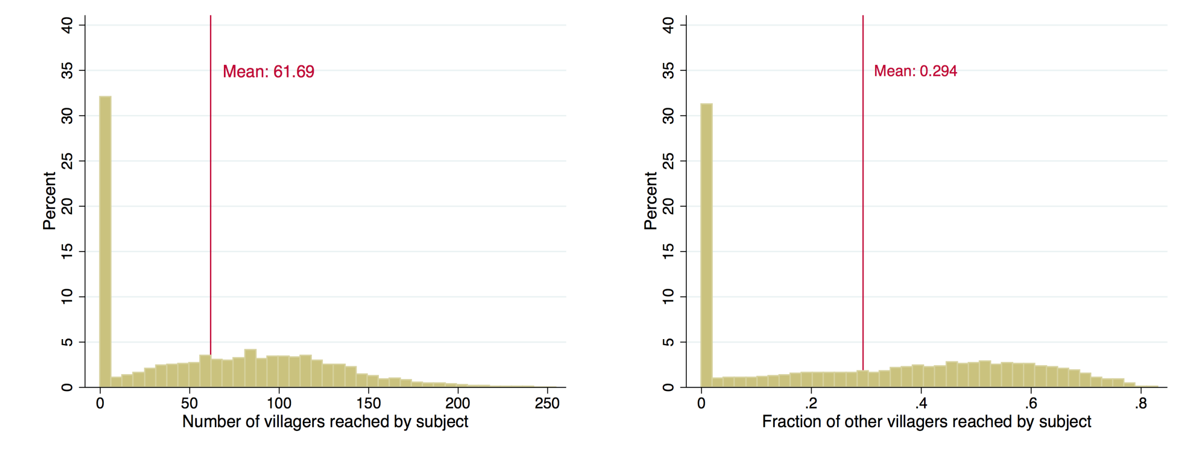
(a) Number of people



(b) Fraction of people

Authors' calculations using data from Banerjee et al. (2013).

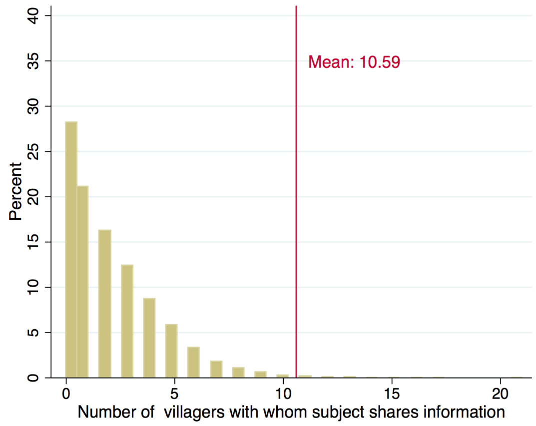
Figure B2: Reach distribution in farmers' information networks



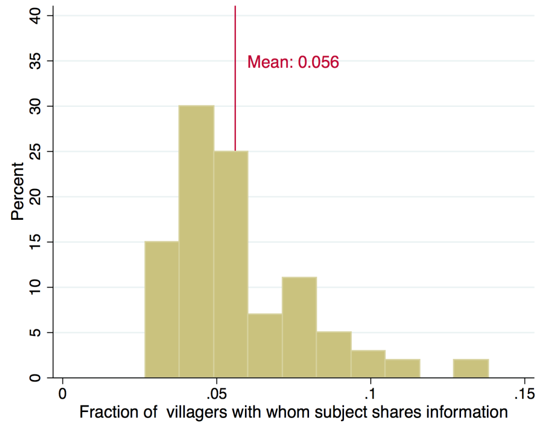
(a) Number of people *(b) Fraction of people*

Authors' calculations using data from Banerjee et al. (2013). Reach is defined as the number of individuals that a subject reaches in the network with a path of no more than five links.

Figure B3: Degree distribution among most central individuals



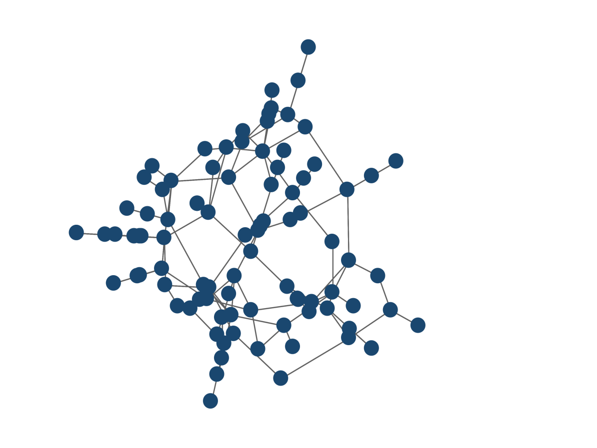
(a) Number of people



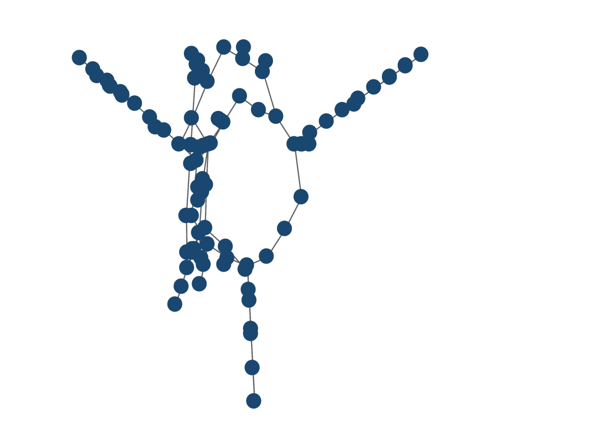
(b) Fraction of people

Authors' calculations using data from Banerjee et al. (2013).

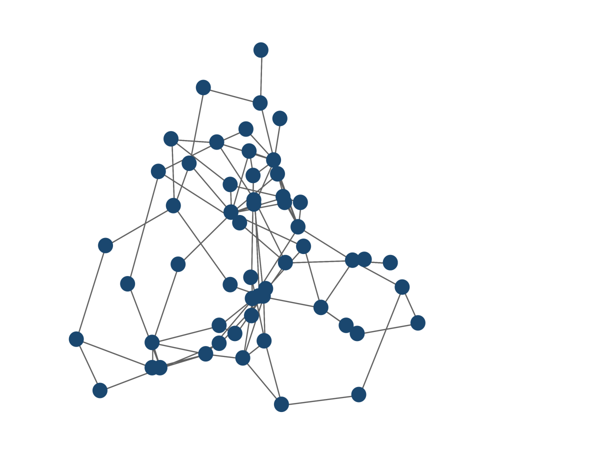
Figure B4: Network maps for selected villages



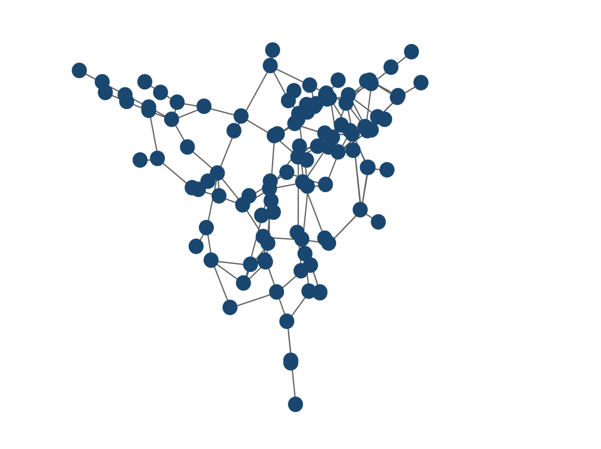
(a)



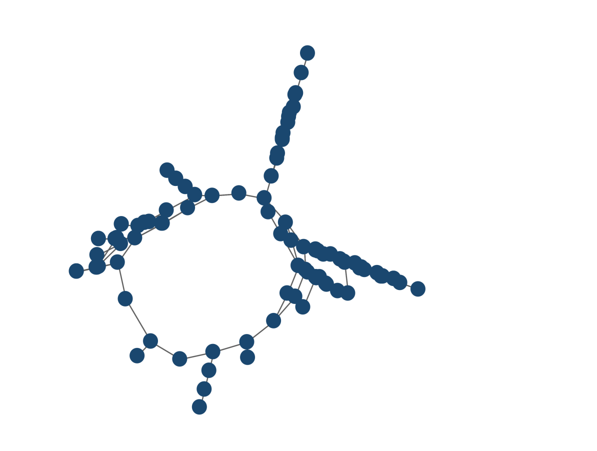
(b)



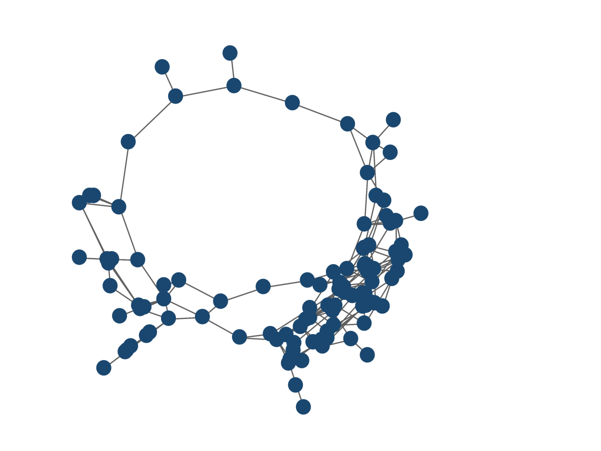
(c)



(d)



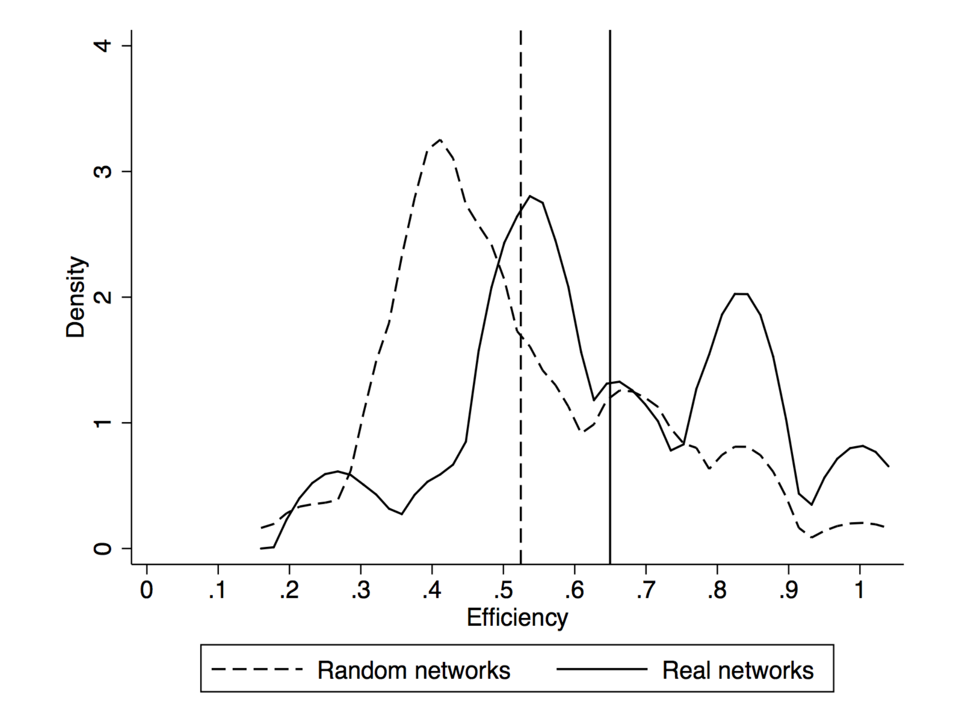
(e)



(f)

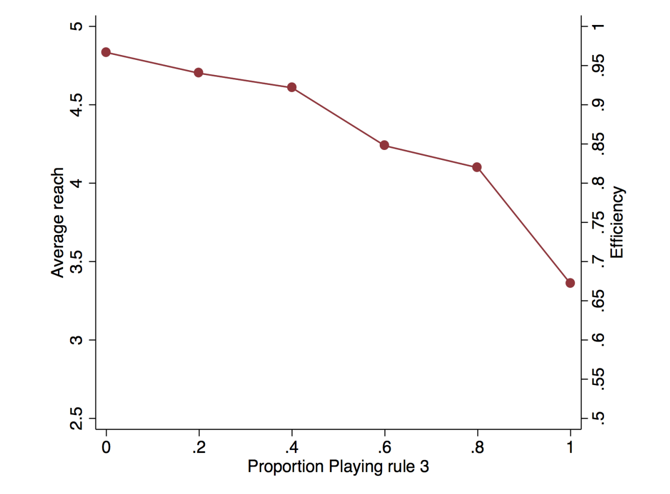
Plots of the largest network component in selected villages using data from Banerjee et al. (2013).

Figure B5: Efficiency in real networks and in simulated random networks



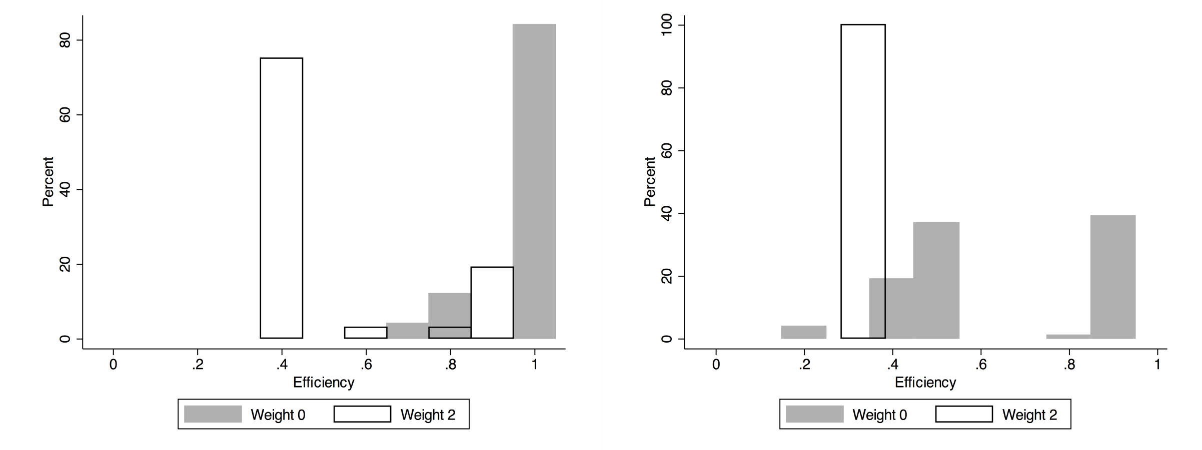
Note. 'Real networks' include all sessions in T1no and T2no. 'Random networks' are networks simulated assuming all players choose links at random.

Figure B6: Link formation process with mixed rules



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

Figure B7: Simulated effect of group identity concerns on network structure



(a) Rule 2 *(b) Rule 3*

Weight 0 simulations show efficiency when all players play rule 2 (panel a) or rule 3 (panel b). Weight 2 simulations show efficiency when players value a link to an in-group player as much as 2 units of in-reach (panel a) or two units of reach (panel b). We report results summarising 100 simulation for each of the 4 rules.

Figure B8: Time series of effciency in T1no and T2no, turns 6-12

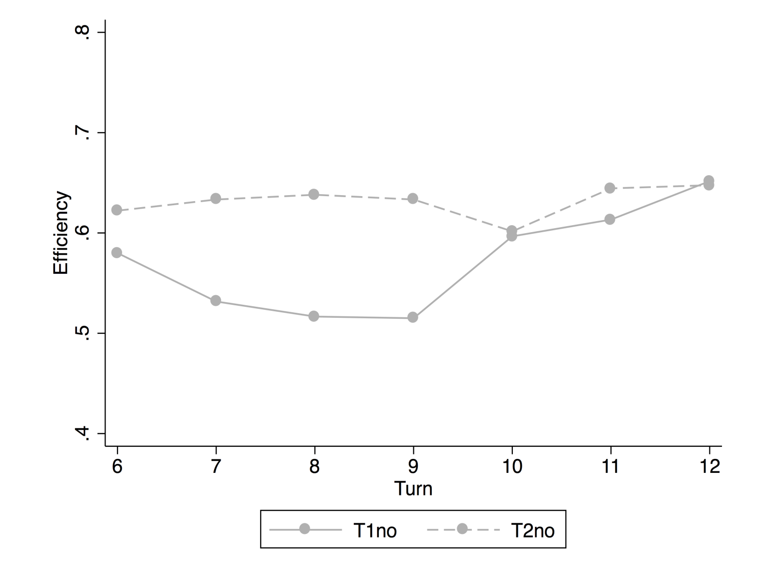


Figure B9: Overlap in choice sets

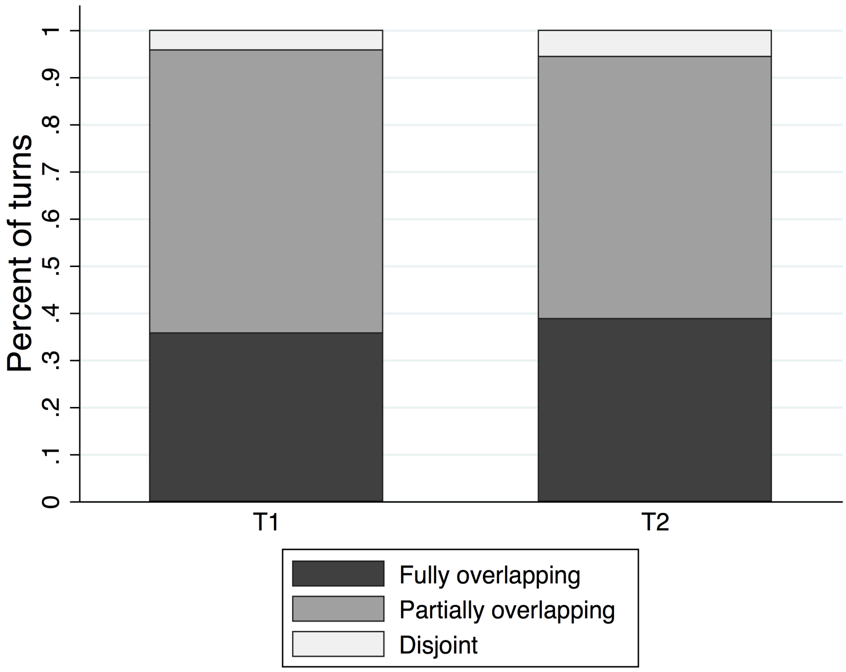
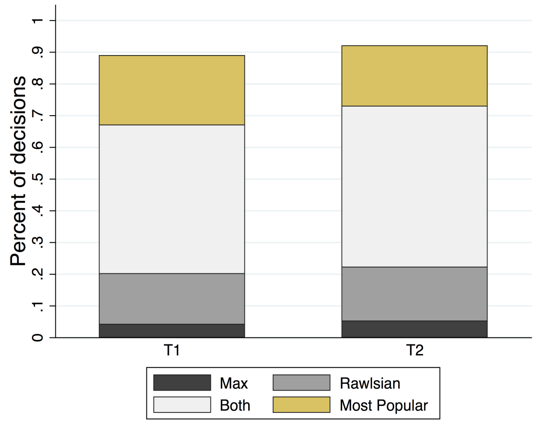
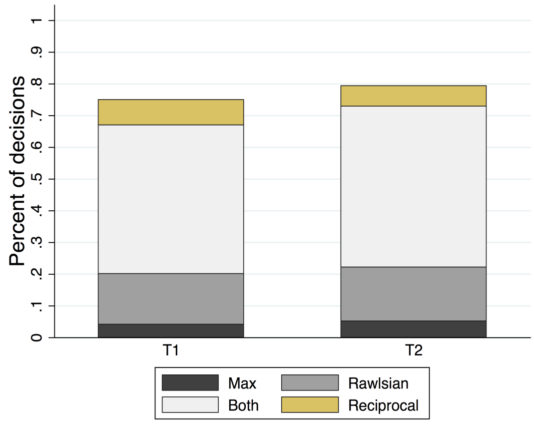


Figure B10: What explains links that are not consistent with link-formation rules 1-4?



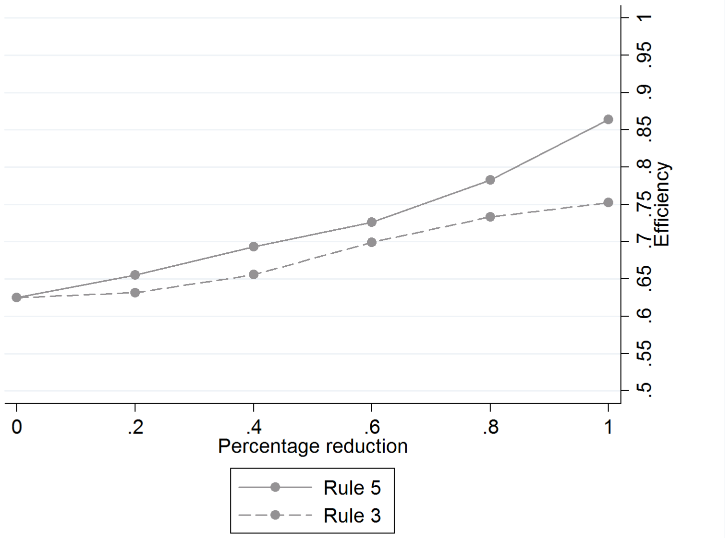
(a) Links to the 'most popular' player



(b) 'Reciprocal' links

The category 'most popular' shows the relative frequency of decisions consistent with rule 5 and not consistent with the 'max' rule nor with the 'Rawlsian' rule. The category 'reciprocal' shows relative frequency of decisions consistent with rule 6 and not consistent with the 'max' rule nor with the 'Rawlsian' rule. Only data for sessions with no knowledge of group identity is shown.

Figure B11: Efficiency simulation



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

Figure B12: Last turn vs previous turns

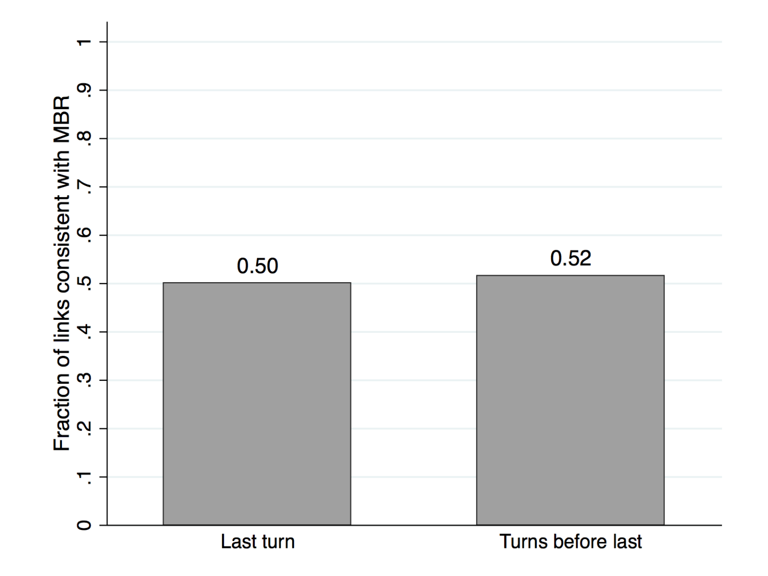
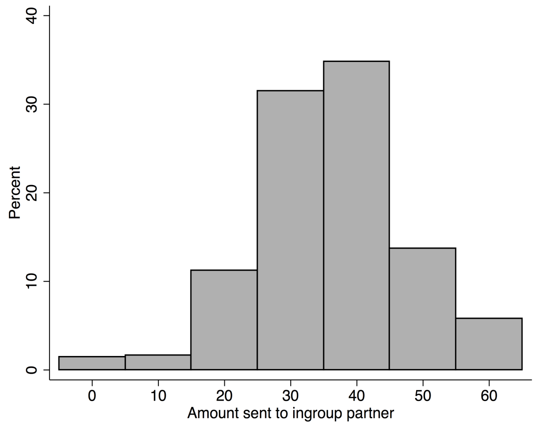
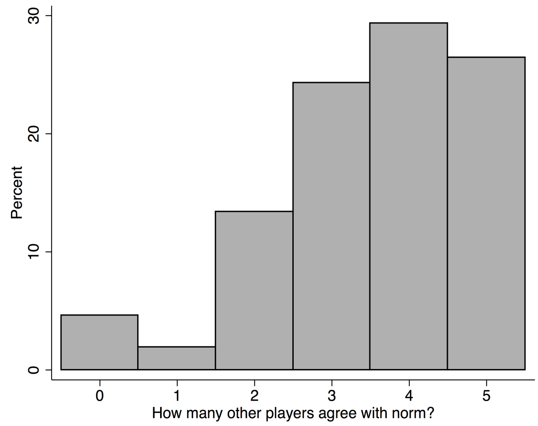


Figure B13: In-group bias

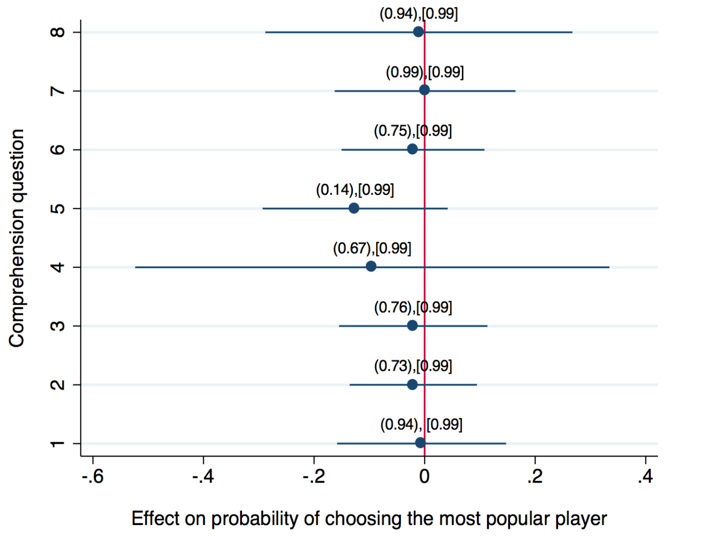


(a) Allocation to in-group partner



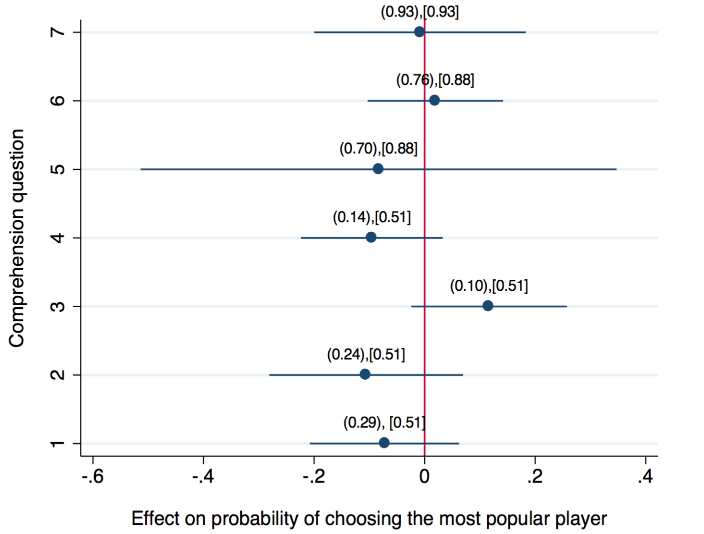
(b) Norm expectation

Figure B14: Comprehension questions and the most-popular-player heuristic in T1



Note. This figure reports the results of a regression of a dummy for whether the player chose a link consistent with the most-popular-player heuristic at least once on a dummy for whether the player answered a given comprehension question correctly. We plot coefficient estimates and 95% confidence intervals. Standard errors are clustered at the level of the session. Further, above each coefficient, we report a p-value (in parentheses), and a false-discovery-rate q-value (Benjamini et al., 2006) (in brackets).

Figure B15: Comprehension questions and the most-popular-player heuristic in T2



Note. This figure reports the results of a regression of a dummy for whether the player chose a link consistent with the most-popular-player heuristic at least once on a dummy for whether the player answered a given comprehension question correctly. We plot coefficient estimates and 95% confidence intervals. Standard errors are clustered at the level of the session. Further, above each coefficient, we report a p-value (in parentheses), and a false-discovery-rate q-value (Benjamini et al., 2006) (in brackets).

C Tables

Table C1: Comprehension questions

Treatment: T1

Question	Topic	Correctly answered (percent)
1	Direction of links	75.3
2		75.7
3		77.0
4	Indirect connections	98.3
5		85.8
6	Identify myopic best response	80.8
7		86.6
8		93.7

Treatment: T2

Question	Topic	Correctly answered (percent)
1	Direction of links	84.8
2		83.2
3		82.8
4	Identify link that benefits most players	61.1
5		98.4
6	Identify player with smallest payoff	47.1
7		83.6

Table C2: Summary statistics on allocations, norms and expectations

Variable	Obs.	Mean	Std. Dev.	Min	Max
Amount allocated to in-group partner	483	18.03	5.83	0	30
Bias in allocation task	483	0.54	0.5	0	1
Agrees with in-group norm	482	1.43	0.5	1	2
No. other players expected to agree with the norm	477	3.51	1.31	0	5

'Amount allocated to in-group partner' is the number of Rupees allocated to the in-group partner in the allocation task. 'In-group bias in allocation' is a dummy equal to 1 if the player has allocated more than half of the endowment to the in-group partner in the allocation task. 'Agrees with in-group norm' 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.

Table C3: The agricultural knowledge contest

	All	Group 1	Group 2
Probability of winning the point	0.387	0.396	0.379

Table C4: Degree centrality and reach

	Number of steps			
	2	3	4	5
Probability degree central node does not have highest reach	.54	.7	.76	.78

Authors calculations of the probability that the degree central node in the village does not have the highest reach, using data from the network survey of Banerjee et al. (2013). Reach is defined differently in each column. In the first column, we define reach as the number of people in the village that each subject reaches with at most two links. We then calculate reach using paths that are at most three, four, and five links. The maximum path length is indicated in the column heading.

Table C5: Balance test

	Age	Education	Upper Caste	Land Owned	Land Cultivated	Network Size
	(1)	(2)	(3)	(4)	(5)	(6)
T1id	-2.378	.091	-.040	.077	.141	.111
	(2.530)	(.081)	(.097)	(.733)	(.653)	(1.095)
T2no	-3.664	.033	-.007	-.069	-.008	1.581
	(2.569)	(.073)	(.091)	(.608)	(.546)	(1.723)
T2id	-1.993	.002	-.146	-.005	.067	1.150
	(2.167)	(.072)	(.101)	(.694)	(.631)	(1.376)
Mean in T1no	45.387	.248	.764	4.076	3.556	10.187
F-test (p-value)	.54	0.68	.43	.99	.99	.69
Obs.	478	465	432	474	469	427

OLS regressions. The dependent variable is indicated in the row's name. Upper caste is a variable that takes value of 1 if the respondent is not from a scheduled 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: *** $\leftrightarrow$ 99%, ** $\leftrightarrow$ 95%, * $\leftrightarrow$ 90%. Standard errors clustered at the session level reported in parentheses. The second to last row reports the p-value of an F-test of the joint null hypothesis that all coefficients are equal to zero.

Table C6: Continuous measures of reach and in-reach

	Link_{ijr}			
	(1)	(2)	(3)	(4)
Reach	.047	.047	-.049	-.048
	(.001)***	(.007)***	(.000)***	(.000)***
In-reach	-.029	-.026	.058	.057
	(.012)**	(.096)*	(.000)***	(.000)***
Const.	.178	.176	.187	.186
	(.000)***	(.000)***	(.000)***	(.010)**
Obs.	1200	910	1260	940
Cluster N	20	20	21	21
Sample	T1no	T1no	T2no	T2no
Controls		✔		✔

Linear probability model. The dependent variable is a dummy which takes a value of one if player i connects to player j in round r of the game. Round fixed effects are included. 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 comprehension questions and dummies for having completed secondary education, for being Hindu, and for belonging to an upper caste. Confidence: *** $\leftrightarrow$ 99%, ** $\leftrightarrow$ 95%, * $\leftrightarrow$ 90%. Standard errors corrected for clustering at session level. P-values obtained with the wild bootstrap-t procedure reported in parentheses.

Table C7: Links to the most popular player and reciprocal links

	Link_{ijr}	
	(1)	(2)
max reach	.126	
	(.044)**	
min in-reach	.015	
	(.029)**	
max in-reach		.803
		(.016)**
min reach		.095
		(.018)**
Most popular	-.019	.006
	(.503)	(.859)
Reciprocal	-.018	-.001
	(.705)	(.978)
Const.	.119	.028
	(.024)**	(.546)
Obs.	910	940
Sample	T1no	T2no
Cluster N	20	21
Controls	✔	✔

Linear probability model. The dependent variable is a dummy which takes a value of one if player i connects to player j in round r of the game. Each regression controls for the round, as well as for age, land owned, land cultivated, number of contacts in real information networks, number of mistakes in the initial comprehension questions and dummies for having completed secondary education, for being Hindu, and for belonging to an upper caste. Confidence: *** $\leftrightarrow$ 99%, ** $\leftrightarrow$ 95%, * $\leftrightarrow$ 90%. Standard errors corrected for clustering at session level. P-values obtained with the wild bootstrap-t procedure reported in parentheses.

Table C8: Why do individuals target the most popular player?

	Link$_{ijr}$								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Most popular node	-.0005	.040	.019	.014	-.034	-.009	-.139	-.105	-.121
	(.962)	(.494)	(.228)	(.744)	(.305)	(.749)	(.084)*	(.199)	(.029)**
Most popular * salient efficient node	-.028	-.100	-.065						
	(.712)	(.108)	(.197)						
Salient efficient node	.009	.045	.026						
	(.774)	(.165)	(.263)						
Most popular * difference				-.021	-.006	-.013			
				(.235)	(.712)	(.258)			
Difference most/least efficient node				.016	.006	.011			
				(.201)	(.625)	(.210)			
Most popular * round							.077	.040	.058
							(.093)*	(.356)	(.077)*
Round	-.005	-.006	-.006	-.008	-.006	-.007	-.039	-.023	-.031
	(.101)	(.122)	(.025)**	(.064)*	(.181)	(.029)**	(.089)*	(.333)	(.053)*
Const.	.209	.191	.201	.189	.218	.203	.268	.256	.262
	(.000)***	(.000)***	(.000)***	(.000)***	(.000)***	(.000)***	(.000)***	(.000)***	(.000)***
Obs.	1200	1260	2460	1200	1260	2460	1200	1260	2460
Cluster N	20	21	41	20	21	41	20	21	41
Sample	T1no	T2no	Both	T1no	T2no	Both	T1no	T2no	Both

Linear probability model. The dependent variable is a dummy which takes a value of one if player i connects to player j in round r of the game. Confidence: *** $\leftrightarrow$ 99%, ** $\leftrightarrow$ 95%, * $\leftrightarrow$ 90%. Standard errors corrected for clustering at the session level. P-values obtained with the wild bootstrap-t procedure reported in parentheses.

Table C9: Link-formation rules and real networks

	Most Popular	Max	Rawlsian	Ingroup Link
	(1)	(2)	(3)	(4)
Average degree of partners	.162	.024	.005	-.127
	(.092)*	(.132)	(.113)	(.093)
Number of partners	-.025	.017	.018	.010
	(.021)	(.028)	(.025)	(.023)
Const.	1.087	.853	1.097	.547
	(.221)***	(.296)***	(.262)***	(.232)**
Obs.	397	397	397	397

OLS regression. The dependent variable is a variable capturing the number of times a player played consistently with a given rule (specified in the column heading). We estimate the model using data from session 9 onwards. We include all players who report to interact outside of the experiment with at least one person in their session. Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%. Robust standard errors reported in parentheses.

Table C10: Deviations from MBR and reach in final network

	Dep var: reach at the end of the experiment					
	(1)	(2)	(3)	(4)	(5)	(6)
Deviated from MBR	-.44	-.424	.075	.078	-.856	-.916
	(0.189)**	(0.192)**	(0.154)	(0.154)	(0.195)***	(0.211)***
Chose most popular player		-.16		-.226		.257
		(0.169)		(0.146)		(0.185)
Heuristic followed in...	both rounds		first round		second round	
Effect size in sd.	-0.36	-0.34	0.06	0.06	-0.69	-0.74
Obs.	239	239	239	239	239	239

Linear probability model. The dependent variable is the reach of player i at the end of the game. Confidence: *** $\leftrightarrow$ 99%, ** $\leftrightarrow$ 95%, * $\leftrightarrow$ 90%. Standard errors corrected for clustering at the session level.

Table C11: Links and past behaviour of other players

	Link_{ijr}	
	(1)	(2)
Max reach	.176 (.01)**	
Min in-reach	.044 (.351)	
Max in-reach		.163 (.003)***
Min reach		.162 (.001)***
Most popular player	.076 (.012)**	.075 (.065)*
Played according to max rule in round 1	.028 (.477)	.013 (.723)
Played according to Rawlsian rule in round 1	.036 (.549)	-.091 (.218)
Played according to most chosen player rule in round 1	-.026 (.602)	-.007 (.849)
Const.	.069 (.294)	.123 (.072)*
Obs.	600	630
Sample	T1	T2
Cluster N	20	21

Linear probability model. The dependent variable is a dummy which takes a value of one if player i connects to player j. Only round two observations are used for estimation. Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%. Standard errors corrected for clustering at the session level.

Table C12: *Choosing in-group links*

	In-group link	
	(1)	(2)
Identity session	.117	.046
	(.055)**	(.056)
Const.	.282	.268
	(.037)***	(.033)***
Obs.	438	447
Sample	T1	T2
Cluster N	40	41

Linear Probability Model. The dependent variable takes the value of 1 if the subject links with an in-group partner. Decisions taken in the first turn of the first round are not included in the sample. Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%. Standard errors clustered at the session level reported in parentheses.

Table C13: *Choosing 'max' links*

	Max link	
	(1)	(2)
Identity session	-.0003	-.080
	(.065)	(.062)
Const.	.473	.519
	(.044)***	(.045)***
Obs.	438	447
Sample	T1	T2
Cluster N	40	41

Linear Probability Model. The dependent variable takes the value of 1 if the subject links with a player with the maximum reach (in T1) and maximum in-reach (in T2). Decisions taken in the first turn of the first round are not included in the sample. Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%. Standard errors clustered at the session level reported in parentheses.

Table C14: *Choosing 'Rawlsian' links*

	Rawlsian link	
	(1)	(2)
Identity session	-.013	-.038
	(.055)	(.055)
Const.	.600	.649
	(.033)***	(.036)***
Obs.	438	447
Sample	T1	T2
Cluster N	40	41

Linear Probability Model. he dependent variable takes the value of 1 if the subject links with a player with the minimum in-reach (in T1) and minimum reach (in T2). Decisions taken in the first turn of the first round are not included in the sample.Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%. Standard errors clustered at the session level reported in parentheses.

Table C15: Choosing in-group links

Restricted sample

	In-group link	
	(1)	(2)
Identity session	.139	.053
	(.066)**	(.079)
Const.	.298	.242
	(.048)***	(.051)***
Obs.	207	215
Sample	T1	T2
Cluster N	40	41

Linear Probability Model. The dependent variable takes the value of 1 if the subject links with an in-group partner. Sample restricted to 'max' links. Decisions taken in the first turn of the first round are not included in the sample. Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%. Standard errors clustered at the session level reported in parentheses.

Table C16: In-group links and player characteristics in T1

	In-group link		
	(1)	(2)	(3)
Identity session	.095 (.073)	.133 (.077)*	-.014 (.115)
Bias in allocation task	.002 (.084)		
Bias * Identity session	.044 (.124)		
In-group norm		.055 (.080)	
In-group norm * Identity session		-.027 (.097)	
In-group norm expectation			-.014 (.024)
Norm expectation * Identity session			.040 (.032)
Const.	.281 (.051)***	.250 (.067)***	.333 (.098)***
Obs.	438	437	435
Sample	T1	T1	T1
Cluster N	40	40	40

Linear Probability Model. The dependent variable takes the value of 1 if the subject links with an in-group partner. “Bias in allocation task" is a dummy equal to one if in the allocation task the player has allocated more than half of the endowment to the in-group partner. “in-group norm" is a dummy equal to one if the player agrees with the in-group norm. “in-group norm expectation" reports the number of players that the subject expects to agree with the in-group norm. Decisions taken in the first turn of the first round are not included in the sample. Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%. Standard errors clustered at the session level reported in parentheses.

Table C17: In-group links and player characteristics in T2

	In-group link		
	(1)	(2)	(3)
Identity session	.014	.001	-.080
	(.080)	(.073)	(.133)
Bias in allocation task	-.015		
	(.074)		
Bias * Identity session	.054		
	(.111)		
In-group norm		-.011	
		(.082)	
In-group norm * Identity session		.083	
		(.103)	
In-group norm expectation			-.032
			(.021)
Norm expectation * Identity session			.034
			(.040)
Const.	.278	.275	.387
	(.057)***	(.056)***	(.076)***
Obs.	447	447	440
Sample	T2	T2	T2
Cluster N	41	41	41

Linear Probability Model. The dependent variable takes the value of 1 if the subject links with an in-group partner. "Bias in allocation task" is a dummy equal to one if in the allocation task the player has allocated more than half of the endowment to the in-group partner. "in-group norm" is a dummy equal to one if the player agrees with the in-group norm. "in-group norm expectation" reports the number of players that the subject expects to agree with the in-group norm. Decisions taken in the first turn of the first round are not included in the sample. Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%. Standard errors clustered at the session level reported in parentheses.

Table C18: Strategies and comprehension

	Max	Rawlsian	Reciprocal	Most Popular	In-group
	(1)	(2)	(3)	(4)	(5)
Comprehension	.018	.007	.007	-.009	.009
	(.018)	(.017)	(.010)	(.013)	(.016)
	[.697]	[.697]	[.697]	[.697]	[.697]
Const.	.467	.609	.100	.370	.310
	(.024)***	(.022)***	(.012)***	(.016)***	(.022)***
Obs.	885	885	885	885	885
Sample	Full	Full	Full	Full	Full
Cluster N	81	81	81	81	81

Linear Probability Model. The dependent variable takes the value of 1 if the link is consistent with the link-formation rule indicated in the heading. Decisions taken in the first turn of the first round are not included in the sample. Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%. Standard errors clustered at the session level reported in parentheses. False-discovery-rate q-values reported in brackets (Benjamini et al., 2006).

Table C19: *Strategies and comprehension in T1 AND T2 sessions*

	Max		Rawlsian		Reciprocal		Most Popular		In-group	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Comprehension	.036	.017	.018	.021	.004	.006	-.010	-.020	-.033	.028
	(.019)*	(.028)	(.024)	(.024)	(.016)	(.014)	(.021)	(.020)	(.027)	(.026)
	[.31]	[.677]	[.78]	[.657]	[.782]	[.677]	[.782]	[.657]	[.52]	[.657]
Const.	.433	.481	.574	.631	.107	.096	.381	.360	.377	.291
	(.033)***	(.031)***	(.034)***	(.027)***	(.022)***	(.017)***	(.030)***	(.021)***	(.044)***	(.028)***
Obs.	438	447	438	447	438	447	438	447	438	447
Sample	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
Cluster N	40	41	40	41	40	41	40	41	40	41

Linear Probability Model. The dependent variable takes the value of 1 if the link is consistent with the link-formation rule indicated in the heading. Decisions taken in the first turn of the first round are not included in the sample. Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%. Standard errors clustered at the session level reported in parentheses. False-discovery-rate q-values reported in brackets (calculated separately for the T1 regressions and the T2 regressions) (Benjamini et al., 2006).

Table C20: *Strategies and comprehension in T1 sessions, by identity treatment*

	Max	Rawlsian	Reciprocal	Most Popular	In-group
	(1)	(2)	(3)	(4)	(5)
Identity	-.018	-.040	-.089	.117	.109
	(.070)	(.076)	(.045)**	(.051)**	(.085)
Comprehension * identity session = 0	.023	-1.25e-09	-.040	.041	-.030
	(.034)	(.047)	(.025)	(.028)	(.035)
	[.628]	[1]	[.343]	[.343]	[.628]
Comprehension * identity session = 1	.044	.027	.028	-.036	-.027
	(.025)*	(.029)	(.018)	(.027)	(.035)
	[.305]	[.444]	[.305]	[.309]	[.444]
Const.	.446	.600	.167	.305	.318
	(.056)***	(.063)***	(.041)***	(.032)***	(.061)***
Obs.	438	438	438	438	438
Sample	T1	T1	T1	T1	T1
Cluster N	40	40	40	40	40

Linear Probability Model. The dependent variable takes the value of 1 if the link is consistent with the link-formation rule indicated in the heading. Decisions taken in the first turn of the first round are not included in the sample. Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%. Standard errors clustered at the session level reported in parentheses. False-discovery-rate q-values reported in brackets (Benjamini et al., 2006).

Table C21: Strategies and comprehension in T2, by identity treatment

	Max	Rawlsian	Reciprocal	Most Popular	In-group
	(1)	(2)	(3)	(4)	(5)
Identity	-.076	-.035	-.006	-.002	.053
	(.062)	(.054)	(.033)	(.044)	(.056)
Comprehension * identity session = 0	-.001	.023	.019	-.027	.019
	(.038)	(.037)	(.020)	(.030)	(.029)
	[.977]	[.675]	[.675]	[.675]	[.675]
Comprehension * identity session = 1	.028	.014	-.008	-.012	.043
	(.040)	(.029)	(.019)	(.027)	(.042)
	[.668]	[.668]	[.668]	[.668]	[.668]
Const.	.520	.647	.098	.362	.267
	(.046)***	(.036)***	(.025)***	(.029)***	(.033)***
Obs.	447	447	447	447	447
Sample	T2	T2	T2	T2	T2
Cluster N	41	41	41	41	41

Linear Probability Model. The dependent variable takes the value of 1 if the link is consistent with the link-formation rule indicated in the heading. Decisions taken in the first turn of the first round are not included in the sample. Confidence: *** ↔ 99%, ** ↔ 95%, * ↔ 90%. Standard errors clustered at the session level reported in parentheses. False-discovery-rate q-values reported in brackets (Benjamini et al., 2006).