

IZA DP No. 5307

## Network Architecture and Mutual Monitoring in Public Goods Experiments

Jeffrey Carpenter  
Shachar Kariv  
Andrew Schotter

November 2010

# Network Architecture and Mutual Monitoring in Public Goods Experiments

**Jeffrey Carpenter**

*Middlebury College  
and IZA*

**Shachar Kariv**

*University of California, Berkeley*

**Andrew Schotter**

*New York University*

Discussion Paper No. 5307  
November 2010

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0  
Fax: +49-228-3894-180  
E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## **ABSTRACT**

### **Network Architecture and Mutual Monitoring in Public Goods Experiments**

Recent experiments show that public goods can be provided at high levels when mutual monitoring and costly punishment are allowed. All these experiments, however, study monitoring and punishment in a setting where all agents can monitor and punish each other (i.e., in a complete network). The architecture of social networks becomes important when individuals can only monitor and punish the other individuals to whom they are connected by the network. We study several non-trivial network architectures that give rise to their own distinctive patterns of behavior. Nevertheless, a number of simple, yet fundamental, properties in graph theory allow us to interpret the variation in the patterns of behavior that arise in the laboratory and to explain the impact of network architecture on the efficiency and dynamics of the experimental outcomes.

JEL Classification: D82, D83, C92

Keywords: experiment, networks, public good, monitoring, punishment

Corresponding author:

Jeffrey P. Carpenter  
Department of Economics  
Middlebury College  
Middlebury, VT 05753  
USA  
E-mail: [jpc@middlebury.edu](mailto:jpc@middlebury.edu)

# 1 Introduction

A perennial question in economics concerns the conditions under which self-interested individuals cooperate to achieve socially efficient outcomes. In a seminal experiment, Fehr and Gächter (2000) showed that public goods can be provided at high levels if individuals can monitor the contributions made by other individuals and punish those who are unwilling to contribute. This stands in stark contrast to the experimental results from the familiar public good game in the literature, in which low provision is common (Ledyard, 1995). A number of experimental papers extend Fehr and Gächter (2000) by making punishment more or less costly to the monitor (Anderson and Putterman, 2005), making punishment only symbolic (Masclot, et al., 2003), or by going in the opposite direction and equating punishment with expulsion from the group (Cinyabuguma, et al., 2005) and continue to find high levels of provision.

A central assumption of nearly all these experiments is *full monitoring* – everyone can monitor all of the other individuals. In reality, individuals living in any society are bound together by a social network, and often they can only monitor the behavior of those who are in their local environment. If each individual monitors the actions of only a small number of other individuals, it is not clear that mutual monitoring can give rise to cooperative outcomes. Clearly, *partial monitoring* can be an obstacle to cooperation if, for example, a critical mass of potential punishers is required to deter shirking or punishers are only emboldened to intervene when they know that they are supported by others.

We represent the partial monitoring structure by a *graph* that specifies the monitoring technology of the group; that is, who monitors whose actions. Each player is located at a node of the graph, and player  $i$  can monitor player  $j$  if and only if there is an edge leading from node  $i$  to node  $j$ . The experiments reported here involve the eight networks [1]-[8] illustrated in Figure 1 below. An arrow pointing from player  $i$  to player  $j$  indicates that  $i$  can monitor and punish  $j$ . The goal of this paper is to identify the impact of network architecture on the effectiveness of mutual monitoring and punishment. The set of networks depicted in Figure 1 has several interesting architectures exemplifying a number of simple yet fundamental concepts in graph theory (defined in the next section) that allow us to interpret variation in the experimental outcomes.

[Figure 1 here]

Our key results are as follows:

- **Cooperation** – Although contributions vary dramatically from network to network, *connected networks* [1]-[4], within which everyone is monitored, demonstrate signifi-

cantly higher contributions than *disconnected networks* [5]-[8]; however, the *complete network* [1] does not elicit significantly more contributions than the other connected networks.

- **Punishment** – More punishment is used to maintain or increase contributions in *directed networks* such as [2], [5], [6], and [7] where the edges point in only one direction relative to *undirected networks*. Our conclusion is that the asymmetry in the relations between any pair of nodes in directed networks gives different monitoring roles to different subjects, which, in turn, increases punishment expenditures.
- **Efficiency** – While there is also considerable variation in net payoffs after subtractions for punishing others and for being punished across networks, the connected networks (such as [1] and [4]) are the most efficient, whereas the disconnected networks (e.g., [6] and [7]) are the least efficient. In addition, adding/removing edges does not necessarily increase/decrease efficiency.

Among our other conclusions, the fact that we find it is necessary to take into account the details of the local neighborhood as well as the entire network architecture to explain individual behavior is particularly relevant for future work. The simple summary characteristics of the networks depicted in Figure 1, such as the average distance between players, do not fully account for the subtle and complicated behaviors that we observe. To determine the important determinants of individual behavior, it will be necessary to investigate a larger class of networks in the laboratory. This is perhaps one of the most important topics for future research.

The paper contributes to the enormous body of experimental work on public goods, but we will not attempt to review this literature. The most closely related papers to ours are those that also allow for costly punishment by Carpenter (2007) and O’Gorman, et al. (2008). Carpenter (2007) compared the complete network [1] to two other connected networks, the directed circle [2] and the undirected circle [3]. Carpenter (2007) found that contributions in the complete network [1] are as high as in the undirected circle [3] but are significantly higher than in the directed circle [2]; however, the number of other potential punishers, not the network structure was emphasized. O’Gorman, et al. (2008) compared the complete network [1] to a version of the directed star [5] in which the responsibility of being the “prison guard” changed randomly each round and found that the complete network is less efficient. These results, though based on different designs and reporting some difference in results are important primarily because they suggest that the network architecture affects the provision of the public good. Nevertheless, a more sophisticated and comprehensive analysis is required to detect the underlying properties of networks that facilitate or hinder

cooperative outcomes. Our paper also contributes to the large and growing literature on the economics of networks (see Jackson, 2008) and the smaller, more recent, literature on network experiments (see Kosfeld, 2004).

The rest of the paper is organized as follows. The next section presents the network concepts that guided our experimental design and the design itself. Section 3 provides our empirical analysis, Section 4 discusses the results and Section 5 provides some concluding thoughts.

## 2 Experimental Design

The voluntary contribution game that we study can be interpreted as follows. At the first stage, players simultaneously make voluntary contributions to a public good. The payoff for each player at the first stage equals his consumption of the public good plus his remaining endowment. At the end of the first stage, players are allowed to monitor the contributions of the players to whom they are connected by a social network, where nodes represent players and edges represent the possibility of one player monitoring the player to whom the edge is pointing. Thus, we drop the standard assumption that individual contributions are public information and assume that players can monitor the contributions of some, but not necessarily all, of the other players. In the second stage, players are given the opportunity to punish, at some cost, the other players to whom they are connected by the network. The terminal payoff for each player from both stages is given by the maximum of either zero or her payoff from the first stage minus the punishment received and the cost of punishing other players.

### 2.1 The networks

We restrict attention to the case of four-person networks, which has several non-trivial architectures. Each network is represented by a graph with four nodes, indexed by  $i = A, B, C, D$ , where at each node there is a single player. An edge between any two players represents that they are connected and the arrowhead points to the player whose action can be monitored. For each player  $i$ ,  $N_i$  denotes the set of players  $j \neq i$  who can be monitored by  $i$ . We can think of  $N_i$  as representing player  $i$ 's *neighborhood*. The collection of neighborhoods  $\{N_A, N_B, N_C, N_D\}$  completely define a four-person network. The set of networks used in the experimental design is illustrated in Figure 1. Note that edges can be directed: the fact that player  $i$  can monitor player  $j$  ( $j \in N_i$ ) does not necessarily imply that  $j$  can monitor  $i$  ( $i \in N_j$ ).

Next, we define some key graph theoretic concepts to which we refer throughout the

paper. Our notation and definitions are standard, but to avoid ambiguity we present the concepts in some detail.

- A *complete* network is a network in which each pair of nodes is connected by an edge. Otherwise, the network is *incomplete*. Referring to Figure 1, in the complete network [1] every player directly monitors every other player. The rest of the networks [2]-[8] we study are incomplete.
- A network is *connected* if every pair of players  $i$  and  $j$  is linked by a path and *disconnected* otherwise.<sup>1</sup> Obviously, players from disconnected components of a network cannot monitor each other. Referring to Figure 1, networks [1]-[4] are connected whereas networks [5]-[8] are disconnected. Networks [7] and [8] are disconnected, but connectedness is satisfied in a subgraph  $\{N_A, N_B, N_C\}$  in which every pair of players  $i$  and  $j$  are connected.
- A network is *undirected* if the relations between any pairs of nodes is symmetric, so that each edge points in both directions. Otherwise, the network is *directed*. A directed network in which each edge is given a unique direction is called an *oriented* network. In an oriented network, if player  $i$  can monitor player  $j$  ( $j \in N_i$ ) then  $j$  cannot monitor  $i$  ( $i \notin N_j$ ). As shown in Figure 1 above, in our experimental design all directed networks [2], [5], [6], and [7] are oriented. Networks [1], [3], [4], and [8] are undirected as all edges are bidirected and point to both nodes at once.
- The *degree* of a node is the number of edges that end at that node. In a directed graph the degree is usually divided into the *out-degree* and the *in-degree*. The out-degree (resp. in-degree) of node  $i$  is the number of edges with  $i$  as their initial (rep. terminal) node. Clearly, the out-degree of player  $i$  is the number of players  $j$  that can be monitored by  $i$  ( $j \in N_i$ ) and the in-degree of player  $i$  is the number of players  $j$  that can monitor  $i$  ( $i \in N_j$ ).

## 2.2 The game

The game is formally described using the following notation. Each player  $i = A, B, C, D$  is endowed with  $y$  indivisible tokens. At stage one, the players simultaneously choose how many tokens  $0 \leq g_i \leq y$  to contribute to the provision of the public good. The payoff for

---

<sup>1</sup>Put precisely, for any pair of players  $i$  and  $j$ , a *path* from  $i$  to  $j$  is a sequence  $i_1, \dots, i_K$  such that  $i_1 = i$ ,  $i_K = j$  and there is an edge pointing from  $i_k$  to  $i_{k+1}$  for  $k = 1, \dots, K - 1$ . Player  $i$  is connected to  $j$  if there is a path from  $i$  to  $j$ .

each player  $i$  in the first stage can be summarized by

$$\pi_i^1 = y - g_i + \alpha \bar{g}, \quad (1)$$

where

$$\bar{g} = \sum_{j=A,B,C,D} g_j$$

and  $\alpha$  is the marginal per capita return (*MPCR*). Hence, each player receives the value of the public good ( $\alpha \bar{g}$ ) in addition to the number of tokens retained from his endowment ( $y - g_i$ ). To avoid trivialities, we assume that  $1 > \alpha > \frac{1}{4}$ . This condition ensures that contributing is, on one hand, socially efficient, and on the other hand, strictly dominated for any individual player.

At stage two, after players are informed about the individual contributions of the members of their group, they can punish the players to whom they are connected in the network. More precisely, each player  $i$  can punish player  $j \in N_i$  by reducing his payoff from the first stage  $\pi_j^1$  by  $p_i^j$  tokens. Reducing the payoff of other players is costly. The cost of reducing one token from any of the other players is  $0 < c < 1$  tokens. We also assume that each player  $i$  can spend up to his entire payoff from the first stage  $\pi_i^1$  towards reducing the payoff  $\pi_j^1$  of all  $j \in N_i$ , and that  $\pi_i^1$  can be reduced at the second stage to zero but the terminal payoff for the game cannot be negative. The payoff of player  $i$  from both stages of the game can therefore be summarized by

$$\pi_i = \max \left\{ 0, \pi_i^1 - c \sum_{j \in N_i} p_i^j - \sum_{j: i \in N_j} p_j^i \right\}. \quad (2)$$

By backward induction, it follows that punishment cannot defer free riding, in any network architecture. We next briefly illustrate the logic of the backward induction argument and then draw out the important implications of the theory. Since punishing is costly, each player  $i$  will refrain from doing so at the second stage (i.e.,  $p_i^j = 0$  for all  $i$  and  $j \in N_i$ ). Because each player  $j \in N_i$  expects that player  $i$  will never punish him, his best response is to contribute nothing in the first stage ( $g_j = 0$ ). Thus, the addition of the second stage has no effect on the outcome of the first stage which, in this finitely repeated version of the game, is full free-riding and therefore the prediction of standard theory is that  $g_i = 0$  and  $p_i^j = 0$  for all  $i$  and  $j \in N_i$ . Note, however, that the aggregate payoff is maximized if each player  $i$  fully cooperates by contributing his endowment ( $g_i = y$ ).

### 2.3 Procedures

The experiment was run with 264 participants in 14 sessions at the Center for Experimental Social Sciences (CESS) at New York University and at the Experimental Social Science

Lab (Xlab) at the University of California, Berkeley. The subjects in this experiment were recruited from all undergraduate classes and had no previous experience in public good or networks experiments. After subjects read the instructions, they were read aloud by an experimenter.<sup>2</sup> At the end of the instructional period subjects were asked if they had any questions or any difficulties understanding the experiment. No subject reported any difficulty understanding the procedures or using the computer program. Each experimental session lasted about an hour and a half. A \$5 participation fee and subsequent earnings from playing the game were paid in private at the end of the experimental session.

The endowment was given by  $y = 25$  and the marginal per capita return and the cost of punishing were fixed at  $\alpha = 0.4$  and  $c = 0.5$ , respectively. The network was held constant throughout a given experimental session. In each session, the network positions were labeled  $A$ ,  $B$ ,  $C$ , and  $D$ . The participant's type ( $A$ ,  $B$ ,  $C$ , or  $D$ ) remained constant throughout the session. Each experimental session consisted of 15 independent decision-rounds. To minimize the investment in reputations, each round started with the computer randomly forming four-person networks by selecting one participant of each type. The networks formed in each round depended solely upon chance and were independent of the networks formed in any of the other rounds.

Each round of the experiment consists of two stages, the contribution stage and the punishment stage. The contribution decision was to allocate the endowment between a private good which only benefited the individual and a public good which benefited everyone in the group, according to the payoff function (1). Once all the contributions were recorded, subjects observed the contributions of the participants to whom they were connected by the network. In addition, all subjects were informed about the sum of the contributions to the public good by all the participants in their group (including themselves). In the punishment stage, subjects choose if and by how much to reduce the first stage payoff of each of the other participants with whom they were connected by the network. If they did not wish to reduce the earnings of another participant they had to enter zeros. At the completion of the two stages, the computer informed the subjects of their total payoffs according to the payoff function (2).

The various experimental roles are summarized in Table 1 below. The entries in the right hand column have the form  $a/b$  where  $a$  is the number of observations and  $b$  the number of subjects per type.

*[Table 1 here]*

---

<sup>2</sup>Sample experimental instructions, including the computer program dialog windows are available in the Appendix and online at ([http://emlab.berkeley.edu/~kariv/CKS\\_II\\_A.pdf](http://emlab.berkeley.edu/~kariv/CKS_II_A.pdf)).

### 3 Results

In this section we report the results of our experiment. Our analysis focuses on two important questions. First, we ask the obvious question of whether the graph theoretic constructs defined in Section 2 above systematically affect the performance of our networks. Performance is measured in three ways: contribution levels, punishment given and overall payoff efficiency (i.e., payoffs net of punishment expenses). Second, we ask the more subtle question of whether networks “matter” in that people’s behavior is affected not only by their immediate surroundings (their node) but also by the larger network in which that node functions. Put differently, if a participant’s node is defined as the number of out- and in-links connecting him to the network, we ask whether his behavior changes as we take that node and embed it in different networks.

Our experimental design allows us to make direct comparisons to test the importance of each of the properties of network architecture. For example to test the importance of the network being complete, we compare [1] the Complete network to two other networks that are incomplete: [3] the Undirected Circle and [4] the Undirected Star. This comparison is salient because, while both [3] and [4] are incomplete, they are connected and undirected like [1]. Considering connectedness, we compare [2] the Directed Circle to [6] the Line because removing just one link renders the Directed Circle disconnected. We also compare [2] to [7] the disconnected Directed Circle in which player D stands alone, and [3] the Undirected Circle to [8] the disconnected Undirected Circle, another comparison in which D can escape any punishment. To examine the impact of networks being directed, we compare [2] the Directed Circle to [3] the Undirected Circle, [5] the Prison Guard to [4] the Undirected Star, and [7] the disconnected Directed Circle to [8] the disconnected Undirected Circle. The only complication here is that the comparison of [5] to [4] might be confounded by the fact that [5] is also disconnected while [4] is not.

Because degree is mostly a property of nodes (i.e., the number of out- and in-links), its examination effectively coincides with and motivates our second question: does the behavior at a node depend on the network in which it is embedded? In all networks any given agent operates in a neighborhood defined by the number and directionality of the links connected to his node in the network. This agent’s node can be embedded in many different networks, however. Return to Figure 1, for example, and take player B in the Undirected Circle. This player has two out-links and two in-links to agents A and C. This defines the degree of his node. Now take the Disconnected Undirected Circle and look at player B there. That player occupies exactly the same node as agent B in the Undirected Circle. In fact, players A and C in both networks also occupy the same node as agent B in the Undirected Circle.

What differs is the fact that these neighborhoods are embedded in different networks since player D is isolated in the Disconnected Undirected Circle but connected symmetrically to the network in the Undirected Circle. If networks matter then we would expect that the behavior of subjects at identical nodes but different networks would differ. This would mean that, when determining their behavior, subjects look beyond the narrow boundaries of their local neighborhood and take into account the more global aspects of the network their node is embedded in.

We proceed by systematically evaluating these four properties of network architecture. Along the way we will use a combination of nonparametric rank sum tests ( $|z|$ ) and proportions tests ( $|r|$ ). Where appropriate, we will also run parametric regressions that account for individual heterogeneity (using random effects), learning (with period fixed effects) and add the appropriate controls.

### 3.1 Completeness

Before examining the effect of completeness within our experiment, it is important to demonstrate that there is nothing special about our procedures that could confound our results. In other words, our results are more convincing if we can show that behavior in our complete network is similar to what has been found in other experiments. Figure 1 is an effective summary of our results. Statistics for three metrics of group performance are reported directly next to each network so that one can begin to map outcomes onto architecture. Average contributions in Figure 1 are measured as a fraction of the total endowment. On average, participants contributed 56% of their endowment in our complete network. This behavior is similar to the “stranger” contribution levels found by Fehr and Gaechter (2000) and Carpenter (2007) of 58% and 61%, respectively.

*[Figure 2 here]*

While the Complete network elicits contributions in line with other experiments it does not yield the highest average contributions. Indeed, the Directed Circle generates the highest average contribution of 60% of the endowment, a rate that is statistically greater than all networks except the Complete. The comparisons of our undirected connected networks [1] to [3] and [4] are the most interesting, however, because the only difference is “completeness”. Considering the summary statistics in Figure 1 and the direct comparisons in Figure 2(a), it does not appear that the complete network robustly results in higher contributions. While the mean contribution pooled across rounds is slightly higher in [1] it is not significantly higher than [3] and only marginally significantly higher than [4] ( $|z| = 0.64$ ,  $p = 0.52$ ;  $|z| = 1.99$ ,  $p = 0.05$ , respectively). In fact, when we control for individual heterogeneity and

repeated play in column (1) of Table 2 (in which [1] is the omitted network), we do not find any significant differences in contributions.

*[Table 2 here]*

Figure 1 also lists the unconditional probability that a “target” who contributes nothing will be punished in each of the networks. In the Complete network [1] total free-riders are punished 48% of the time and this rate is only slightly lower in [3] (42%) or [4] (47%). Proportions test suggest that these rates are not different ( $|r| = 0.47$ ,  $p = 0.64$ ;  $|r| = 0.18$ ,  $p = 0.86$ , respectively). While the incidence of punishment might not vary between [1] and [3] or [4], the severity does. Summarizing punishment is tricky because punishment only makes sense in the context of contributions. In Figure 2(b) we plot the estimated relationship between a target’s contribution and how much the target was punished by an individual for networks [1], [3], and [4]. As in Carpenter and Matthews (2009) we utilize a spline specification to allow punishment to diminish more quickly above the implied contribution norm. In this case the “knot” that maximizes the regression F statistic, and in this sense fits the data best, occurs when 10 units are contributed - participants are punished increasingly more below ten than above it. As one can see, punishment levels are much lower in the Complete network. This finding is confirmed in column (4) of Table 2 in which we regress positive punishment amounts on network indicators and find that the estimated punishments in [3] and [4] are significantly larger than in the baseline [1].

There are two possible explanations for the punishment differences we see between the Complete network and its closest comparisons. First, punishment levels might be lower in complete networks because monitors are more inclined to free-ride on the punishment efforts of the other players when they know that everyone is watching everyone else. However, this does not "jibe" with the fact that the incidences of punishment do not seem to differ across networks [1], [3], and [4]. Perhaps a better explanation is that monitors face a coordination problem when there are many sets of eyes watching a target. Without directly communicating to settle how much punishment should be levied and how it should be shared among the monitors, everyone continues to punish but they are each forced estimate, on their own, how much to reduce the severity of punishment to account for the actions of the other monitors.

Combined, relatively high contribution levels and low punishment expenditures make the complete network one of the most efficient architectures. The mean pooled payoff (net of the costs of being punished and punishing) in [1] is significantly higher than in [3] ( $|z| = 4.13$ ,  $p < 0.01$ ) and marginally significantly higher than in [4] ( $|z| = 1.61$ ,  $p = 0.10$ ). These results are mostly replicated in column (5) of Table 2 wherein individual random effects and time

period fixed effects are controlled for. As one can see in Figure 2(c) the performance of the Complete network tends to improve relative to [3] and [4] as the experiment proceeds.

In sum, the Complete network does not seem to be better at eliciting contributions but, because its punishment levels tend to be lower, it does achieve higher than average efficiency.

### 3.2 Connectedness

To compare connected to disconnected networks, one can compare the left column of Figure 1 to the right column. The contribution differences are striking: there is no disconnected network that yields higher mean pooled contributions than the lowest performing connected network, the Undirected Star. However, to conduct our analysis systematically, remember we need to compare [2] to [6], [2] to [7] and [3] to [8] directly which we do in panels (a-c) of Figure 3. In each comparison, the connected graph yields significantly higher pooled contributions ( $|z_{2-6}| = 14.00$ ,  $p < 0.01$ ;  $|z_{2-7}| = 7.31$ ,  $p < 0.01$ ;  $|z_{3-8}| = 4.38$ ,  $p < 0.01$ ) and these results are mostly robust to an analysis that controls for individual heterogeneity and repeated game effects. If one compares the point estimates in column (1) of Table 2 we find that the coefficient on [2] is higher than [6] ( $p < 0.01$ ), [2] is higher than [7] ( $p < 0.01$ ) but while the difference between the point estimates on [3] and [8] goes in the anticipated direction, it is not significant at standard levels ( $p = 0.23$ ).

*[Figure 3 here]*

Although the evidence is mixed, connected networks also appear to elicit at least as much punishment, both in terms of the incidence and the level. Figure 1 suggests that the probability of punishing a complete free-rider is higher in [2] than in [6] ( $|r| = 3.49$ ,  $p < 0.01$ ), higher in [3] than in [8] ( $|r| = 4.38$ ,  $p < 0.01$ ) and no lower in [2] than in [7] ( $|r| = 1.10$ ,  $p = 0.27$ ). The full punishment splines, panels (a-c) of Figure 4, which are based on an analysis of all the punishment data (i.e., they include the zeros) seem to indicate that there is more punishment in connected networks. However, considering only the positive observations in column (4) of Table 2 the differences do not appear to be robustly significant.

*[Figure 4 here]*

There also appears to be mixed evidence of an efficiency advantage in connected networks. While Figure 1 indicates that [2] yields higher mean payoffs than [6] or [7] ( $|z| = 5.20$ ,  $p < 0.01$ ;  $|z| = 2.49$ ,  $p < 0.01$ , respectively), the payoffs in [8] actually tend to be higher than in [3] ( $|z| = 1.81$ ,  $p = 0.07$ ). Looking at the difference in mean payoff over the 15 rounds of the game in Figure 5(a-c), we see little evidence of differences, results which are

substantiated by comparing the coefficients in column (5) of Table 2. Here only the [2]-[6] payoff difference is significant.

To summarize, although it is clear that connected networks tend to achieve higher contribution levels, connectedness does not necessarily lead to more punishment or higher average efficiency.

*[Figure 5 here]*

### 3.3 Directedness

Directed links occur when one person can monitor another. Links become undirected when both people can monitor each other. By comparing [2] to [3], [5] to [4] and [7] to [8] we can examine the effect of monitoring being one-way or mutual. The role of directedness seems to depend, more than connectedness, on the basic architecture of the network. While studying Figures 1 and 3(d-f) one can see that in the connected setting the directed [2] elicits higher average contributions than the undirected [3] ( $|z| = 2.20$ ,  $p = 0.03$ ) and the undirected [4] and [8] yield higher contributions in the star structure [5] and when one group member is isolated [7] ( $|z| = 11.13$ ,  $p < 0.01$ ;  $|z| = 3.76$ ,  $p < 0.01$ , respectively). Only the [5] versus [4] difference remains significant ( $p < 0.01$ ), however, when, in column (1) of Table 2, we include individual random effects and time period fixed effects.

Comparing the probabilities of punishing a total free-rider in Figure 1 and the punishment splines in Figure 4(d-f) we also find that the effect of directedness on punishment is architecture-dependent. Because [2] yields the highest likelihood of a total free-rider being punished, it is clearly higher than [3]. However, the difference is not quite statistically significant ( $|r| = 1.39$ ,  $p = 0.16$ ) because there are very few observations of total free-riding in these two networks. The effect of directedness reverses in this domain as well (i.e., free-riders are more likely to be punished in the undirected network) when one compares [5] and [4] but does not achieve statistical significance ( $|r| = 1.52$ ,  $p = 0.13$ ). As one can see in Figures 1 and 4(f), there is a large difference in the chances of a total free-rider being punished in the disconnected circles. Here the directed graph yields both a higher instance of punishment ( $|r| = 3.38$ ,  $p < 0.01$ ) and, according to the comparison of point estimates from column (4) of Table 2, a significantly higher level of punishment ( $p < 0.01$ ).

The marginally significant contribution and punishment differences between the directed and undirected networks combine to provide significant payoff differences. Although not obvious from Figures 5(d-f), the directed [2] results in higher average payoffs than the undirected [3] ( $|z| = 2.06$ ,  $p = 0.04$ ), the undirected [4] yields higher payoffs than the directed [5] ( $|z| = 7.00$ ,  $p < 0.01$ ) and the undirected [8] yields higher payoffs than the directed [7]

( $|z| = 3.38, p < 0.01$ ). Except for the comparison between [2] and [3], these payoff differences remain significant in column (5) of Table 2.

More than both completeness and connectedness, the effect of directedness seems to depend on the basic structure of the network. When the underlying network is connected, there is some evidence that directed links lead to more cooperation, punishment and higher payoffs. When the underlying structure is star-shaped, having a disconnected “prison guard” is particularly bad for contributions and payoffs. Lastly, when both underlying networks include a completely disconnected agent we find the opposite of the connected case: directed links in this structure lead to much more punishment but not more cooperation or higher payoffs.

### 3.4 Degree

Recall that degree is mostly a nodal property and while our experiment was not designed to systematically add or remove links (mostly because this would confound comparisons of the other properties) we can now use the concept to transition from an analysis of the overall performance of a network to analyzing the effect of broader network structures on the behavior of agents who occupy nodes of common degree, i.e., local “neighborhoods”. We return to an examination of the effects of the total number of links, *per se*, in the discussion. The question then is do networks have an impact on the contribution and punishment behavior of participants inhabiting identical nodes? The answer is that when the underlying structures change significantly they clearly do.

Since nodes can be defined by the number of out- and in-links facing an agent, we can index them as  $N_{n,o,i}$ ,  $n = 1, 2, \dots, 8$ ,  $o = 0, 1, 2, 3$  and  $i = 0, 1, 2, 3$  where  $N_{n,o,i}$  indicates a subject in network  $n$  who has  $o$  out-links and  $i$  in-links. For example,  $N_{6,1,1}$  is the neighborhood of subjects B and C in network [6] (the Line) since each of them have 1 out- and 1 in-link. Note that players A, B, C, and D in network [2] also have 1 out- and 1 in-link (they are all defined as  $N_{2,1,1}$ ) so they also share the same node as subjects B and C in network [6] but they are in different networks. For notational convenience, we simply leave  $n$  unspecified when we consider nodes outside the context of their networks. For example,  $N_{n,1,1}$  denotes the generic 1 out- 1 in-link node. Table 3 catalogues the nodes that exist in our experiments, the networks they are part of and the subjects types that inhabit them.

[Table 3 here]

By design, most nodes exist in several networks; however, some nodes are rare and have only one representative as, for example,  $N_{n,3,0}$  whose only inhabitant is subject A in the Prison Guard [5]. The most frequent node is  $N_{n,1,1}$  since it occurs in four networks: [2], [4],

[6], and [7]. Other flexible neighborhoods are  $N_{n,3,3}$  in networks [1] and [4],  $N_{n,2,2}$  in both networks [3] and [8], the  $N_{n,0,0}$  isolated subjects in networks [7] and [8], and  $N_{n,0,1}$  existing in networks [5] and [6]. Any other nodes exist in only one network and hence will not be discussed.

As catalogued in Table 3, five of the seven nodes we study exist in more than one network. Differences in the behavior and outcomes of agents who occupy the same node in different networks are depicted in Figure 6. To test for differences in nodal outcomes by network we use nonparametric tests and regress contributions, positive instances of punishment, and payoffs on node indicators in columns (1-3) of Table 4.

*[Figure 6 here]*

Let us first discuss the  $N_{n,3,3}$  neighborhoods. Because neither the summary test ( $|z| = 0.43$ ,  $p = 0.67$ ) indicates a difference nor are the point estimates different ( $p = 0.98$ ) in the first column of Table 4, contributions do not seem to differ for  $N_{n,3,3}$  players who find themselves in [1] or [4]. However, the  $N_{n,3,3}$  players who find themselves at the center of the star in [4] do punish more according to the results in the second column of Table 4 ( $p = 0.05$ ) and as a result accrue lower payoffs ( $p = 0.01$ ) than the players in [1]. Clearly, the punishment responsibilities are more salient to the player at the star's center. In this sense the structure changes significantly between  $N_{1,3,3}$  and  $N_{4,3,3}$ .

*[Table 4 here]*

The  $N_{n,2,2}$  node exists in networks [3] and [8]. The difference between these networks is that one player is completely disconnected in [8] and all players are connected in [3]. Figure 6(b) and the summary tests suggest differences in contributions and punishment between occupiers of the  $N_{n,2,2}$  node in these two networks. Players in the connected network contribute and punish more ( $|z| = 3.88$ ,  $p < 0.01$ ,  $|r| = 1.81$ ,  $p = 0.07$ )<sup>3</sup>. However, in terms of payoffs, the larger  $N_{3,2,2}$  contributions and punishments tend to “net out” out because there is no evidence of payoff differences between the two  $N_{n,2,2}$  nodes ( $|z| = 1.11$ ,  $p = 0.27$ ). These results suggest that having someone completely disconnected from the monitoring network affects the outcomes for the subset of connected nodes. In other words, the connected players in [8] do not simply ignore player D. In fact, one could imagine that D becomes a scapegoat because she does not have to fear punishment and therefore it would be easy for her to free-ride. Ironically, Ds in [8] actually appear to play with considerable integrity because,

---

<sup>3</sup>While these differences survive the inclusion of period fixed effects in columns (1) and (2) of Table 4, the relatively large rhos reported in Table 4 suggest that the substantial individual heterogeneity can reduce their significance.

in fact, they contribute at levels comparable to the other three connected members of the network ( $|z| = 0.86$ ,  $p = 0.39$ ).

We see the most variation in behavior among the  $N_{n,1,1}$  nodes. The mean fraction contributed of subjects in the neighborhoods of the Directed Circle ( $N_{2,1,1}$ ) was 0.60 while that same node, when existing in either the Undirected Star ( $N_{4,1,1}$ ), the Line ( $N_{6,1,1}$ ), or the Disconnected Directed Circle ( $N_{7,1,1}$ ), led to contribution levels of just 0.49, 0.27, and 0.38 respectively. Using a Kruskal Wallis test we find a significant impact of network on nodal performance at better than the 1% level ( $\chi^2 = 177.37$ ,  $p < 0.01$ ).<sup>4</sup> In short, 1-out and 1-in link nodes elicit high contributions when embedded in connected networks with the Directed Circle [2] being the most hospitable network for such neighborhoods and the Line [6] being the most inhospitable network.

The impact of network structure on nodal performance can also be seen in the punishment behavior of  $N_{n,1,1}$  participants since while the mean punishment sent by subjects in the  $N_{2,1,1}$  neighborhoods of [2] is 2.67 it is, 0.78, 1.10, and 2.26, respectively for the Undirected star ( $N_{4,1,1}$ ), the Line ( $N_{6,1,1}$ ), and the disconnect Directed Circle ( $N_{7,1,1}$ ). Another Kruskal-Wallis test indicates that these differences are significant ( $\chi^2 = 24.33$ ,  $p < 0.01$ ).<sup>5</sup> Note that while punishments are highest in the connected [2], they are lowest in [4] which is also connected. This is not too surprising since in that network there are three subjects who have the opportunity to punish the one subject in the center of the star and such a plethora of punishers provides another coordination problem as to who will be responsible for punishing the subject in the center of the star. Another factor that might account for the relatively low punishment in [4] is the fact that  $N_{n,1,1}$  players here have mutual links: not only can they monitor and punish their neighbors, their neighbors can punish them. Occupiers of the same node in [2], [6], and [7] punish one neighbor and are punished by another.  $N_{n,1,1}$  monitors might show more restraint compared to the others just because they are afraid of engaging in punishment feuds (Nikiforakis, 2008). It is also interesting that the mean punishment levels do not differ much between [2] and [7], especially because the monitors in [7] form their own 3-person version of [2]. Again, however, it appears that the existence of a disconnected player (D) reduces contributions.

In terms of payoffs, the  $N_{n,1,1}$  node clearly does best in connected networks.<sup>6</sup> If one were to look for a common feature to explain this difference it might be that in disconnected networks

---

<sup>4</sup>If done pairwise, all the differences are significant using the nonparametric rank sum tests, and most of them survive the addition of individual random effects and time period fixed effects in column (1) of Table 4.

<sup>5</sup>Here the  $N_{2,1,1}$  versus  $N_{7,1,1}$  and  $N_{4,1,1}$  versus  $N_{6,1,1}$  comparisons do not survive when the analysis is done pairwise.

<sup>6</sup>While all the pairwise comparisons are significant using the nonparametric test, only the differences between the connected and disconnected networks survive the analysis in column (3) of Table 4.

like 6 and 7 there exists one person who is not monitored and cannot be punished. So despite the fact that locally subjects in these nodes seem to be playing the same monitoring-punishment game, globally they realize that there is one player that has no incentive to contribute. In this sense, the broader network matters.

There are two remaining types of nodes to compare: the  $N_{n,0,1}$  and the  $N_{n,0,0}$ . Because they all involve agents who do not punish, the four instances of these nodes are combined in Figure 6(d). The nonparametric test suggests that players B, C, and D contribute significantly more in [5] than players D do in [6] ( $|z| = 4.18, p < 0.01$ ). This can be accounted for by the higher incidence of punishment in [5]. The behavior of players D does not appear to depend on being in network [7] or [8]. To some extent this result makes sense because while [5] and [6] are very different architectures (i.e., the punishment responsibilities are consolidated in [5]), there is no reason to believe *ex ante* that D’s experience should be different: she is completely isolated in both networks and only receives feedback on the group mean contribution.

The final aspect of nodal behavior that we can examine is the question of which links matter the most for contributions. Is it in-links or out-links? As one might expect, in general subjects who are not monitored contribute less than those who are and the more players are monitored, the more they contribute. For example, in the Line subjects A ( $N_{6,1,0}$ ) cannot be monitored but can monitor subjects B. Since  $N_{6,1,0}$  cannot be monitored we might expect that she contributes very little, which in fact is true since these players contribute only 20% of their endowments on average and this is the lowest mean contribution of any type. Regressing contributions on the number of in-links, out-links and indicators for time periods supports the fact that contributions are increasing in the number of in-links associated with a subject’s neighborhood but rejects the hypothesis that out-links affect contributions ( $\beta_{in-link} = 3.51, p < 0.01$ ). For example, compared to A in [6] who has one out-link, D in [7] is more isolated -  $N_{7,0,0}$  can neither punish nor be punished and yet subjects in that position have a mean contribution level of 0.37.

Networks do seem to matter because nodal behavior appears to be affected by significant changes in the architecture. For example, When punishment responsibilities are consolidated in player  $N_{4,3,3}$ , she appears to “step up” and punish more compared to the  $N_{1,3,3}$  monitors. Further, there is some evidence that nodes placed in one-way monitoring relationships elicit more punishment than when the same node is placed in mutual monitoring relationships.<sup>7</sup> The simple explanation for this is that people want to avoid sparking feuds with their neighbors. Lastly, we see that the existence of a completely disconnected player

---

<sup>7</sup>The incidence of punishing a total free-rider is significantly higher when the links are one-way ( $|r| = 2.53, p = 0.01$ ).

can also significantly affect performance at other nodes in the network.

## 4 Discussion

In the previous section we examined each network property in detail, one at a time. However, it might be helpful to combine the effects of all the properties in one summary analysis. Is it the case that these properties provide the foundation for the successes and failures seen in Figure 1? In Table 5 we regress our three measures of network performance on all four factors. Recall that there can be only one complete network for any fixed group size so “complete” is an indicator for network [1]. “connected” is an indicator that is one for the networks on the left side of Figure 1 and it is zero for those on the right. Four of the eight graphs (2, 5, 6, 7) are “directed” and we add an aggregate measure of “degree” by calculating the total number of edges for each structure.

*[Table 5 here]*

Considering contributions in column one, when we control for the other properties it appears that only one factor seems to matter: connected networks yield significantly more contributions than disconnected networks. The effect is not only statistically significant, five more tokens contributed in the connected networks out of a twenty-five token endowment is substantial.

Two factors, Complete and Directed, affect the amount of punishment that is doled out to targets. Together, these point estimates suggest an interesting interpretation. The fact that complete networks, in which the policing of free riders is very decentralized yield lower sanctions and directed networks in which it is the responsibility of just one person to punish a free rider yield higher sanctions suggests that punishers face the coordination problem mentioned above. Everyone agrees that free-riders should be punished, but without communicating directly, it is hard to know how much to punish in decentralized networks like the Complete [1].

When considering efficiency in column 3, it might be the case that punishment is too severe in directed networks and a little punishment ambiguity helps payoffs in the Complete network. When combined, all four properties significantly affect final net payoffs. Not only do those in the Complete [1] do well and those in directed graphs suffer because of differences in the amount of punishment, more links reduces payoffs perhaps because they tend to be used too often to punish and connected graphs do better. The effect of connected graphs is particularly interesting because it appears even after controlling for the punishment

dynamics. In other words, part of the success of connected networks occurs for reasons beyond their ability to distribute punishment.

Given the differences between networks operate on the ability to monitor and punish other players, it is important to examine the extent to which network differences in punishment that arise can explain the differences that we see in contributions. Returning to the contribution regression analysis in Table 3, in the second column we add the lag of received punishment and the lag of contributions. We add the lag of contributions to control for level differences: while free riders can increase their contributions substantially, high contributors can only increase their contributions slightly regardless of how much punishment they receive. As expected, punishment is highly significant. For each (lab) dollar of punishment the average player receives, she increases her contribution by twenty cents. As important, however, is the fact that the addition of punishment has reduced, compared to column (1), all the coefficients on the network indicators, some substantially. Clearly, a large part of the variation previously attributed to the networks is really due to differences in the amount of punishment generated by the different networks. In column (3) of Table 3 we examine a robustness check on our punishment analysis. The model in column (2) assumes that participants have a common response to punishment. In column 3 we add (unreported) interactions that allow the response to punishment to vary by network. In addition to the added interactions also not changing the results much, a chi-squared test of the joint significance of the interactions suggests that they add nothing to column 2 ( $\chi^2 = 2.85$ ,  $p = 0.90$ ) and so assuming a common response to punishment seems reasonable. In short, although some differences remain, a major reason for the differences seen in contributions across the networks is that the some networks elicit a lot of punishment and others do not.<sup>8</sup>

*[Figure 7 here]*

We end our discussion by speculating, based on our results, about the “optimal” number of links in the standard four-person public good game with punishment. If, in the end, we are mostly concerned about the efficiency with which public goods are provided then we have to look at payoffs net of the costs of punishing and being punished. If we estimate the effect of the number of links on net payoffs we find that all three coefficients of a third order polynomial specification are significant. The derivative of this function provides us with an estimate of the marginal benefit of adding links to the monitoring network. The marginal benefit is plotted in Figure 7. If adding links was free, as it was in the experiment, then the

---

<sup>8</sup>We also made sure that the mediation of the network affects on contributions by adding punishment is not due to the fact that the data from all the participants who cannot be punished (e.g., player A in the Prison Guard) are dropped in column 2. In another regression we replaced the missing punishment values with zeros and see that the resulting point estimates are almost identical.

marginal cost of a link is zero. The shape is interesting and informative. We see that the marginal benefit actually becomes negative between five and nine links. In other words, it makes sense to add links 1 through 4 or links 9 through 12 but adding links to networks that have between 4 and 8 already seems to actually reduce average payoffs. The precision of this estimate nicely summarizes the implications of our results. In three-link networks like the Prison Guard [5] or the Line [6], the problem is straight-forward: there is not enough monitoring, more links should be added. At the same time, networks like the Undirected Circle [3] with eight links actually have the opposite problem. Here there is more punishment than what can be supported by the level of contributions. Ironically, one way to reduce the amount of punishment is to add more links because this will lower the responsibility of any given monitor and this ambiguity seems to induce some monitors to reduce punishment to a degree that ends up being closer to optimal.

## 5 Conclusions

Our experiment demonstrates that network architecture has a significant impact on the behavior of agents in public goods games and, therefore, on the outcomes achieved in different networks. While the major determinant of contribution levels appears to be whether or not all the agents are connected, not all connected networks are equally efficient since they elicit different punishment behavior. The Complete network, for example, yields high contribution levels and high efficiencies both because and despite the fact that this architecture elicits less punishment than other networks like the Directed Circle where punishment responsibilities are not shared. In short, while the previous literature was correct in pointing out that punishment may increase contributions, it failed to investigate the subtle relationship between network structure and performance. This paper has taken a first step in that direction.

## 6 Tables and Figures

Table 1: Experimental Design		
Network	Players	Obs./Subjects
1	A,B,C,D	240/16
2	A,B,C,D	240/16
3	A,B,C,D	300/20
4	B,C,D	405/27
4	A	135/9
5	B,C,D	765/51
5	A	255/17
6	D	240/16
6	A	240/16
6	B,C	480/32
7	D	60/4
7	A,B,C	180/12
8	D	105/7
8	A,B,C	315/21
	totals	3960/264

TABLE 2: Comparing Individual Behavior and Outcomes across Networks

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Contribution	Contribution	Contribution	Punishment	Payoff
[2] Directed Circle	3.974 (3.396)	2.374 (2.072)	2.102 (2.095)	4.471*** (1.574)	-1.621 (1.382)
[3] Undirected Circle	0.301 (3.195)	-0.168 (1.946)	-0.268 (1.967)	2.811* (1.557)	-3.005** (1.311)
[4] Undirected Star	-1.673 (2.868)	-0.908 (1.748)	-1.108 (1.767)	3.069** (1.443)	-0.964 (1.174)
[5] Prison Guard	-7.117*** (2.656)	-3.411** (1.665)	-3.716** (1.687)	2.607* (1.494)	-3.307*** (1.086)
[6] Line	-10.832*** (2.676)	-5.926*** (1.701)	-6.237*** (1.721)	3.984*** (1.425)	-4.049*** (1.093)
[7] Disconnected Directed Circle	-4.857 (3.383)	-2.383 (2.224)	-2.724 (2.240)	10.058*** (1.977)	-3.513*** (1.382)
[8] Disconnected Undirected Circle	-3.022 (2.987)	-1.385 (1.926)	-1.740 (1.943)	3.187** (1.572)	-1.264 (1.225)
Lagged Punishment Received		0.214*** (0.03)	-0.003 (0.25)		
Lagged Contribution		0.530*** (0.03)	0.530*** (0.03)		
Intercept	17.016*** (2.423)	3.408** (1.546)	3.702** (1.570)	1.283 (1.254)	32.937*** (1.047)
Includes time period fixed effects	Yes	Yes	Yes	Yes	Yes
Includes Punishment×Network interactions	No	No	Yes	No	No
rho	0.68	0.49	0.49	0.41	0.24
Prob > Chi <sup>2</sup>	<0.01	<0.01	<0.01	<0.01	<0.01
Obs.	3960	3080	3080	886	3960

Note: Random effect tobit with censoring at zero and full contributions or GLS regressions . \*, \*\*, \*\*\* indicates significance at the 10, 5 and 1 percent levels.

---

TABLE 3: Nodes and Network Structures

---

Links	Nodes (Players)
0 out   0 in	$N_{7,0,0} (D), N_{8,0,0} (D)$
0 out   1 in	$N_{5,0,1} (B, C, D), N_{6,0,1} (D)$
1 out   0 in	$N_{6,1,0} (A)$
1 out   1 in	$N_{2,1,1} (A, B, C, D), N_{4,1,1} (B, C, D), N_{6,1,1} (B, C), N_{7,1,1} (A, B, C)$
2 out   2 in	$N_{3,2,2} (A, B, C, D), N_{8,2,2} (A, B, C)$
3 out   0 in	$N_{5,3,0} (A)$
3 out   3 in	$N_{1,3,3} (A, B, C, D), N_{4,3,3} (A, B, C, D)$

---

TABLE 4: Comparing Individual Behavior and Outcomes across Nodes

	(1)	(2)	(3)
Dependent Variable	Contribution	Punishment	Payoff
$N_{5,3,0}$	3.883 (2.801)	0.483 (1.685)	-0.668 (1.241)
$N_{1,3,3}$	13.399*** (3.365)	-2.124 (1.815)	1.988 (1.367)
$N_{4,3,3}$	13.514*** (3.941)	1.295 (1.939)	-2.021 (1.611)
$N_{3,2,2}$	13.701*** (3.184)	0.686 (1.743)	-1.017 (1.297)
$N_{8,2,2}$	10.796*** (3.150)	1.063 (1.754)	0.055 (1.283)
$N_{2,1,1}$	17.360*** (3.384)	2.345 (1.757)	0.367 (1.367)
$N_{4,1,1}$	11.131*** (3.002)	0.695 (1.790)	2.043* (1.220)
$N_{6,1,1}$	3.749 (2.921)	2.911* (1.742)	-3.004** (1.184)
$N_{7,1,1}$	9.687*** (3.630)	7.942*** (2.130)	-2.438* (1.477)
$N_{5,0,1}$	7.070*** (2.668)		-1.536 (1.094)
$N_{6,0,1}$	2.803 (3.393)		-2.237* (1.367)
$N_{7,0,0}$	5.002 (5.377)		1.276 (2.162)
$N_{8,0,0}$	9.126** (4.288)		2.728 (1.752)
Intercept	3.611 (2.422)	3.382** (1.476)	30.947*** (1.038)
Includes time period fixed effects	Yes	Yes	Yes
rho	0.68	0.41	0.23
Prob > Chi <sup>2</sup>	<0.01	<0.01	<0.01
Obs.	3960	886	3960

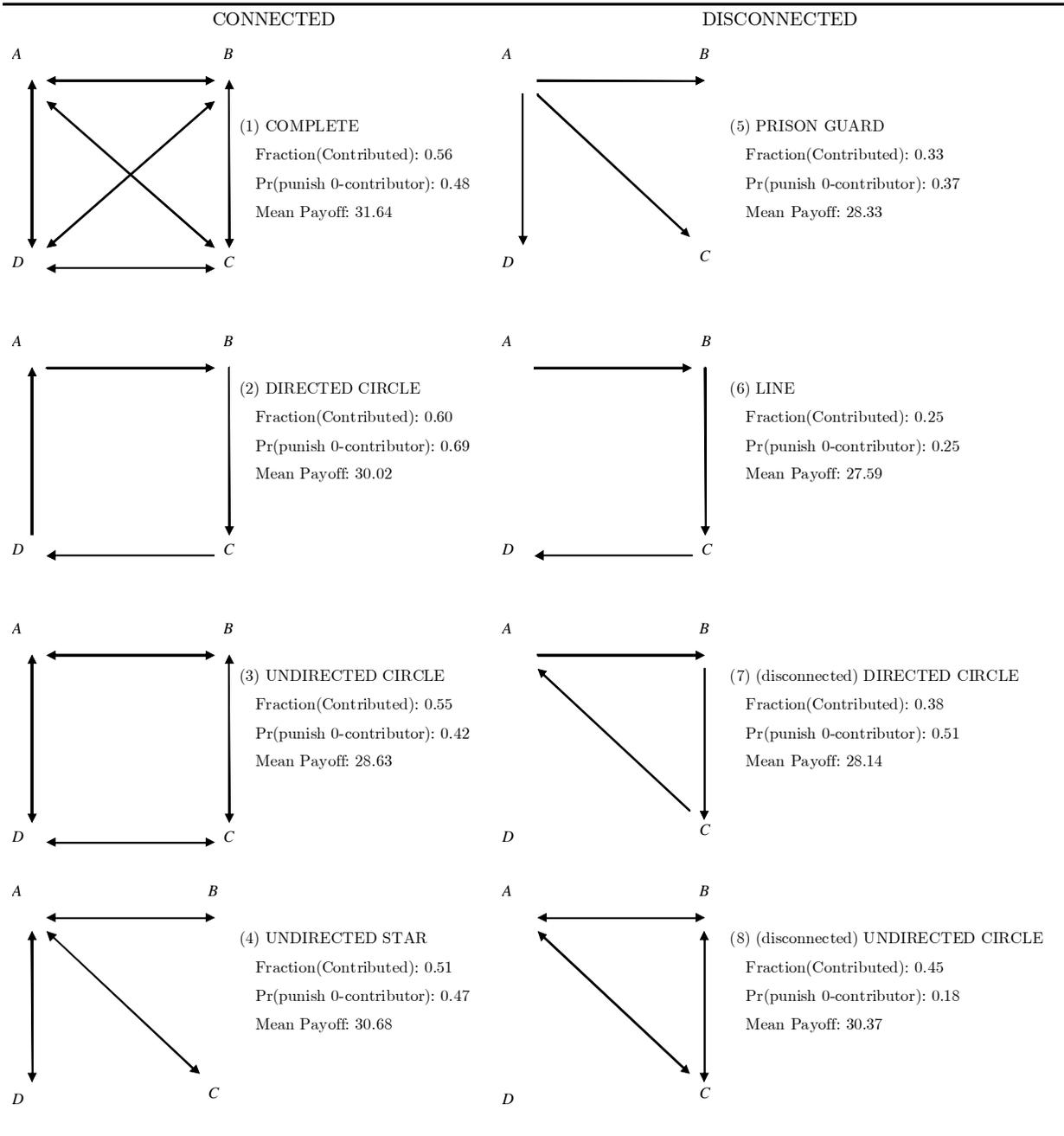
Note:  $N_{6,1,0}$  is the baseline. Random effect tobit regressions with censoring at zero and full contributions or GLS. \*, \*\*, \*\*\* indicates significance at the 10, 5 and 1 percent levels.

TABLE 5: The Effect of Network Properties

	Contributions	Punishment	Efficiency
Complete	-8.042 (7.626)	-2.811* (1.487)	6.050** (3.099)
Connected	5.336*** (2.055)	0.565 (0.427)	1.398* (0.832)
Directed	2.267 (4.267)	2.143*** (0.848)	-4.145** (1.733)
Degree (total)	1.435 (1.357)	0.379 (0.266)	-0.899* (0.552)
rho	0.698	0.288	0.247
Prob > Chi <sup>2</sup>	<0.01	0.10	<0.01
Obs.	3960	4647	3960

Note: Each regression includes time period fixed effects and individual random effects. Two-limit Tobits are used for contributions and GLS is used for punishment and efficiency. Significance levels are indicated by \*\*\* (1%), \*\* (5%) and \* (10%).

FIGURE 1: Network Architecture and Summary Statistics



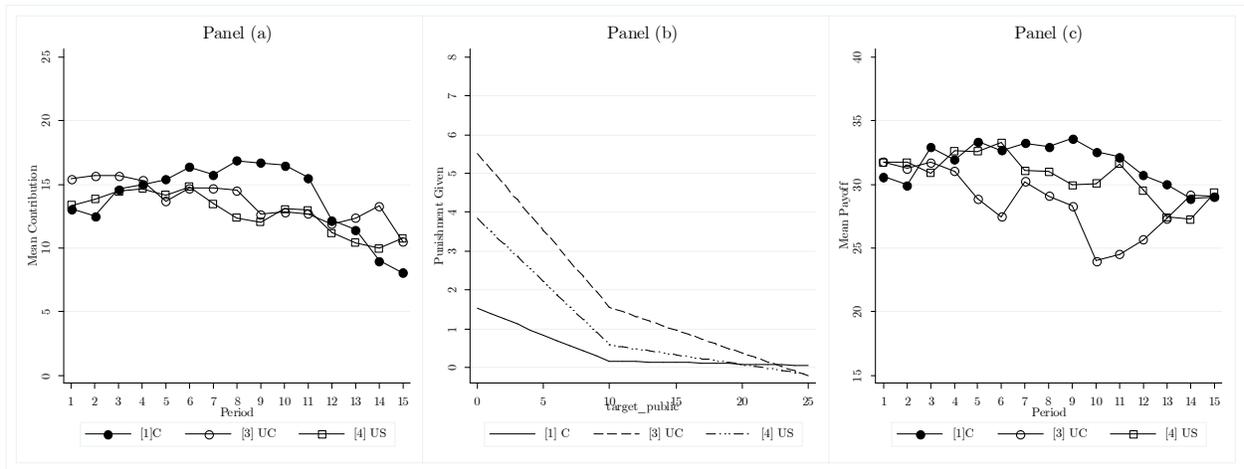


Figure 2. Comparing complete and incomplete networks.

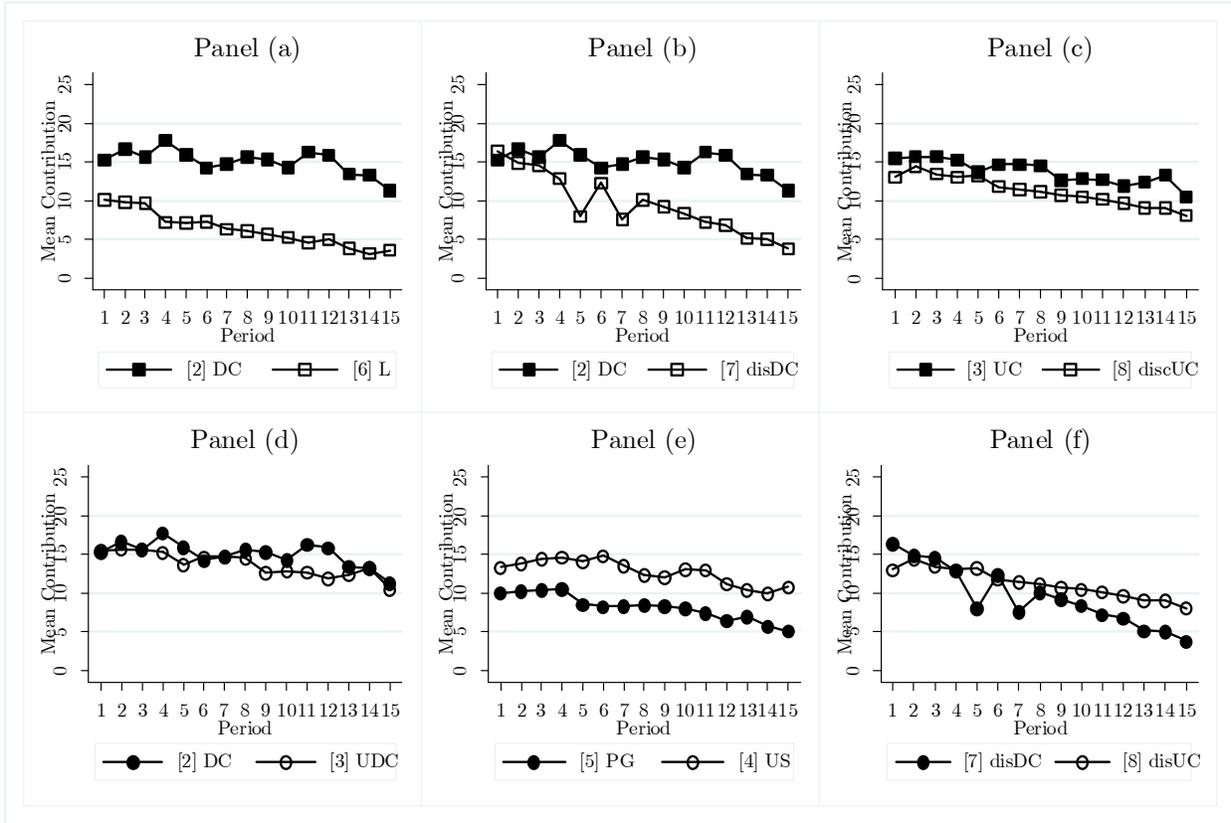


Figure 3. Comparing contributions: Panel (a)-(c) connected versus unconnected networks. Panels (d)-(f) directed versus undirected networks.

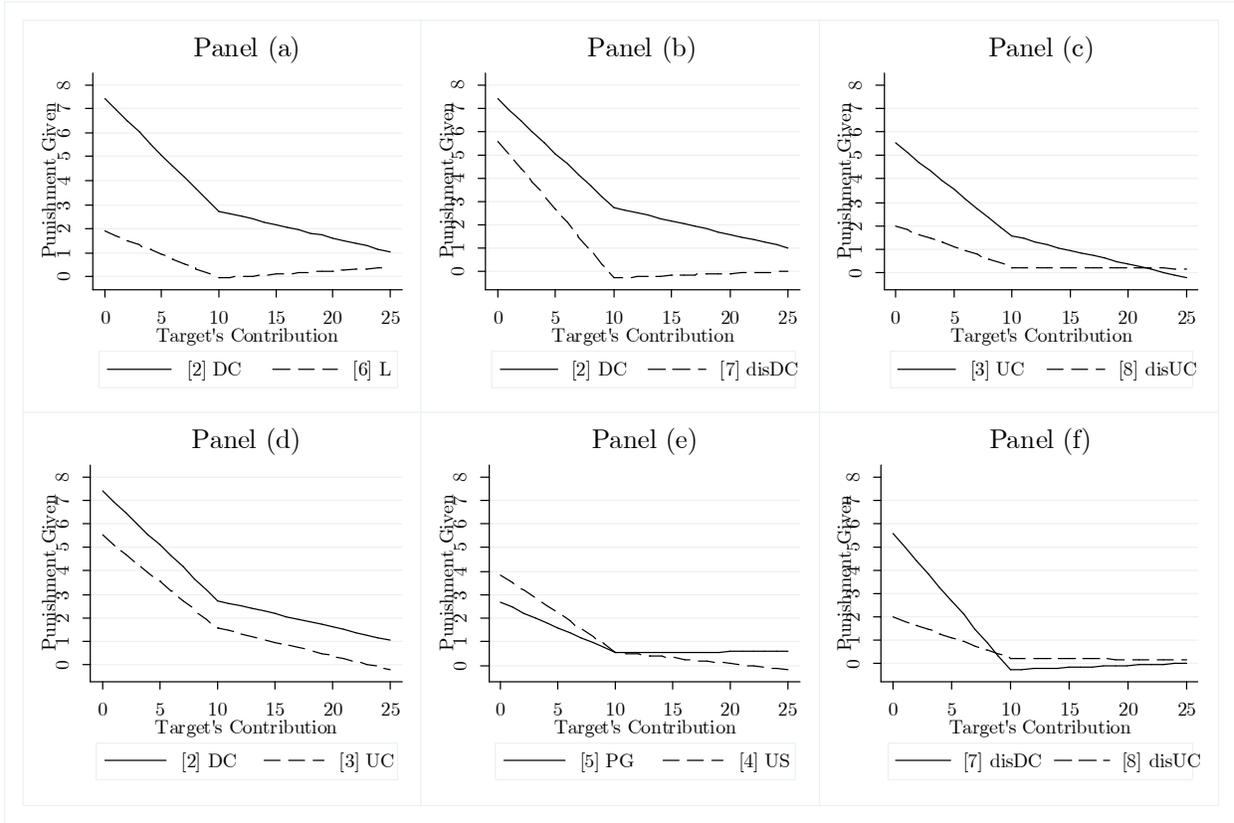


Figure 4. Comparing punishment given (estimated punishment splines): Panel (a)-(c) connected versus unconnected networks. Panels (d)-(f) directed versus indirected networks.

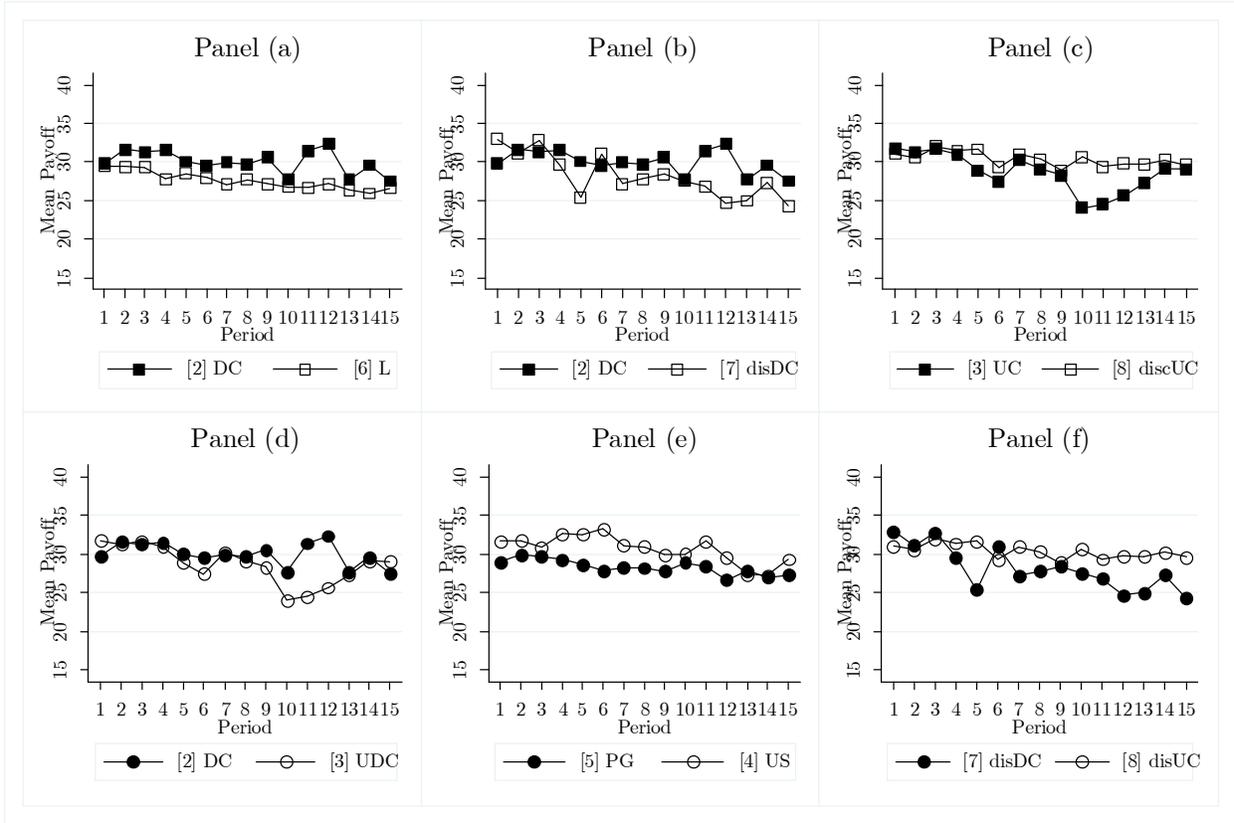


Figure 5. Comparing efficiencies: Panel (a)-(c) connected versus unconnected networks. Panels (d)-(f) directed versus indirected networks.

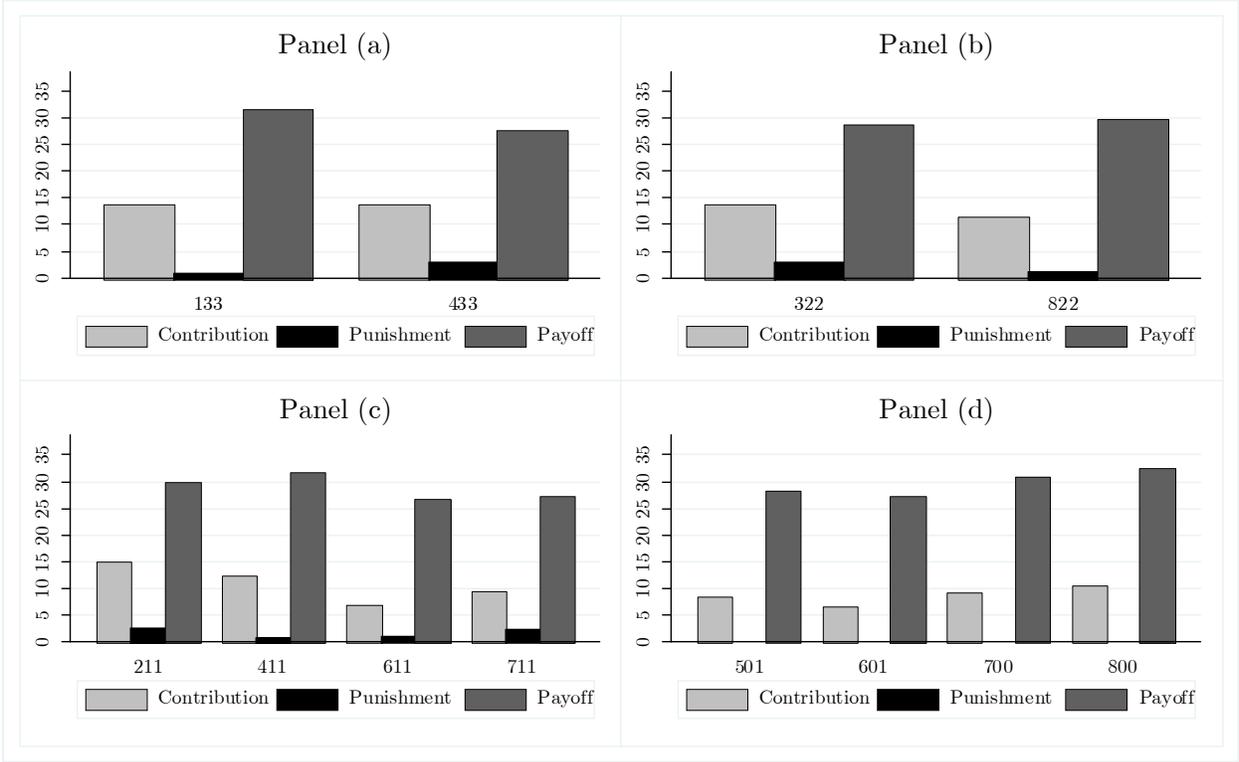


Figure 6. Comparing nodes: Panel (a) nodes with 3 out-links and 3 in-links. Panel (b) nodes with 2 out- and in-links. Panel (c) nodes with 1 out- and in-link and panel (d) those nodes with no out-links.

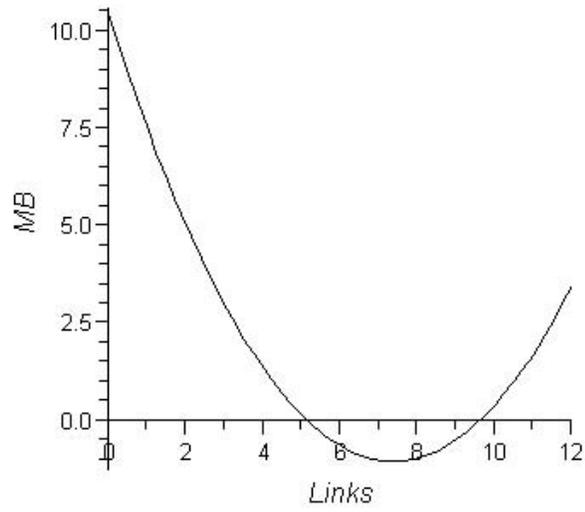


Figure 7. The estimated marginal benefit of adding punishment links.

## 7 Appendix - Instructions for the Complete Network

### *Introduction*

This is an experiment in the economics of decision-making in networks. A research foundation has provided funds for this research. Your earnings will depend partly on your decisions and partly on the decisions of the other participants in the experiment. If you follow the instructions and make decisions carefully, you may earn a considerable amount of money.

At this point, take a minute to write down the number of the computer you are using (it appears on the top of the monitor). At the end of the experiment, you should use your computer number to claim your earnings. At this time, you will receive \$5 as a participation fee (simply for showing up on time).

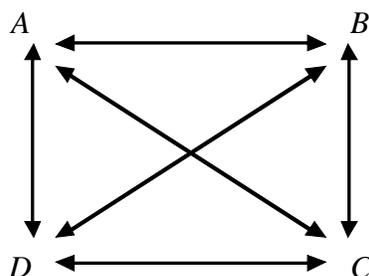
During the experiment your payoff will be denominated in experimental tokens that will be converted into dollars at the end of the experiment at the following rate: 25 Tokens = 1 Dollar.

### *The Experiment*

In this experiment, you will participate in 15 independent and identical (of the same form) periods each divided into two decision-stages. In each period, you will be assigned to a position in a four-person network. You will only be able to observe the choices of the other participants to whom you are connected in this network.

Before the first period, you will be randomly assigned to one of the network positions labeled A, B, C, or D. A fourth of the participants in the room will be designated as type-A participants, one fourth as type-B participants, one fourth as type-C participants and one fourth as type-D participants.

Your type depends solely upon chance and will remain constant throughout the experiment. When you are asked to make your first decision you will be informed of your type on the left hand side of the dialog window (see attachment 1). The network is given to you in the window that appears to the right of it and is illustrated in the scheme below.



A line segment between any two types indicates that they are connected and the arrowhead points to the participant whose action can be observed. A two-way arrowhead indicates that both connected participants can observe the actions of the other. For example, the two-way arrowhead between subjects A and D indicates that each can observe the actions of the other. In the network used in this experiment each type can observe the choices of all the other types.

Each period of the experiment consists of two stages. In the first stage, you will decide how many tokens you want to allocate in each of two accounts. One account is a private account which only you benefit from, and the other is a public account which benefits all four participants in your group equally.

In the second stage of the period, you will observe the choices at the first stage of the participants to whom you are connected by the network. You can then decide whether and by how much to reduce their earnings from the first stage.

Next, we will describe in detail the process that will be repeated in all 15 periods.

### *Stage One*

Each period starts by having the computer randomly form four-person networks by selecting one participant of type-A, one of type-B, one of type-C and one of type-D. The networks formed in each period depend solely upon chance and are independent of the networks formed in any of the other periods.

That is, in any network each participant of type-A is equally likely to be chosen for that network, and similarly with participants of type-B, type-C and type-D. This is done by the computer. Hence, if there are 20 participants in the experiment the computer will randomly form five groups of four participants in each period.

At the beginning of each period each participant receives an endowment of 25 tokens. Your task is to decide how to use your endowment. You have to decide how many of the 25 tokens you want to allocate to the public account and how many to your private account. Your earning in the first stage will be determined by how many tokens you allocate to the private and public accounts and total number of tokens the other participants in your group allocate to the public account.

To make your allocation decision use the mouse to click on either the Private or Public input fields on the lower left hand corner of the dialog window (see attachment 1) and use the keyboard to fill the number (no decimals) of tokens between 0 and 25 that you wish to allocate to that account. Note that any of the 25 tokens not assigned to one account will automatically be allocated to the other. After that, confirm your decision by clicking on the Submit button. Once you have done this, your decision cannot be revised. This completes the first stage in the period.

Your earnings at the first stage are composed to two parts:

(1) Your earnings from your private account simply equal the tokens in your private account.

(2) Your earnings from the public account are the total tokens allocated by all the participants in your group (including yourself) to the public account times 0.4.

Your earnings are given to you on the right hand side of the dialog window (see attachment 1).

Your total earnings in the first stage can therefore be summarized by (tokens in private account) $+0.4\times$ (group total tokens in public account).

The earnings of each participant in your group are calculated in the same way. Therefore, each participant receives the same amount from the public account.

For example, suppose that the total tokens allocated by all the participants in your group to the public account is 60. Then, each participant in the group receives  $0.4\times 60=24$  tokens from the public account. If you allocated 10 tokens to the public account then your final payoff would be  $(25-10)+(0.4\times 60) = 15+24$  or 39 tokens.

Similarly, if the total tokens allocated by all the participants in your group to the public account is 20 then each participant in the group receives  $0.4\times 20=8$  tokens from the public account. If you allocated 10 tokens to the public account your final payoff would be  $(25-10)+(0.4\times 20) = 15+8 = 23$  tokens.

For each token you allocate to your private account you earn one token. If instead you allocate this token to the public account the group total would rise by one token. The earnings of all participants in your group, including yourself, from the public account would rise by  $0.4\times 1=0.4$  tokens each. Thus, the total earnings of the participants in your group would rise by 1.6 tokens. Also remember that you earn 0.4 tokens for each token allocated to the public account by any of the other participants in your group.

### *Stage Two*

When everyone in your group has made a decision in the first stage, you will observe the choices of the participants to whom you are connected by the network. For example, if you are a type-A participant, you will be informed what allocation the type-B, type-C, and the type-D participants have chosen.

You will observe how many tokens each participant with whom you are connected by the network allocated to both his or her private and public accounts. This information is given to you in the large window that appears at the top of the dialog window (see attachment 2).

Each of the other participants will also observe the choices of the participants to whom he or she is connected by the network. In addition, all participants will be informed about the total number of tokens allocated by all the participants in your group (including yourself)

to the public account. This information is given to you right below the large window that appears at the top of the dialog window (see attachment 2).

At this stage, you have to decide if and by how much you wish to reduce the earnings from the first period of each of the other participants with whom you are connected by the network. At the same time each of the other participants will decide if and by how much he or she wishes to reduce the earnings from the first period of each of the other participants with whom he or she is connected by the network.

When you are ready to make your decision, use the mouse to choose the type of participant in the stage two window (see attachment 2) and use the keyboard to fill the number of tokens (no decimals) you wish to reduce from that participant's earnings into the input-text box that appears to the right of the participant's type. After that, confirm your decision by clicking on the Set button.

You must decide by how much you wish to reduce the earnings of each of the other participants with whom you are connected by the network by filling a number of tokens for each of them. If you do not wish to reduce the earnings of another participant you must enter 0.

Reducing the earnings of other participants is costly. The cost of reducing one token from any of the other participants is  $1/2$  token. Each participant can spend up to his or her first stage earnings on reducing the earnings of the other participants with whom he or she is connected by the network. To keep track of both your current period earnings and your accumulated earnings from each of the previous periods, there is a text box near the bottom of the screen (see attachment 2).

After you have entered the amount that you wish to reduce the earnings of each of the other participants to whom you are connected by the network, confirm your decisions by clicking on the Submit button (see attachment 2). Once you have done this, your decisions cannot be revised. This completes the second stage in this period.

When a period ends, the computer will inform all participants of their total earnings, earnings from private account, earnings from public account, total tokens spent on reducing others' earnings, and total tokens reduced by other participants (see attachment 3). After letting you observe the results, the next period will start by having the computer randomly forming new groups of participants in networks.

### *Earnings*

Your final earnings at the end of each period are determined as follows:

Stage one earnings (your private account balance + your share of the group account).

Minus: the reductions directed at you.

Minus: your expenditure on the reductions directed at other participants.

Because this sum can be negative, if your expenditures on reductions and the reductions directed at you are greater than your token earnings from the first stage, we will truncate your earnings at zero. Thus, your earnings from the first stage can be reduced at the second stage to zero but total earnings for the period cannot be negative.

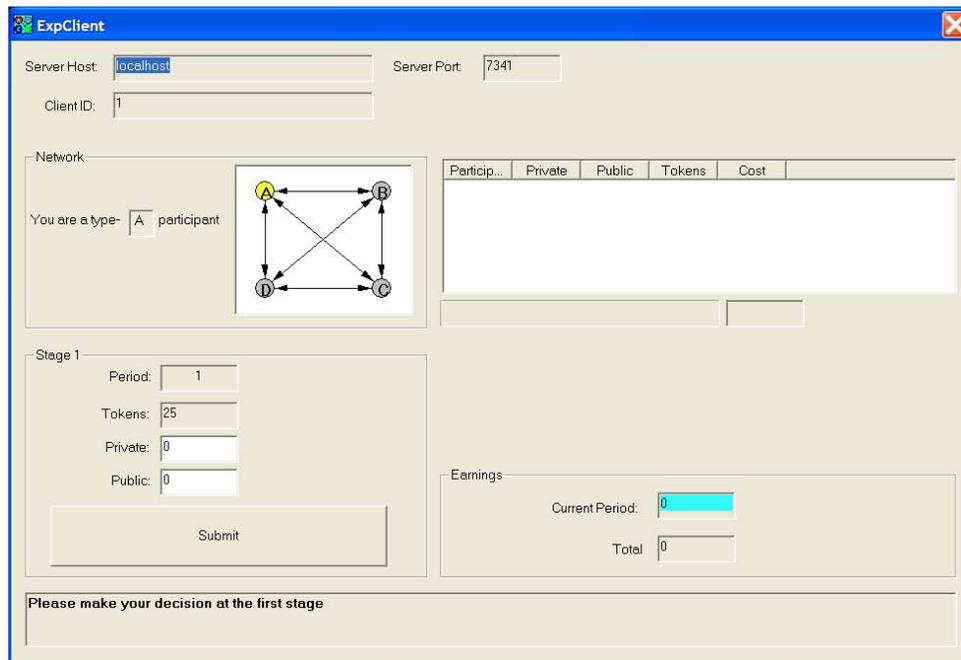
The process will be repeated until all the 15 periods are completed. At the end of the last period, you will be informed the session has ended. Your final earnings will be the sum of your earnings over the 15 periods.

### Rules

Please do not talk with anyone during the experiment. We ask everyone to remain silent until the end of the last period and then wait for further instructions.

Your participation in the experiment and any information about your earnings will be kept strictly confidential. Your payments receipt and participant form are the only places in which your name and social security number are recorded.

If there are no further questions, you are ready to start. An instructor will approach your desk and activate your program.



Attachment 1

ExpClient

Server Host: localhost Server Port: 7341

Client ID: 2

Network

You are a type- B participant

Particip...	Private	Public	Tokens	Cost
A	5	20		
B	5	20		
C	5	20		
D	5	20		

Total Public Contributions: 00

Participant  reduce by  tokens

Cost of reducing tokens:

Stage 1

Period:

Tokens:

Private:

Public:

Earnings

Current Period:

Total

Please make your decisions at the second stage

## Attachment 2

ExpClient

Period 1 ended  
 Your total earnings in this period are 37 tokens  
 Your earnings from private account are 5 tokens and your earnings from public account are 32 tokens  
 The total tokens spent on reducing others' earnings is 0. The total tokens reduced by other participants is 0

## Attachment 3

## 8 References

- Anderson, Christopher and Putterman, Louis. 2005. Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. *Games and Economic Behavior*. 54(1), 1-24.
- Carpenter, Jeffrey, 2007. Punishing Free-Riders: how group size affects mutual monitoring and the provision of public goods. *Games and Economic Behavior*, 60(1), 31-51.
- Carpenter, Jeffrey and Matthews, Peter, 2009. What norms trigger punishment? *Experimental Economics*, 12(3), 272-288.
- Cinyabuguma, Matthias, Page, Talbot and Putterman, Louis, 2005. Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics*. 89(8), 1421-1435.
- Fehr, Ernst and Gächter, Simon, 2000. Cooperation and Punishment in Public Goods Experiments. *American Economic Review*, 90(4), 980-994.
- Jackson, Matthew, 2005, A Survey of Models of Network Formation: Stability and Efficiency. In: Gabrielle Demange and Myrna Wooders (Eds.), *Group Formation in Economics: Networks, Clubs and Coalitions*. Cambridge University Press, Cambridge, pp. 11-57.
- Kosfeld, Michael, 2004. Economic Networks in the Laboratory: A Survey. *Review of Network Economics*, 3, 20-41.
- Ledyard, John, 1995. Public Goods: a survey of experimental research. In: John Kagel and Alvin Roth (Eds.), *The Handbook of Experimental Economics*. Princeton University Press, Princeton, pp. 111-194.
- Masclét, David, Noussair, Charles, Tucker, Steven and Villeval, Marie-Claire, 2003. Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism. *American Economic Review*, 93(1), 366-380.
- Nikiforakis, Nikos, 2008. Punishment and Counter-punishment in Public Goods Games: Can we still govern ourselves? *Journal of Public Economics*, 92, 91-112.
- O’Gorman, Rick, Henrich, Joseph and Van Vugt, Mark, 2008. Constraining Free Riding in Public Goods Games: Designated Solitary Punishers can Sustain Human Cooperation. *Proceedings of the Royal Society B*.