Journal of Economic Behavior & Organization 132 (2016) 26–41

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

# Asymmetric network monitoring and punishment in public goods experiments

Luke Boosey*, R. Mark Isaac

*Department of Economics, Florida State University, 288 Bellamy Building, Tallahassee, FL 32306-2180, United States*

ARTICLE INFO

*Article history:*
Received 14 March 2015
Received in revised form 1 September 2016
Accepted 19 September 2016
Available online 3 October 2016

*JEL classification:*
C72
C91
C92
H41

*Keywords:*
Networks
Public goods
Monitoring
Punishment
Asymmetry
Retaliation

ABSTRACT

We extend the recent experimental literature on incomplete punishment networks in linear public goods games. In these games, we use an exogenous network to restrict both monitoring and the set of feasible punishment flows. In addition to two baseline structures (the Complete network and the Circle network), we examine a novel Asymmetric network in which both punishment responsibility and exposure differ across players. Average contributions are significantly lower in the Asymmetric network, driven entirely by the under-monitored player who faces only one potential punisher. We formulate and examine the hypothesis that asymmetry among a player's potential punishment targets may lead to discriminatory patterns of punishment. In particular, players might wish to punish targets for whom they are solely responsible discriminately more than targets for whom they share responsibility. The experimental data do not support this hypothesis, although they do suggest a compelling explanation as to why. Specifically, we find that the under-monitored player in the network retaliates against previous punishment significantly more often than others in the group, which deters their only potential punisher from issuing stronger sanctions. Thus, an additional complication of asymmetry in the network is that it may lead to more instances of anti-social retaliation, inhibiting the effectiveness of the decentralized punishment institution.



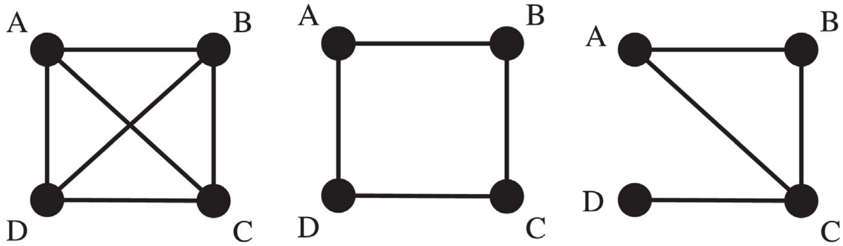
## 1. Introduction

In recent years, many studies have shown that cooperation in social dilemma environments can be enhanced by a range of decentralized institutions. A seminal contribution by Ostrom et al. (1992) demonstrated that in a common pool resource experiment, cooperative players were willing to use costly sanctions to deter over-extraction of the resource. Similarly, Fehr and Gächter (2000) showed that costly punishment can be used to sustain higher contributions in voluntary contribution public goods games, in stark contrast to the tendency for contributions to decay over time when punishment is not possible.[1] Several subsequent studies have reproduced this result, while showing that the impact of punishment on cooperation may depend on several different features of the institution. For example, the use of punishment is sensitive to the relative cost (or price) of punishment (Anderson and Putterman, 2006; Carpenter, 2007b; Egas and Riedl, 2008; Nikiforakis and Normann, 2007), the formal or informal nature of the sanctions (Masclet et al., 2003; Noussair and Tucker, 2005), the interaction of

* Corresponding author.
*E-mail addresses:* lboosey@fsu.edu (L. Boosey), misaac@fsu.edu (R. Mark Isaac).

[1] Ledyard (1995) and Chaudhuri (2010) provide two excellent surveys of the experimental public goods literature.

http://dx.doi.org/10.1016/j.jebo.2016.09.015




(a) Complete Network (b) Circle Network (c) Asymmetric Network

**Fig. 1.** The networks.

punishment with different forms of communication (Bochet et al., 2006), the threat of expulsion or exclusion (Cinyabuguma et al., 2005), the availability of rewards (Sefton et al., 2007), underlying social norms (Carpenter and Matthews, 2009), and the type of feedback provided to subjects (Nikiforakis, 2010).

Recent work by Carpenter et al. (2012) and Leibbrandt et al. (2015) has extended the analysis to environments where punishment opportunities are restricted by an underlying network. Carpenter et al. (2012) show that cooperation tends to be greater in connected networks than disconnected networks, while punishment tends to be used more heavily in directed networks than in undirected networks. Similarly, Leibbrandt et al. (2015) find that the structure of the network significantly affects contributions and punishment decisions, but not overall efficiency levels. In this paper, we extend the literature on network punishment by showing how an asymmetric distribution of punishment responsibilities may undermine the impact of punishment on cooperation. In particular, a player who is under-monitored (relative to the other group members) may seek to exploit his position by using anti-social punishment to retaliate and deter future sanctions.

We report the results from a laboratory experiment where groups of 4 players participated in a two-stage linear public goods game with punishment, under three different network structures. To serve as two separate baselines, we examined the Complete network and the (undirected) Circle network. However, the main treatment of interest is an Asymmetric network, in which both the monitoring and punishment opportunities depend on the player's position in the network. The three networks are shown in Fig. 1.

Our approach differs from the two previous studies in a couple of ways. First, we introduce a novel network structure that allows us to explore a more subtle degree of asymmetry in the distribution of punishment responsibility across players. Second, we restrict attention to undirected networks, which leaves open the possibility of retaliation in repeated play. Furthermore, in contrast with Carpenter et al. (2012), but consistent with Leibbrandt et al. (2015), we use a partners matching protocol, keeping both groups and players' locations fixed across all periods. This aspect of the design is important for two reasons. First, it gives us a better chance to observe retaliation against punishment in subsequent periods. Second, as suggested by Leibbrandt et al. (2015), we feel that the partners design allows greater insight into the network as a social structure, rather than as an architecture.

The overall impact of restricting monitoring and punishment opportunities in a group is not immediately apparent, since there are competing effects on the level of deterrence to free-riding behavior. On the one hand, some individuals in the group may face fewer potential punishers, reducing the incentives for them to make higher contributions. On the other hand, some punishers will have fewer targets to monitor, which could increase their propensity to punish the targets they do monitor. This effect may be particularly important in groups where coordination failure or second-order free-riding among punishers is a problem.

The potential importance of asymmetry in the network can be illustrated by comparing our Asymmetric network with two of the network structures studied in Carpenter et al. (2012) and Leibbrandt et al. (2015). The first is the *Untouchable* network in Leibbrandt et al. (2015) (referred to as the *Disconnected Undirected Circle* by Carpenter et al., 2012) where one player is completely unmonitored and disconnected from the other three players (who form a completely connected clique). Neither Leibbrandt et al. (2015) nor Carpenter et al. (2012) find any significant difference between average contributions in the Complete and Untouchable networks. Our Asymmetric network can be formed from this *Untouchable* network by adding a link between the untouchable player and one of the other three players. The addition of this link is a subtle change, yet it has an important effect on the structure of punishment responsibility. First, the previously untouchable player is still relatively under-monitored, and his only potential punisher is also (nominally) responsible for monitoring all other players. This suggests that the threat of punishment is still much weaker for the under-monitored player (player D in our network), than for the other three players. Second, the player who can now punish everyone (player C in our network) is solely responsible for one target (player D) but shares the responsibility of monitoring the other two targets (players A and B). As such, player C may concentrate on sanctioning player D more severely than he punishes A or B for comparable transgressions. Thus, there are potentially competing effects on the contributions made by player D in our Asymmetric network, compared with the other players. Neither of these effects are present in the *Untouchable* network studied by the existing literature.

A similar argument can be made with respect to the *Undirected Star* network studied by Carpenter et al. (2012).[2] In this case, the Asymmetric network is formed by adding a link between two of the peripheral players (players A and B). From player C's perspective, this additional link creates the same shift in punishment responsibility discussed above – he is solely responsible for punishing player D, but shares responsibility for punishing the other two potential targets. Thus, whereas before there would be no reason for the intensity of punishment to differ across targets (controlling for the same behavior), player C may be tempted to choose to concentrate his limited punishment resources on player D, for whom he is solely responsible. Although we concentrate in this paper on the Asymmetric network shown in Fig. 1, there are other possible networks with different kinds of asymmetry that would also be interesting to examine. For the *Undirected Star* network, we rely on the findings reported by Carpenter et al. (2012), which suggest that there are no significant differences in average contributions compared with the *Complete* network. However, there is considerably less scope for retaliation in their experiment, since players are not matched with the same partners in subsequent rounds. In our environment, the central player in the *Star* network may be subjected to retaliation or pre-emptive anti-social punishment by all three of the peripheral players in the network, making effective punishment much more difficult to implement.

Based on the above arguments, we develop and test the hypothesis that player C issues discriminatory punishment, by punishing player D more frequently or more severely than players A and B, for comparable contribution decisions. The experimental data do not support this hypothesis, however, they do provide a compelling explanation as to why the differences do not emerge. Specifically, player D exploits her position as an under-monitored player and in many cases *protects* that position by retaliating against past punishment received from player C. Thus, while player C may wish to discriminately punish player D more than the others, he is deterred from doing so by the threat of retaliation. The result is that cooperation is more difficult to sustain in the Asymmetric network. Average contributions are significantly lower than in the Complete and Circle networks, and the difference is driven almost entirely by the under-monitored player D.

In broader terms, our paper adds to the growing literature that studies strategic interaction over networks. The role of networks in fostering cooperation and trust has been carefully examined in the context of favor exchange, informal borrowing, and insurance networks (see, e.g. Jackson et al., 2012; Karlan et al., 2009; Bloch et al., 2008; Bramoullé and Kranton, 2007b). In general, this literature has emphasized the way in which the network serves as a form of social collateral that facilitates economic transactions that might not otherwise occur.

Also related to our study is the work by Bramoullé and Kranton (2007a) on public goods provision in networks. They examine the network environment in which benefits of public good provision are shared across network links, and show that there are, in general, many equilibria of the game. However, the only stable equilibrium profiles are those characterized by a form of specialization, where the sets of contributors correspond to maximal independent sets for the network. There is also a steady stream of experimental research on games played over networks.[3] Rosenkranz and Weitzel (2012) provide an experimental test of the model developed by Bramoullé and Kranton (2007a). They find some evidence to support the emergence of the stable specialized equilibria, but predominantly in the complete and star network structures, with little equilibrium convergence in other four-person network structures with an intermediate number of links.

Our focus in this paper is also on the provision of public goods, although our attention is directed towards the punishment institution and the role of network connections in facilitating the costly enforcement of cooperation. In particular, the network helps to establish a distribution of punishment power, which leads to heterogeneity in both the players' exposure to, and responsibility for punishment. Thus, our investigation is also closely related to several other studies examining the effectiveness of decentralized punishment. One early contribution by Carpenter (2007a) shows that punishment is far less effective when players can only monitor and punish one other person than when they can punish all others or even half of the others in the group. Our result concerning player D's contributions in the Asymmetric network reaffirms the notion that under-monitoring is less effective at increasing contributions.

Another earlier study by O'Gorman et al. (2009) explores the case where the responsibility for punishment is delegated to a single monitor in the group.[4] This may be considered a particular kind of asymmetry, although all of the non-punishers are perfectly symmetric. Under this centralized punishment institution, they find that contributions are just as high as when everyone can punish everyone, while the overall efficiency is higher. Thus, they conclude that groups may benefit from having a single punisher whose behavior "can efficiently galvanize group cooperation" (O'Gorman et al., 2009, p. 323). In this case, players who face a single potential punisher are responsive to the threat of sanctions, while the solitary punisher is able to solve coordination and second-order free-riding problems that may occur with decentralized punishment.

An obvious difference between our study and the approach taken by O'Gorman et al. (2009) is that we allow for punishment to flow in both directions. This aspect of our design enables under-monitored players, such as player D in the Asymmetric network, to also punish their solitary potential punishers. Furthermore, since players interact repeatedly in fixed groups, there are opportunities for all players to implement retaliatory punishment. Our results show that this strategy is adopted significantly more often by the under-monitored player D in our Asymmetric network. A compelling explanation

[2] As for the Disconnected Undirected Circle, they do not find a significant difference between average contributions in the Undirected Star and the Complete network.

[3] For a comprehensive survey of experimental work involving networks, see Choi et al. (2016).

[4] Such an environment is isomorphic to a directed star network with the punisher at the center and the other players on the periphery.

for this is that D needs only to deter one player (C) from punishing her in order to escape the consequences of contributing less, while others must deter two or three potential punishers.

Other than Carpenter et al. (2012) and Leibbrandt et al. (2015), the closest related study to ours is by Nikiforakis et al. (2010). They also examine asymmetric punishment institutions, although the asymmetry across players manifests in a much different way. Rather than considering asymmetry in the set of feasible punishment flows, they allow for different players in the group to differ in the effectiveness of their punishment. Despite facing a reduced threat of punishment (since the authors control for average punishment effectiveness), players with high punishment effectiveness, who they refer to as strong players, do not exploit the privilege of their position by contributing less than others. Our approach considers a different kind of asymmetry; specifically, different players face different levels of responsibility for and exposure to punishment. When punishment is feasible, the effectiveness of punishment is symmetric across all players. Contrasting with Nikiforakis et al. (2010), in our experiments, we find that the players who are in the 'privileged' position (of being under-monitored) are able to exploit their advantage by using the threat of retaliation to deter future punishment.

Finally, given the opportunity for players in our repeated game to implement targeted revenge, our results also contribute to the literature on counter-punishment in social dilemmas. Nikiforakis (2008) shows that when explicit opportunities for counter-punishment are provided, cooperation breaks down and efficiency levels are lower than in the standard game without punishment. Indeed, he presents evidence that counter-punishment is sometimes assigned as retaliation, and sometimes as a strategic attempt by players to deter their neighbors from punishing them in the first place. Other studies by Hermann et al. (2008), Denant-Boemont et al. (2007), and Cinyabuguma et al. (2006) have found a similar effect on cooperation and efficiency when counter-punishment opportunities are available. Although we do not incorporate an explicit counter-punishment stage, we find significant evidence of anti-social punishment by players in the periods after they are punished, especially by player D in the Asymmetric network. This suggests that asymmetry in the punishment network may also facilitate more frequent use of anti-social punishment by under-monitored players, thereby reducing the effectiveness of the punishment institution.

The rest of the paper is organized as follows. In Section 2, we summarize the experimental design and procedures, while Section 3 introduces our main hypotheses. In Section 4, we present the results, beginning with an overview of the data, and proceeding with both aggregate and individual-level analysis of contributions and punishment decisions. Section 5 discusses the implications of our results and concludes with some suggestions for future research.

## 2. The experiment

### 2.1. The game

In the experiment, all subjects participated in seven independent matches. Before each match, they were randomly assigned into groups of 4, then played 10 periods of a two-stage game. The first stage consisted of the standard linear VCM game. Each player received an endowment of 10 tokens, which they could invest in a group exchange (equivalent to a public good contribution) and an individual exchange (equivalent to private consumption). Each token invested by the player in the individual exchange earned her 1 token, while each token invested by the group in the group exchange yielded 0.4 tokens to every player in the group. This setup creates the classical social dilemma in which individuals have a dominant strategy to invest nothing in the group exchange, while the joint group earnings would be maximized if everyone invested their full endowment in the group exchange. Using $c_i$ to denote player *i*'s contribution (investment in the group exchange), the Stage 1 payoff to player *i* was given by

$$\pi_i^1 = 10 - c_i + 0.4\sum_{j=1}^{4} c_j. \tag{1}$$

After the first stage, each subject was shown the contribution decisions made by each of his immediate neighbors in the network, which was exogenously determined and fixed throughout all 10 periods of a given match.[5]

In the second stage, the players were given the option to assign punishment to any or all of their immediate neighbors.[6] Each player received an additional endowment of 6 tokens. Using these tokens, he could buy and assign deduction points to any of his neighbors. Each deduction point cost the player (the punisher) 1 token and reduced the target player's earnings by 3 tokens. Players were allowed to assign a maximum of 2 deduction points to each of their neighbors, including partial

[5] Subjects also observed the total contributions made by the group and were shown the calculation of their Stage 1 payoff before beginning Stage 2. This means that all players in the Circle network, and players A and B in the Asymmetric network, could indirectly deduce the individual contribution made by the group member with whom they were not linked. However, they were only *explicitly* shown the individual contributions made by their direct neighbors.

[6] In the experiment, the term "punishment" was never used. Rather, the players were given the option to assign "deduction points". In this way, we kept the language as neutral as possible, without confusing the interpretation of the deduction points as a way to sanction other players.

deduction points (up to two decimal places). Thus, using $d_{ij}$ to denote the number of deduction points assigned by player $i$ to player $j$ and $N_i$ to denote the set of player $i$'s neighbors, player $i$'s payoff from Stage 2 was given by

$$\pi_i^2 = 6 - \sum_{j \in N_i} (d_{ij} + 3d_{ji}). \tag{2}$$

After the second stage, the players observed the number of deduction points they assigned and the number of deduction points they received from each of their neighbors in the network. However, they were not shown any information about punishment flows between other players or about the final payoffs received by others. Player $i$'s payoff from the entire period was the sum of his payoffs from the two stages, $\Pi_i = \pi_i^1 + \pi_i^2$. Subjects were paid the sum of their payoffs from all 10 periods in a match,[7] and were paid for all seven matches.

### 2.2. Design and procedures

We conducted 8 sessions with a total of 128 subjects who were recruited using ORSEE (Greiner, 2015) from a pool of undergraduate students at Florida State University (FSU). Subjects were screened at the announcement stage to allow only students who had not previously participated in a similar public goods or network experiment. The experiment was programmed using z-Tree (Fischbacher, 2007) and run in the XS/FS laboratory at FSU. Each session lasted for 2 hours and subjects earned an average payment of US$26.78 (including a $10 showup fee).

At the beginning of the session, subjects were randomly seated at computer terminals and given a set of written instructions describing the experiment.[8] Instructions were read aloud by the experimenter, then the subjects participated in an unpaid practice period to familiarize themselves with the computer software. All 16 subjects in a session participated in 7 independent matches. The treatment variable in the experiment is the underlying monitoring and punishment network. The 3 treatments we used are called Complete, Circle, and Asymmetric, corresponding to the networks shown in Fig. 1.[9] Each match consisted of 10 periods of the repeated two-stage game described above, and between matches, the subjects were randomly regrouped into new groups of 4 players.[10] Both the group assignment and the network treatment were fixed for all 10 periods of a given match. All of the networks are undirected, so that an edge represents mutual opportunities for monitoring and punishment. Once subjects were assigned to their group, they were randomly assigned to a position in the network. The positions were labeled *A*, *B*, *C*, and *D*, as shown in Fig. 1.

## 3. Hypotheses

In this section, we present several hypotheses for aggregate and individual behavior. First, previous results suggest that average contributions will be similar in the Complete and Circle networks (Carpenter et al., 2012). In the Asymmetric network, the overall effect on average contributions is unclear. On the one hand, player D answers to only one potential punisher (C) who is already partly responsible for two other players. Thus, it seems plausible that player D may perceive the threat of punishment to be very weak and choose to contribute less. On the other hand, given the structure of punishment responsibility across the network, it is also possible that player C may choose to focus more attention on player D, that player D may anticipate this increased scrutiny and respond by contributing more. Given these competing arguments, we propose the following as a null hypothesis.

**Hypothesis 1.** Average contributions are the same in all three network structures.

Our Asymmetric network also motivates an additional hypothesis concerning the effect of a player's location within the network on his contribution. In the Complete and Circle networks, there is no reason to expect behavior to depend on the player's location, since all locations are symmetric. On the other hand, in the Asymmetric network, the level of deterrence due to the threat of punishment is systematically different across locations, as is the level of punishment responsibility attributed to a player. Thus, we investigate whether there are any differences in contribution decisions across player location in the asymmetric network.

[7] Note that it was possible for a player to receive a negative payoff in a given period. Thus, it would also be possible (although unlikely) for a player to receive a negative payoff from all 10 periods in a given match. We instituted a bankruptcy rule specifying that if the sum of a player's payoffs from all 10 periods in a match was negative, his actual earnings from the match would be set equal to zero. As it turned out, we did not need to invoke the bankruptcy rule in any instance.

[8] Sample instructions are provided in the Online Supplement.

[9] In 5 sessions, the order of the treatments went Complete, Circle, Asymmetric × 3, Circle, Complete. In the other 3 sessions, we changed the order to be Asymmetric × 2, Circle, Complete × 2, Circle, Asymmetric. We do not observe any significant order effects between these two groups of sessions, and so we do not distinguish between the orders in the rest of the paper. However, it is possible that our decision to run only two predetermined orders may lead to some learning effects or other dependence across matches. To the extent that there may be learning over time, we believe that the effects are likely to be similar across both orders used in our design. An alternative approach that might minimize the potential learning effects would be to randomize the assignment of network treatments in each session.

[10] Players were randomly regrouped according to an absolute stranger matching protocol for Matches 2 through 6, ensuring that subjects were not matched with any of the same group members twice during those matches. Moreover, even though subjects could be matched twice with a member of their group from Match 1 or Match 7, there was no way for them to identify their group members or to know if their neighbors were the same.

**Hypothesis 2.** In the symmetric network treatments (Complete and Circle), average contributions are the same across player locations. In the Asymmetric network, contributions differ across the players' locations in the network.

The intuition is less straightforward for punishment decisions across networks. The reason is that higher contributions may be sustained by either greater actual punishment, or by the greater threat of punishment. In the former case, higher contributions in a given network may be attributed to greater actual punishment than in the other networks. In the latter case, the argument contends that the greater threat of punishment sustains the higher contributions, which necessitate less actual punishment than in the other networks. Furthermore, at the aggregate level, the comparison of absolute punishment levels across networks is complicated by the differences in total punishment capacities and in the distribution of punishment responsibility among the players. Thus, we focus our analysis on punishment decisions at the individual level, and on punishment flows between pairs of punisher and target.

In each of the networks, we also examine (for robustness) previous results concerning the effects of different factors on punishment decisions. For instance, lower contributors are punished more often (and more severely) than higher contributors. Similarly, players who contribute further below the group mean are punished more than players who contribute closer to (but still below) the group mean.

**Hypothesis 3.** In each network treatment, players assign more punishment to lower contributors than higher contributors. Furthermore, both the probability of punishment and the severity of punishment increase the further the target player's contribution falls below the group mean.

Finally, we examine the hypothesis that, due to the asymmetry in the structure of monitoring and punishment opportunities, certain players choose to discriminate between their potential targets. For instance, in the Asymmetric network, player C is solely responsible for monitoring and punishing player D, while she shares the responsibility for punishing her other two targets, A and B. Given this structure, player C may concentrate on sanctioning player D, while relying on players A and B to monitor and punish each other.

To a lesser extent, we might also expect players A and B to concentrate more on punishing each other than on punishing player C (even for comparable behavior), choosing to defer the monitoring and punishment of C to player D (for whom C is the only potential target). This generates two specific hypotheses about discriminatory punishment between potential targets.

**Hypothesis 4.** In the Asymmetric network, after controlling for target behavior, (i) player C punishes player D more often or more severely than he punishes either player A or player B; and (ii) players A and B punish each other more often or more severely than they punish player C.

## 4. Results

There are 3 main results from our study. First, we find that average contributions are lower in the Asymmetric network. Consistent with previous studies, players in the Circle network contribute similar amounts as in the Complete network. Moreover, the lower contributions in the Asymmetric network can be attributed entirely to player D, who faces only one potential punisher (player C).

Second, after controlling for contributions and for players' deviations from the group mean, we find little evidence of discriminatory punishment across targets by player C in the Asymmetric network. In other words, player C does not treat player D more harshly than he treats players A or B, even though he is solely responsible for D and shares responsibility for A and B.

Third, in the Asymmetric network, player D demonstrates a significantly higher propensity to issue anti-social punishment against player C for punishment received in previous periods. This suggests the possibility that player D uses targeted revenge to protect the advantage of being relatively under-monitored. Thus, while it is possible that player C may wish to discriminately punish player D more than his other targets, he could be deterred from doing so by the greater threat of retaliation.

### *4.1. Aggregate results*

We start by presenting some aggregate results for contributions and punishment. Fig. 2 illustrates that the average individual contribution pooled across sessions is lower for the Asymmetric treatment in all periods. In Table 1, we report summary statistics (pooled over all 10 rounds) for contributions and punishment flows in each network treatment. Consistent with Fig. 2, average contribution is almost a full token lower (out of a possible 10 tokens) in the Asymmetric network than in the other two networks.

Table 2 reports the coefficient estimates for the effects of the two incomplete network structures on contributions. We estimate the model with subject-level random effects and include session-level dummies in order to control for any potential session fixed effects. Compared with the omitted Complete network, contributions are not significantly different in the Circle network, ($p = 0.130$), but are significantly lower in the Asymmetric network ($p = 0.000$). We also find a statistically significant difference between the coefficients on the Circle network dummy and the Asymmetric network dummy ($F$-test, $p = 0.000$). In addition, we test for treatment differences using the Wilcoxon signrank test. Since the subjects in each session participated in

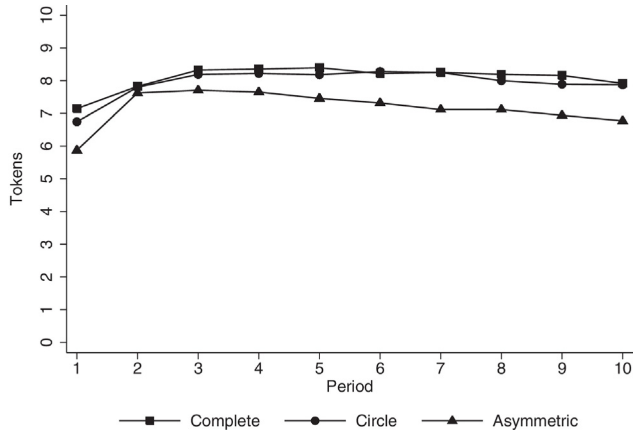



**Fig. 2.** Average contribution by period for each network.

**Table 1**
Summary statistics for the 3 network treatments.

| | Complete | Circle | Asymmetric |
|---|---|---|---|
| Average contribution | 8.0799 | 7.9444 | 7.1586 |
| | (2.8781) | (2.9939) | (3.3263) |
| Pr(Punished) | 0.1950 | 0.1646 | 0.1850 |
| | (0.3963) | (0.3709) | (0.3884) |
| Average punishment received | | | |
| Overall | 0.3057 | 0.2343 | 0.2578 |
| | (0.8753) | (0.6767) | (0.6884) |
| Conditional on being punished | 1.5675 | 1.4233 | 1.3933 |
| | (1.3978) | (1.0446) | (0.9895) |
| No. of observations | 2400 | 2400 | 3680 |

*Notes*: Averages pooled across all 10 periods. Standard deviations reported in parentheses.

**Table 2**
Estimating the effect of the network treatment on contributions.

| Dependent variable = contribution | Estimate (s.e.) |
|---|---|
| Circle | −0.1355 |
| | (0.0895) |
| Asymmetric | −0.9030*** |
| | (0.1501) |
| Constant | 7.3874*** |
| | (0.5289) |
| Session dummies | Yes |
| No. of observations | 8480 |
| $R^2$ | 0.1704 |

*Notes*: The estimated model is a panel regression with subject-level random effects and session-level dummies (to control for session-level fixed effects). Standard errors are shown in parentheses, and are clustered at the subject level. Significance levels are indicated as follows. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

all three network treatments, we compute the average contribution for each treatment by session (pooling across all subjects and all periods). Thus, we have 8 observations of average contributions for each network treatment. Consistent with Fig. 2 and Table 2, we find significant differences between Asymmetric and Complete ($p = 0.0357$) and between Asymmetric and Circle ($p = 0.0357$), but no differences between Complete and Circle ($p = 0.4838$).[11] We summarize these observations in the following result.

[11] Using the Wilcoxon ranksum test, where one observation is the average contribution in a group over all 10 periods of a Match, we also find significant differences between the Complete and Asymmetric networks ($p = 0.0087$) and between the Circle and Asymmetric networks ($p = 0.0268$), while there is no significant difference between the Complete and Circle networks ($p = 0.5903$). However, the ranksum test is not a clean test in our setting due to the potential dependence between different matches within a session.

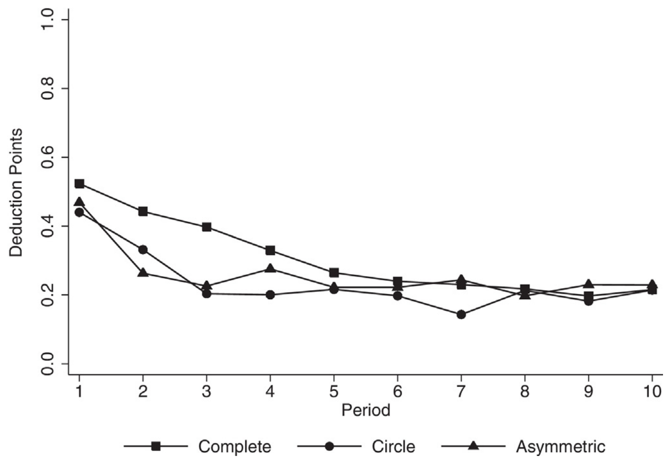



**Fig. 3.** Average punishment received by period for each network.

**Result 1.** Average contributions are lower in the Asymmetric network than in the other networks. There is no difference between the Complete and Circle networks.

Next, we consider the summary statistics for punishment reported in Table 1. There is little variation in either the probability of being punished or the intensity of punishment received across networks. In Fig. 3, we plot the average punishment received by an individual in each period for the three network treatments. In each network, the average punishment declines over time, from between 0.45 and 0.55 deduction points, to just over 0.2 deduction points.

However, at the aggregate level, the analysis of punishment flows does not provide us with much insight, since it does not control for the differences in contributions, nor does it control for the differences in punishment capacities across networks. Thus, in order to better understand the impact of the network on contributions and punishment decisions, we shift our analysis to the individual-level data.

### *4.2. Individual-level contributions*

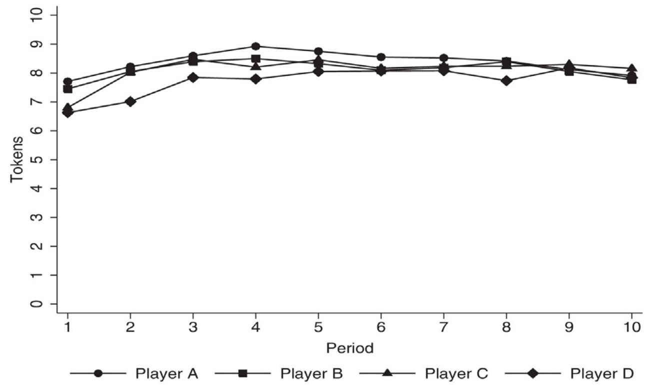
By focusing on the individual-level data, we are able to address much of the heterogeneity that is observed across different individuals. This is particularly important in regards to the Asymmetric network, where players at different locations in the network have different opportunities for, and exposure to monitoring and punishment. Fig. 4 illustrates that in the Complete and Circle networks, there are no apparent differences in contributions between player locations, while in the Asymmetric network, player D's average contribution is substantially lower than the others.

**Result 2.** Individual contributions do not differ across player location in the Complete or Circle networks. In the Asymmetric network, player D contributes significantly less than the other three players.

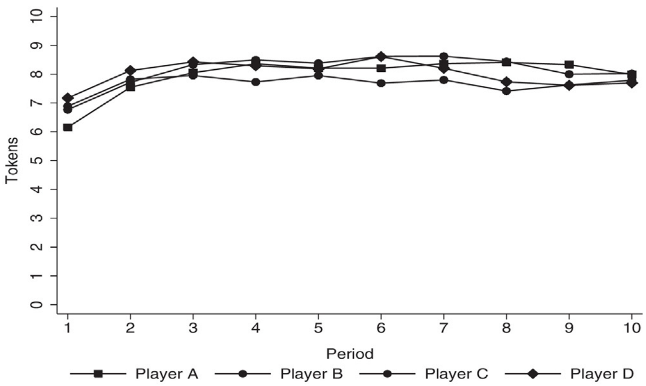
We provide statistical support for this result by estimating the effects of player location on individual contributions. We include network treatment dummies, location dummies (for player C and player D), and various interaction terms between these and the other explanatory variables. We also include the player's lagged contribution $c_{it-1}$, the positive (and negative) deviation from the average contribution in the previous period, denoted by $posdev_{it-1}$ (and $negdev_{it-1}$, respectively), the amount of punishment received in the previous period, and a variable controlling for the *period* number. The omitted network is the Complete network and the omitted player locations are positions A and B. The estimation results are reported in Table 3.

Both of the models we estimated are pooled OLS regressions with session dummy variables and robust standard errors clustered at the subject level.[12] Model (1) in Table 3 shows that individual contributions are indeed significantly lower for player D in the Asymmetric network than for the other players in the Asymmetric network as well as those in either of the other two network structures we considered. With players A and B in the Complete network as the omitted category, we find that the only significant effect is captured by the coefficient on $Asymmetric \times l_D$. We also find that individual contributions in period $t$ are significantly positively correlated with the previous period's contribution. Players who contribute higher than the average in the previous period ($posdev_{it-1}$) tended to reduce their contribution, which is consistent with conditional cooperation (perhaps driven by reciprocity or inequality aversion), and similarly, players who contributed less than the average in the previous period ($negdev_{it-1}$) significantly increased their contributions. We do not find the amount

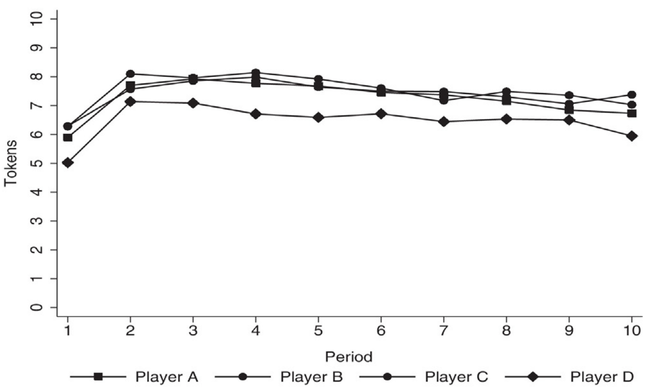
[12] We also estimated panel regression models with subject-level random effects. However, the unobservable effect term did not explain any of the variation. Thus, the results do not differ from using the pooled OLS estimator.




(a) Complete network




(b) Circle network




(c) Asymmetric network

**Fig. 4.** Average contributions by player location across networks.

of punishment received in the previous period to be significant, although this could easily be caused by the fact that lagged punishment received is significantly correlated with the other three explanatory variables (as we establish in the next set of regression results).

In model (2), we include additional interaction terms between the network treatment, the location, and two of the main explanatory variables, $c_{it-1}$ (lagged contribution) and $punrec_{it-1}$ (lagged punishment received). Compared to the excluded players A and B in the Complete network, the only significant coefficient is on the interaction between the *Asymmetric* treatment dummy, location *D* dummy, and lagged punishment received. That is, the level effect identified by model (1) can

**Table 3**
The effects of player location on individual contributions.

| Dependent variable = subject *i*'s contribution in period *t* | (1) | | (2) | |
|---|---|---|---|---|
| | Coefficient | s.e. | Coefficient | s.e. |
| $c_{it-1}$ | 0.748*** | 0.025 | 0.759*** | 0.046 |
| $posdev_{it-1}$ | −0.430*** | 0.039 | −0.438*** | 0.039 |
| $negdev_{it-1}$ | 0.184*** | 0.067 | 0.204*** | 0.069 |
| $punrec_{it-1}$ | 0.122 | 0.083 | 0.033 | 0.192 |
| *period* | −0.116*** | 0.013 | −0.114*** | 0.013 |
| *Circle* | 0.080 | 0.087 | 0.451 | 0.513 |
| *Asymmetric* | −0.020 | 0.089 | −0.094 | 0.505 |
| $l_C$ | −0.001 | 0.162 | 0.384 | 0.735 |
| $l_D$ | −0.153 | 0.140 | −0.520 | 0.853 |
| $Circle \times l_C$ | −0.302 | 0.212 | −1.320 | 1.052 |
| $Circle \times l_D$ | 0.034 | 0.178 | −0.155 | 0.811 |
| $Asymmetric \times l_C$ | −0.093 | 0.220 | −0.916 | 0.917 |
| $Asymmetric \times l_D$ | −0.477** | 0.202 | 0.457 | 0.978 |
| $Circle \times c_{it-1}$ | | | −0.056 | 0.055 |
| $Asymmetric \times c_{it-1}$ | | | 0.002 | 0.055 |
| $l_C \times c_{it-1}$ | | | −0.026 | 0.075 |
| $l_D \times c_{it-1}$ | | | 0.034 | 0.090 |
| $Circle \times l_C \times c_{it-1}$ | | | 0.117 | 0.117 |
| $Circle \times l_D \times c_{it-1}$ | | | 0.035 | 0.088 |
| $Asymmetric \times l_C \times c_{it-1}$ | | | 0.084 | 0.098 |
| $Asymmetric \times l_D \times c_{it-1}$ | | | −0.096 | 0.111 |
| $Circle \times punrec_{it-1}$ | | | 0.390 | 0.266 |
| $Asymmetric \times punrec_{it-1}$ | | | 0.307 | 0.257 |
| $l_C \times punrec_{it-1}$ | | | −0.419 | 0.264 |
| $l_D \times punrec_{it-1}$ | | | 0.331 | 0.314 |
| $Circle \times l_C \times punrec_{it-1}$ | | | −0.001 | 0.400 |
| $Circle \times l_D \times punrec_{it-1}$ | | | −0.403 | 0.445 |
| $Asymmetric \times l_C \times punrec_{it-1}$ | | | 0.376 | 0.412 |
| $Asymmetric \times l_D \times punrec_{it-1}$ | | | −1.137*** | 0.392 |
| Constant | 2.921*** | 0.339 | 2.827*** | 0.534 |
| Session dummies | Yes | | Yes | |
| No. of observations | 6480 | | 6480 | |
| $R^2$ | 0.5647 | | 0.5711 | |

*Notes*: The estimated models are both pooled OLS regressions. The Complete network is the omitted category for the network dummies. Players at locations A and B were pooled and treated as the omitted category for location dummies. In model (1), only interactions between network and location are included. In model (2), we included network and location interactions with lagged contributions and lagged punishment received. The significance of the interaction term with lagged punishment received for Asymmetric player D remains if we also include interaction terms with $posdev_{it-1}$ and $negdev_{it-1}$. Standard errors are clustered at the subject level. Significance levels are indicated as follows. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

be further sharpened. Player *D* in the *Asymmetric* network is significantly less likely to increase contributions after being punished than all of the other players.

### 4.3. Individual-level punishment

Next, we investigate the factors that influence the amount of punishment received by an individual. Fig. 5 shows that, for all 3 network treatments, the frequency with which a player receives positive punishment (from at least one of their neighbors) is decreasing in the player's contribution. Similarly, Fig. 6 shows that the mean amount of punishment received (conditional on being punished) is also declining with the player's contribution level. Thus, as should be expected, on average, higher contributors are punished less frequently and less severely than lower contributors. We then test for systematic differences in the frequency and intensity of punishment received across player locations. While Hypothesis 4 predicts differences in the Asymmetric network, player location should have no effect on punishment received in the Complete or Circle networks, after controlling for the recipient's contribution decision.

For both the Complete and Circle networks, we estimate the following panel tobit regression with individual random effects.

$$punrec^*_{it} = \beta_0 + \beta_1 c_{it} + \beta_2 posdev_{it} + \beta_3 negdev_{it} + \beta_4 period + \alpha_1 l_C + \alpha_2 l_D + u_i + \varepsilon_{it} \quad (3)$$

where $punrec^*_{it}$ is the latent variable for punishment received by player *i* in period *t*. The observed variable, $punrec_{it}$ is censored below (at 0) and above (at 6 in the Complete network, and at 4 in the Circle network). The results, reported for

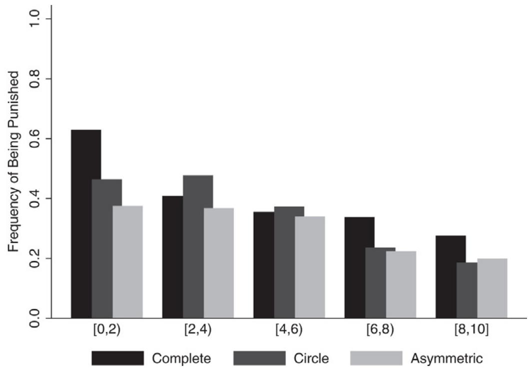



**Fig. 5.** Frequency of being punished across different contribution ranges.

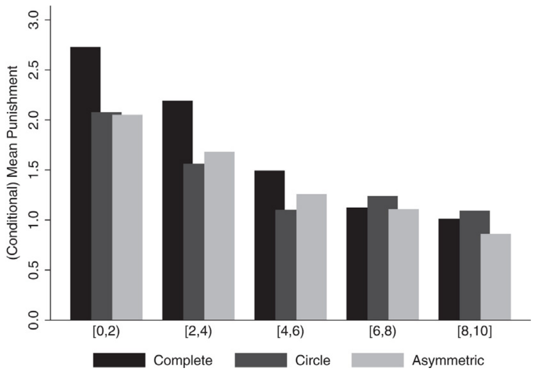



**Fig. 6.** Mean punishment received (conditional) across different contribution ranges.

both networks in Table 4, show that none of the location dummy variables are significant.[13] The regression estimates thus provide statistical support for our next result, confirming the first part of Hypothesis 4.

**Result 3.** Player location does not affect punishment received in the Complete or Circle networks.

In addition to Result 3, the marginal effects computed for the panel Tobit models reported in Table 4 also confirm several other established results. Lower contributions increase the expected amount of punishment received and the probability of being punished, especially for those who are already contributing less than the group mean.[14] In contrast, for players who contribute more than the group mean ($posdev_{it} > 0$), the mean probability of being punished (and the conditional severity of punishment) is close to 0. As a result, we need to interpret the significant, positive coefficient on $posdev_{it}$ in conjunction with the coefficient on contribution, $c_{it}$. In the Complete and Circle networks, the lower bound on punishment received is mostly binding for players who contribute above the group average contribution. Thus, the overall negative effect of contributing a higher amount on the probability and severity of punishment is counteracted by a positive, significant coefficient on $posdev_{it}$.

### 4.4. Individual punishment flows in the asymmetric network

In the Asymmetric network, the maximum possible amount of punishment received varies across player location. Thus, we cannot run the comparable regression analysis used for the Complete and Circle networks reported in Table 4. In order to better understand the punishment decisions in the Asymmetric network, we examine individual punishment flows between pairs of players. In particular, we are interested in determining whether players discriminate among their punishment targets based on asymmetries in the network. In Hypothesis 4, we identified two predictions about punishment flows in the Asymmetric network. First, since player C is solely responsible for monitoring and punishing player D, we ask whether C punishes D disproportionately more than he punishes A or B for the same contribution decisions. Second, since player

[13] We also checked for significance of the interaction terms between player location and the other explanatory variables. In the Complete network, none of the interaction terms are significant, while in the Circle network, the term interacting location *C* with $posdev_{it}$ is negative and marginally significant ($p = 0.043$). No other interaction terms were significant for the Circle network.

[14] We compute marginal effects of each regressor on both $\mathbb{E}[punrec^*_{it}]$ and $\mathbf{Pr}[punrec^*_{it} > 0]$. Since the sign and significance levels of all the coefficients translate to each of the marginal effect, we do not report them here.

**Table 4**
Punishment received in the complete and circle networks.

| | Complete network | Circle network |
|---|---|---|
| Dependent variable = $punrec^*_{it}$ | (1) | (2) |
| $c_{it}$ | −0.301*** | −0.296*** |
| | (0.054) | (0.039) |
| $posdev_{it}$ | 0.375*** | 0.385*** |
| | (0.101) | (0.060) |
| $negdev_{it}$ | 0.672*** | 0.277*** |
| | (0.091) | (0.066) |
| $period$ | −0.134*** | −0.055*** |
| | (0.028) | (0.021) |
| $l_C$ | 0.084 | −0.190 |
| | (0.291) | (0.201) |
| $l_D$ | −0.053 | 0.152 |
| | (0.288) | (0.211) |
| Constant | 0.322 | 0.293 |
| | (0.596) | (0.440) |
| Session dummies | Yes | Yes |
| No. of observations | 1600 | 2400 |
| Log-likelihood | −749.73 | −1266.41 |

*Notes*: Columns (1) and (2) report coefficient estimates for the latent dependent variable $punrec^*_{it}$ in a panel tobit model with subject level random effects. Standard errors are reported in parentheses. Significance levels are indicated as follows. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

C is also monitored by player D, we ask whether players A and B punish each other more than they punish player C, after controlling for contribution decisions.

We define $d_{ijt}$ to be the number of deduction points assigned by player $i$ to player $j$ in period $t$. Then we estimate the following panel tobit model for the Asymmetric network,

$$d_{ijt} = \sum_{k\in\mathcal{L}}\beta_{1k}(l_i^k \times c_{it}) + \sum_{(p,r)\in\mathcal{P}}\beta_2^{(p,r)}(l_i^p \times l_j^r \times c_{jt}) + \sum_{(p,r)\in\mathcal{P}}\beta_3^{(p,r)}(l_i^p \times l_j^r \times posdev_{jt}) + \sum_{(p,r)\in\mathcal{P}}\beta_4^{(p,r)}(l_i^p \times l_j^q \times negdev_{jt})$$
$$+ \beta_5 \cdot period + u_i + \varepsilon_{ijt}$$

where $\mathcal{L} = \{A, B, C, D\}$ is the set of locations and $\mathcal{P} \subset \mathcal{L}\times\mathcal{L}$ is the set of all feasible pairs of punisher location ($p$) and recipient location ($r$) in the Asymmetric network. That is, the model estimates the effects, on the number of deduction points assigned, of the punishing player's contribution interacted with location, the recipient's contribution interacted with the pair of locations for punisher and recipient, the recipient's deviation from the group mean (both positive and negative) interacted with the pair of locations, and the period number. We report the results in Table 5.

Comparing the coefficients across player C's targets ('C to A' vs. 'C to B' vs. 'C to D'), we find no significant difference in the effect of the recipient's contribution $c_{jt}$ using a joint test ($\chi^2(2) = 3.31$, $p = 0.19$), although the pairwise comparison between 'C to B' and 'C to D' is marginally significant ($\chi^2(1) = 3.09$, $p = 0.079$). The coefficients on the recipient's positive deviation are not significantly different from zero for any of player C's potential targets. On the other hand, the effect of the recipient's negative deviation from the mean is weaker for punishment assigned by player C to player A than for punishment assigned to B or D.[15] Similarly, comparing across player A's targets ('A to B' vs. 'A to C') and across player B's targets ('B to A' vs. 'B to C'), we find no differences between the coefficients for recipient's contribution, recipient's positive deviation, or recipient's negative deviation. Thus, contrary to Hypothesis 4, we find no evidence that the players discriminate between their targets.

**Result 4.** In the Asymmetric network, controlling for the target's contribution and deviation from the mean, (i) Player C does not discriminately punish player D more than he punishes player A or player B, and (ii) Players A and B do not punish each other discriminately more than they punish player C.

Nevertheless, as discussed in Section 1, the data provide a potential explanation for these results. Specifically, we find evidence to suggest that player C is deterred from issuing more punishment to player D because D retaliates against punishment received from C. The estimates reported in Table 5 provide initial evidence to support this explanation.

If player D were engaging in a considerable amount of targeted revenge, this would tend to generate perverse or counter-intuitive coefficient estimates for the punishment flows from 'D to C'. There are three such perverse estimates that we find in the data for punishment issued by D to C. First, punishment assigned by player D is decreasing in his own (punisher's) contribution (significant at the 1% level). This has the unexpected implication that at location D, lower contributors issue more punishment than higher contributors. However, this would be the case if lower contributing player Ds were issuing anti-social punishment out of retaliation. Furthermore, this perverse effect does not arise for any other location.

[15] We report the $\chi^2$ statistics and corresponding $p$ values for these comparisons in the Online Supplement.

**Table 5**
Pairwise punishment flows in the Asymmetric network.

| Dependent variable = $d_{ijt}$ | | Coefficient | Std. Dev. |
|---|---|---|---|
| Punisher's contribution, $c_{it}$ | $l_A$ | 0.081 | (0.093) |
| | $l_B$ | 0.262*** | (0.081) |
| | $l_C$ | 0.072 | (0.071) |
| | $l_D$ | −0.227*** | (0.086) |
| Recipient's contribution, $c_{jt}$ | A to B | −0.332*** | (0.111) |
| | A to C | −0.282*** | (0.109) |
| | B to A | −0.506*** | (0.100) |
| | B to C | −0.493*** | (0.099) |
| | C to A | −0.272*** | (0.087) |
| | C to B | −0.322*** | (0.089) |
| | C to D | −0.221*** | (0.085) |
| | D to C | −0.054 | (0.089) |
| Recipient's positive deviation, $posdev_{jt}$ | A to B | −0.059 | (0.286) |
| | A to C | 0.008 | (0.276) |
| | B to A | 0.634*** | (0.234) |
| | B to C | 0.252 | (0.266) |
| | C to A | −0.172 | (0.287) |
| | C to B | 0.147 | (0.244) |
| | C to D | −0.063 | (0.299) |
| | D to C | 0.523** | (0.248) |
| Recipient's negative deviation, $negdev_{jt}$ | A to B | 0.629*** | (0.186) |
| | A to C | 0.875*** | (0.178) |
| | B to A | 0.520*** | (0.176) |
| | B to C | 0.467*** | (0.174) |
| | C to A | 0.427** | (0.167) |
| | C to B | 0.959*** | (0.172) |
| | C to D | 0.816*** | (0.122) |
| | D to C | 1.084*** | (0.176) |
| Period | | −0.082*** | (0.028) |
| Session dummies | | Yes | |
| No. of observations | | 6400 | |
| Log-likelihood | | −2223.91 | |

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Second, recipient's contribution has a significant negative effect on punishment for every pairwise punishment flow *except* for the punishment from D to C. The fact that this effect does not show up for punishment from player D to player C suggests that some high contributing player Cs must also be punished, as would be the case if D were retaliating against previous punishment received, since that previous punishment would most likely have been issued by a high contributing player C.[16] Third, recipient's positive deviation from the mean has a significant positive effect on punishment from player D to player C. Again, this perverse coefficient estimate is indicative of anti-social punishment by player D, in retaliation against a high-contributing player C for previous punishment.[17]

We provide more formal evidence of retaliation by player D using a more direct approach. First, we define a pairwise punishment flow from player $i$ to player $j$ in period $t$ to be an *opportunity for retaliation* if

(i) in the previous period, $t-1$, player $j$ assigned positive punishment to player $i$;
(ii) in period $t$, player $j$ contributed more than the group mean contribution; and
(iii) in period $t$, player $j$ contributed more than player $i$.

When the three conditions in this definition are satisfied, punishment from player $i$ (the punisher) to player $j$ (the recipient) in the current period is not easily justified in the typical sense (as a sanction for low contribution by the recipient), but rather is consistent with retaliation.[18] Table 6 summarizes the opportunities for retaliation and the actual incidence of retaliation observed overall, by network, and for the punishment flow from player D to player C in the Asymmetric network.

First, notice that the numbers and percentages are very similar whether or not we exclude the data from Match 1. For consistency with the rest of the analysis in this paper, we concentrate the discussion on the results after excluding Match 1

[16] Alternatively, this might also reflect an attempt by player D to punish player C "pre-emptively" in order to deter sanctions, rather than retaliation. In either case, we can interpret player D's behavior as anti-social, in the sense that player D is not using punishment with the motive of getting player C to contribute more. We thank the editor for providing these additional insights.

[17] For all other punishment flows except one, the recipient's positive deviation has no effect. Surprisingly, the effect of $posdev_{jt}$ is also significantly positive for punishment assigned by player B to player A. Nevertheless, player B's behavior does not exhibit any of the other perverse patterns that we identify.

[18] The identification of retaliation is not as clean as when there is a separate counter-punishment stage, as in Nikiforakis (2008), but our definition should understate the extent to which retaliation occurs.

**Table 6**
Frequency of revenge punishment.

| | Obs. | Opportunity for retaliation | | Retaliation | |
|---|---|---|---|---|---|
| | | Frequency | % of observations | Frequency | % of opportunities |
| Overall | 14,400 | 873 | 6.06% | 121 | 13.86% |
| Complete | 4320 | 210 | 4.86% | 35 | 16.67% |
| Circle | 4320 | 285 | 6.60% | 28 | 9.82% |
| Asymmetric | 5760 | 378 | 6.56% | 58 | 15.34% |
| "D to C" | 720 | 110 | 15.28% | 26 | 23.64% |
| Others | 5040 | 268 | 5.32% | 32 | 11.94% |

*Notes*: (1) Obs. gives the number of observations of punishment flows (between period 2 and period 10); (2) % of Obs. gives the percentage of observations that are classified as opportunities for retaliation; (3) % of Opportunities gives the percentage of opportunities for retaliation that are actually taken.

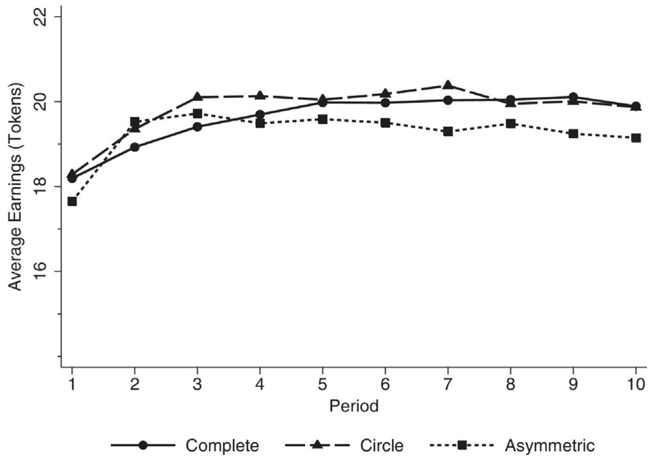



**Fig. 7.** Average earnings across networks by period.

data. Overall, out of 14,400 punishment decisions, 873 (6%) give the punisher an opportunity to exact revenge for punishment received in the previous period. This opportunity is taken 14% of the time (121 observations). The percentage of observations that give the punisher an opportunity to retaliate are similar across networks, if slightly lower in the Complete network.

However, within the Asymmetric network, the punishment flow from 'D to C' provides a revenge opportunity 15% of the time, while other punishment flows provide the opportunity for revenge only 5.3% of the time. Moreover, player D takes advantage of the opportunity to retaliate 23.6% of the time it is available, compared with 12% of the time for other punishment flows. Thus, not only does player D have the opportunity to retaliate with a much higher percentage of his punishment decisions, he also takes advantage of those opportunities much more frequently. We summarize these observations in the following result.

**Result 5.** Player D exacts more targeted revenge than other players, using anti-social counter-punishment to retaliate against previous punishment received from player C.

### 4.5. *Efficiency*

Finally, we compare the level of efficiency (average payoffs) across networks and, in the Asymmetric network, across player locations, after accounting for the costs of punishment. Fig. 7 shows the average group payoffs by period in the 3 networks. In the first 5 periods, there does not appear to be any difference between the networks. However, in the last 5 periods, when punishment levels are similar across networks, the lower contributions in the Asymmetric network also lead to lower average earnings (lower efficiency). Using the Wilcoxon signrank test to compare the average earnings for each network treatment (pooled across all subjects and all periods in a session), we find no significant difference between the Complete and Circle networks ($p = 0.4838$) or between the Complete and Asymmetric networks ($p = 0.2626$), while there is weak evidence of a difference between the Circle and Asymmetric networks ($p = 0.0687$). In part, this reflects the fact that we have only 8 observations per treatment for the signrank test.[19] In Fig. 8, we plot the average earnings over time for the

[19] Using the Wilcoxon rank-sum test to compare the average earnings across networks, when pooled over all periods, reveals no significant difference between Complete and Circle ($p = 0.3436$), but statistically lower average earnings in the Asymmetric network than in both Complete ($p = 0.0141$) and Circle ($p = 0.0748$). However, as noted earlier in the paper, the ranksum test is not a clean test in our setting due to the potential dependence between different matches within a session.

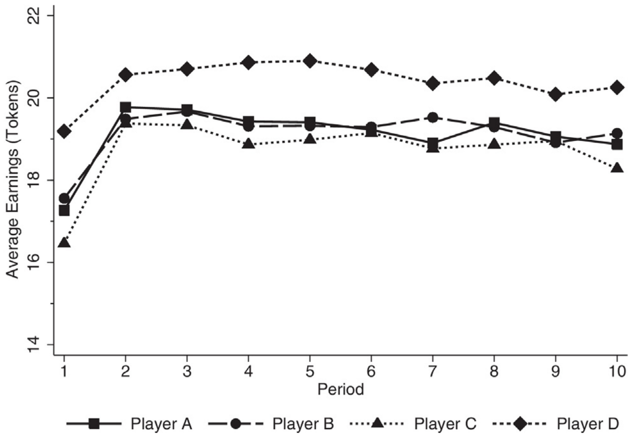



**Fig. 8.** Average earnings across player location in the Asymmetric network.

different player locations in the Asymmetric network. The figure illustrates that average earnings are considerably higher for player D than for the other 3 players. Thus, the efficiency comparison tracks our results on contributions quite closely, with the under-monitored player D enjoying higher average earnings than the others.

## 5. Conclusion

Our study contributes to the recent literature exploring the effect of alternative network structures on the effectiveness of monitoring and punishment in public goods games. In particular, we show that asymmetry in the network can inhibit the use of punishment to sustain higher contributions; in part because of the free-riding behavior of under-monitored group members, and in part because these under-monitored players have greater incentives to use anti-social counter-punishment as a means of exacting revenge. In comparing our Asymmetric network with the Complete and Circle networks, we found that average contributions are significantly lower. These findings suggest that the distribution and concentration of punishment responsibilities are critically important factors for sustaining cooperation. While asymmetry may suggest that players might wish to punish different targets discriminately, based on whether or not they share the monitoring responsibility, it also facilitates greater retaliation by under-monitored players who seek to protect their positional advantage. In future research, our understanding of the effectiveness of decentralized punishment may benefit greatly from further examination into the effects of different kinds of asymmetry in the network.

While our study provides some initial evidence, two natural extensions would be to examine the effects of asymmetry in other network structures and in environments where players are given explicit opportunities for counter-punishment (as in Nikiforakis, 2008). In addition, while our study restricts the feedback observed by players about punishment to the flows in which they are involved, it may be interesting to examine the effect of providing information about punishment flows between other players on the coordination of punishment by players who share responsibility for monitoring a particular target. Finally, an additional interesting extension, closely related to endogenous network formation, might allow subjects to sever links from an initial exogenous network structure. If there is a cost to severing existing links, we can interpret this setup as one in which players can invest in hiding their behavior from potential punishers in the group.

## Acknowledgements

We would like to thank two anonymous referees and the handling Editor for their many helpful comments, as well as Sebastian Goerg, Carl Kitchens, and participants at the 2013 North American Econometric Society Meetings in Los Angeles, the 2013 SEA Meetings in Tampa, and the 2014 ESA Conference in Ft. Lauderdale. We gratefully acknowledge financial support from the College of Social Sciences and Public Policy at Florida State University.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jebo.2016.09.015

## References

Anderson, C.M., Putterman, L., 2006. Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. Games Econ. Behav. 54 (1), 1–24.

Bloch, F., Genicot, G., Ray, D., 2008. Informal insurance in social networks. J. Econ. Theory 143 (1), 36–58.

Bochet, O., Page, T., Putterman, L., 2006. Communication and punishment in voluntary contribution experiments. J. Econ. Behav. Organ. 60 (1), 11–26.
Bramoullé, Y., Kranton, R., 2007a. Public goods in networks. J. Econ. Theory 135 (1), 478–494.
Bramoullé, Y., Kranton, R., 2007b. Risk-sharing networks. J. Econ. Behav. Organ. 64 (3–4), 275–294.
Carpenter, J., Kariv, S., Schotter, A., 2012. Network architecture, cooperation and punishment in public good experiments. Rev. Econ. Des. 16 (2–3), 93–118.
Carpenter, J., Matthews, P.H., 2009. What norms trigger punishment? Exp. Econ. 12 (3), 272–288.
Carpenter, J.P., 2007a. Punishing free-riders: how group size affects mutual monitoring and the provision of public goods. Games Econ. Behav. 60 (1), 31–51.
Carpenter, J.P., 2007b. The demand for punishment. J. Econ. Behav. Organ. 62 (4), 522–542.
Chaudhuri, A., 2010. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. Exp. Econ. 14 (1), 47–83.
Choi, S., Gallo, E., Kariv, S., 2016. Networks in the laboratory. In: Bramoullé, Y., Galleotti, A., Rogers, B. (Eds.), The Oxford Handbook on the Economics of Networks. Oxford University Press, Oxford; New York.
Cinyabuguma, M., Page, T., Putterman, L., 2005. Cooperation under the threat of expulsion in a public goods experiment. J. Public Econ. 89 (8), 1421–1435.
Cinyabuguma, M., Page, T., Putterman, L., 2006. Can second-order punishment deter perverse punishment? Exp. Econ. 9 (3), 265–279.
Denant-Boemont, L., Masclet, D., Noussair, C.N., 2007. Punishment, counter punishment and sanction enforcement in a social dilemma experiment. Econ. Theory 33 (1), 145–167.
Egas, M., Riedl, A., 2008. The economics of altruistic punishment and the maintenance of cooperation. Proc. R. Soc. B: Biol. Sci. 275 (1637), 871–878.
Fehr, E., Gächter, S., 2000. Cooperation and punishment in public goods experiments. Am. Econ. Rev. 90 (4), 980–994.
Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10 (2), 171–178.
Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1 (1), 114–125.
Hermann, B., Thoni, C., Gächter, S., 2008. Antisocial punishments across societies. Science 319, 1362–1367.
Jackson, M.O., Rodriguez-Barraquer, T., Tan, X., 2012. Social capital and social quilts: network patterns of favor exchange. Am. Econ. Rev. 102 (5), 1857–1897.
Karlan, D., Mobius, M., Rosenblat, T., Szeidl, A., 2009. Trust and social collateral. Q. J. Econ. 124 (3), 1307–1361.
Ledyard, J.O., 1995. Public goods: some experimental results. In: Kagel, J., Roth, A. (Eds.), Handbook of Experimental Economics. Princeton, Princeton University Press (Chapter 2).
Leibbrandt, A., Ramalingam, A., Sääksvuori, L., Walker, J.M., 2015. Incomplete punishment networks in public goods games: experimental evidence. Exp. Econ. 18 (1), 15–37.
Masclet, D., Noussair, C., Tucker, S., Villeval, M.-C., 2003. Monetary and nonmonetary punishment in the voluntary contributions mechanism. Am. Econ. Rev. 93 (1), 366–380.
Nikiforakis, N., 2008. Punishment and counter-punishment in public good games: can we really govern ourselves? J. Public Econ. 92 (1–2), 91–112.
Nikiforakis, N., 2010. Feedback, punishment and cooperation in public goods experiments. Games Econ. Behav. 68 (2), 689–702.
Nikiforakis, N., Normann, H.-T., 2007. A comparative statics analysis of punishment in public-good experiments. Exp. Econ. 11 (4), 358–369.
Nikiforakis, N., Normann, H.-T., Wallace, B., 2010. Asymmetric enforcement of cooperation in a social dilemma. South. Econ. J. 76 (3), 638–659.
Noussair, C., Tucker, S., 2005. Combining monetary and social sanctions to promote cooperation. Econ. Inq. 43 (3), 649–660.
O'Gorman, R., Henrich, J., Van Vugt, M., 2009. Constraining free riding in public goods games: designated solitary punishers can sustain human cooperation. Proc. R. Soc. B: Biol. Sci. 276 (1655), 323–329.
Ostrom, E., Walker, J., Gardner, R., 1992. Covenants with and without a sword: self-governance is possible. Am. Polit. Sci. Rev. 86 (2), 404–417.
Rosenkranz, S., Weitzel, U., 2012. Network structure and strategic investments: an experimental analysis. Games Econ. Behav. 75 (2), 898–920.
Sefton, M., Shupp, R., Walker, J.M., 2007. The effect of rewards and sanctions in provision of public goods. Econ. Inq. 45 (4), 671–690.