WORKING PAPERS

N° 15-23

Authority and Centrality

Power and Cooperation in Social Dilemma Networks

Boris van Leeuwen, Abhijit Ramalingam,

David Rojo Arjona and Arthur Schram

IAST

INSTITUTE for ADVANCED STUDY in TOULOUSE

Authority and Centrality

Power and Cooperation in Social Dilemma Networks

Boris van Leeuwen[a]
Abhijit Ramalingam[b]
David Rojo Arjona[c]
Arthur Schram[d]

March 3, 2015

Abstract: We investigate the effects of power on cooperation in repeated social dilemma settings. Groups of five players play either multi-player trust games or VCM-games on a fixed network. Power stems from having the authority to allocate funds raised through voluntary contributions by all members and/or from having a pivotal position in the network (centrality). We compare environments with and without ostracism by allowing players in some treatments to exclude others from further participation in the network. Our results show that power matters but that its effects hinge strongly on the type involved. Reminiscent of the literature on leadership, players with authority often act more cooperatively than those without such power. Nevertheless, when possible, they are quickly ostracized from the group. Thus, this kind of power is not tolerated by the powerless. In stark contrast, centrality leads to less cooperative behavior and this free riding is not punished; conditional on cooperativeness, players with power from centrality are less likely to be ostracized than those without. Hence, not only is this type of power tolerated, but so is the free riding it leads to.

JEL Codes: C91, D02, D03, H41
Keywords: power, cooperation, networks, public goods

[a] Institute for Advanced Studies in Toulouse, Toulouse School of Economics, 21 Allée de Brienne, 31000 Toulouse, France; Boris.vanLeeuwen@iast.fr
[b] School of Economics and CBESS, University of East Anglia, Norwich, UK; A.Ramalingam@uea.ac.uk
[c] Department of Economics, University of Leicester, Leicester, UK; email: dra9@le.ac.uk
[d] Amsterdam School of Economics, CREED and Tinbergen Institute, Amsterdam, The Netherlands; Schram@uva.nl

Acknowledgements: For useful comments, the authors would like to thank Jordi Brandts, Jeff Carpenter, Ron Harstad, Andreas Leibbrandt, Jeroen van de Ven and participants at seminars at UC Riverside, UC Irvine, University of Pennsylvania (SAS), University of East Anglia, ESI Chapman University, University of Málaga, Jadavpur University and MPI Bonn, as well as participants at the workshop on Power, Games, and Fairness (Turku 2013), the 2013 ESA North American meetings in Tucson, the 2013 Southern Economic Association meeting, the 2014 Incentives and Behavior Change Workshop at the University of Amsterdam and the 2014 CCC-meeting at the University of Nottingham. Part of this paper was written while Arthur Schram was visiting the Institute of Economic Analysis (CSIC-IAE) in Barcelona. He thanks the IAE for their hospitality. Financial support from the Research Priority Area Behavioral Economics of the University of Amsterdam, the University of East Anglia, Chapman University and the ANR - Labex IAST is gratefully acknowledged.

"Nearly all men can stand adversity, but if you want to test a man's character, give him power."

attributed to Abraham Lincoln

1 Introduction

It has long been a widely shared 'truism' that the possession of power leads to the dilution of an individual's moral character (see DeCelles et al. 2012 and the references therein). A famous example is the abuse of power by 'prison guards' in the 1971 Stanford prison experiment (Zimbardo 2000). A common presumption is that the actions of powerful individuals will be more strongly motivated by self-interest than the actions of others (e.g., Persson et al. 2003). Such self-interested behavior is often harmful to society at large. Indeed, due to the belief that "power is apt to corrupt the minds of those who possess it" (William Pitt the Elder, 1848:94) many social and public bodies are designed to divide authority and responsibility among individuals and across levels within the institution.

Little is known, however, about the relationship between having power and an increase in self-interested actions. Moreover, different sources of power may have distinct effects. Even if they have similar effects on individuals with power, those without may respond differently to the (ab)use of distinct kinds of power. In this paper we will focus on what we believe are two of the main sources of power: (i) control over the allocation of surplus and (ii) a pivotal position in society or an organization. The power to allocate surplus (authority) is observed in many organizations, ranging from wage determination in small firms to government spending at the national level. Power through centrality arises when someone's presence is vital for keeping a group together. For example, centrality may arise in networks due to social connections, particular skills or location. Think of a firm where one team leader brings together two teams with diverse skill sets. Or geographically, if several nations or regions want to collaborate but can only do so if a centrally located region is involved.

Though it is widely believed that power 'corrupts', it is difficult to obtain empirical validation of this belief because reliable data are difficult to obtain in the field (for one thing, because most people do not like to report that they have been affected in this way). We will use the laboratory to collect data that will allow us to investigate the consequences of introducing asymmetries in power. We choose an environment where there is a tension between one's own interest and the common good, to wit, cooperation in social dilemma games. An important advantage of our approach is that cooperation and free riding in social dilemmas are phenomena that are well understood.

We thus aim to better understand the effects of power asymmetries in social dilemma games. We investigate whether and how power affects cooperative behavior, distinguishing between power from authority and power from centrality. We ask both how having power affects one's actions and how others (without power) respond to this. For the response to power abuse, we study the potential of one mechanism in particular – ostracism of group members – to counterbalance the negative effects of power. Specifically, we address three research questions in this paper: (i) How is cooperative behavior affected by having power? (ii) How do players without power respond to power asymmetries? (iii) Does the abuse of power and the use of ostracism vary between power from authority and power from centrality?

To create power asymmetry in a social dilemma, we impose an exogenous network structure. This allows us to create, in a natural way, the two types of power that we are interested in. To create power derived from *authority*, subjects play a five-player game of trust where one player has discretion over how the surplus created by voluntary contributions is distributed among the members of the group. This is compared to a classic voluntary contribution mechanism (VCM), where every player automatically receives an equal share of the group surplus. Other than this difference in how proceeds are distributed, the two games are equivalent.

To see how people react to power asymmetries in their groups, in some treatments, we allow players the opportunity to ostracize others using a majority-voting rule (as in Cinyabuguma et al. 2005). Excluding people from a group is an obvious way in which abuse of power can be addressed. Think of voting out a politician or removing all social ties with a corrupt leader. In this way, ostracism may be a strong tool in the hands of the powerless to counteract the abuse of power.

The opportunity to ostracize group members also allows us to investigate power from *centrality*. This occurs in networks where one or a few pivotal players are essential for keeping the network connected.[1] Sociologists have long known that such centrality produces power in networks (Bonacich, 1987). The essence of power through centrality is intuitive; leaving out the central player causes the network to fall apart, which is a costly act for other actors involved. For this reason, power though centrality may lead to what anthropologists call 'tolerated theft' (e.g., Blurton Jones, 1984, in a different context). This occurs when

[1]Centrality in a network can be characterized by maximal betweenness, nearness, degree and influence (Freeman, 1979; Jackson, 2008). Betweenness measures the frequency with which a point (i.e., player) in a network falls between pairs of other points. Nearness is defined by the distance of a point to all other points in the network. A point's degree measures the number of other points it is connected to. Influence is closely related to betweenness and measures the fraction of the network that is affected by changes in a point (e.g., decisions by a player).

agents allow others to take advantage of their social position. As far as we are aware, this hypothesis has hitherto not been systematically studied in the economic and game-theoretic literature. Moreover, even if theft by those with power through centrality is indeed tolerated, it remains an open question whether the same holds for other forms of power.

The introduction of players with either type of power creates heterogeneity across players in the social dilemmas that we study. This in itself may affect cooperation. Whereas most work on social dilemmas has focused on homogeneous groups, there are notable exceptions, starting with the seminal work of Ostrom et al. (1994). Players can be heterogeneous in terms of endowments (Cherry et al. 2005), marginal productivity (Tan, 2008; Tan & Noussair, 2011) and marginal benefits (Fisher et al. 1995), for example. How heterogeneity affects cooperation between different types of players and at the group level depends on the particulars of the game (see Reuben and Riedl 2013 for a recent overview of the literature). In contrast to these studies on VCMs, there are very few studies of heterogeneity in trust games, partly due to the predominant restriction to two players. An exception is Anderson et al. (2006) who show that heterogeneity in show-up fees does not affect choices.[2] All in all, no clear picture has yet emerged about the effects of various types of heterogeneity on behavior in social dilemmas.[3]

Our results show that players with authority try to 'lead' their groups towards more efficient outcomes; in the absence of ostracism, they are even more cooperative than are those without power. With ostracism, they are initially as cooperative as are the powerless. However, when ostracism is available, players with authority are often excluded, and very early in the game. This suggests that, though authority has the potential to raise efficiency, those without power resent it and do not allow the realization of this potential. In contrast, centrality is often used as a license to act uncooperatively. Though the possibility of ostracism does increase cooperation levels, central players free ride more than others even though they can be ostracized. The others tolerate such behavior; despite their lower cooperativeness, central players are not ostracized more often.

Our results for authority put the previous experimental literature in a new perspective.

[2] Also related to our authority treatments are the 3-person trust games studied by Cassar & Rigdon (2011) and Buskens et al. (2010). An important difference with these two studies is that in their setup the two 'senders' play separate trust games with one 'receiver' whereas in our study all 'senders' play a joint game in the sense that all contributions are added up and the receiver cannot return distinct amounts across the senders.

[3] There are several experimental studies that investigate public good provision in a network. The network determines which contributions can be accessed (Rosenkranz and Weitzel 2013; Charness et al. 2014) who can monitor whom (Eckel et al. 2010; Fatas et al. 2010), who can punish whom (Leibbrandt et al. 2014) or a combination of these (Carpenter et al. 2012). However, none of these studies involves power heterogeneity.

Far from leading to corrupt behavior, this form of power, often referred to as 'leadership', has been shown to have the potential to *raise* cooperation in groups (see, for instance, van der Heijden et al. 2009 and Stoddard et al. 2014). However, previous work has not considered how those without power react to the presence of such 'leaders' in their groups, *other than* in the own cooperation decisions. While leaders may act cooperatively, it cannot be taken for granted that others expect them to do so. If anything, the truism alluded to at the beginning of the paper suggests the opposite. Our results show that others jump on the opportunity to ostracize such leaders. This also adds to our understanding of the role of ostracism in social dilemmas, which has previously been shown to raise cooperation in groups (Cinyabuguma et al. 2005). We show that this may no longer hold under power asymmetry.

Closest to our study is a recent paper by Cox et al. (2013) who also study social dilemmas with power asymmetries. In particular, their treatment where one player (a "King") moves last and can decide to either take from or contribute to a common fund relates to our treatments where the central player has authority over the generated surplus. They find that the presence of such a King leads to significantly lower contributions by the other players compared to the baseline VCM without power asymmetries. However, Cox et al. (2013) does not involve a network structure and as a consequence, there is no power from centrality. Further, they do not allow for ostracism or other mechanisms that allow others to react to power. In contrast, a comparison between types of power and the reactions of the powerless are at the core of the research we present here. To the best of our knowledge, we are the first to study the effects of power derived from authority *and* network centrality in social dilemmas.

The remainder of this paper is organized as follows. The next section presents our experimental design and section 3 presents our testable hypotheses. Section 4 presents our results and section 5 concludes.

2 Experimental Design and Procedures

Subjects interact in fixed groups of five participants, with whom they play five repetitions of a game explained below. Each session consists of five blocks of such five-round games, with new groups in each five-round game. We employ a 2 × 2 × 2 between-subject design, where we vary (i) whether or not the stage game they play induces power from authority; (ii) whether or not the network structure induces power from centrality; (iii) the availability of ostracism as a mechanism to enhance cooperation. Power from centrality, however, can only

Table 1: Summary of treatments

		No Authority	Authority
No Ostracism (⇒ no Centrality)		*nAnO* VCM-no voting N = 85 (n = 6)	*AnO* Trust-no voting N = 85 (n = 6)
Ostracism	**No Centrality**	*nAnC* VCM-voting N = 75 (n = 6)	*AnC* Trust-voting N = 65 (n = 6)
	Centrality	*nAC* VCM-voting N = 90 (n = 6)	*AC* Trust-voting N = 75 (n = 6)

Notes. The first row in each cell gives the treatment acronym. The second indicates the stage game played (a VCM or a multi-person trust game) and whether or not there was a voting stage to exclude players. The last row gives the number of subjects (N) and independent matching groups (n). Each subject played the five-round stage game five times, with new groups formed from a 10- or 15-person matching group.

exist if ostracism is possible; if a central player cannot be removed from the network, she will always connect the others. The treatment combinations no-authority/centrality/no-ostracism and authority/centrality/no-ostracism therefore do not exist. This leaves six treatments cells, which are summarized in Table 1 and discussed in more detail below. Table 1 also defines the acronyms that we will use and shows the number of observations we have for each treatment.

In all treatments, subjects play on one of two networks consisting of five players labeled ‘Center’ (C), ‘North’ (N), ‘East’ (E), ‘South’ (S) and ‘West’ (W) (see Figures 1a and 1d). We will refer to the last four players as ‘periphery players’. Subjects play the game with all their *neighbors* in the network. The neighbors of a player are all the players with whom she is either directly or indirectly connected (a connection is indicated by a line in Figure 1).

In each round, players participate in a social dilemma game. Decisions and earnings are denoted in points. Each player i receives an endowment of 50 points at the beginning of each round. All players decide simultaneously how much of the endowment, $x_i \in [0,50]$, they want to invest in a common fund. The returns that player i receives from the common fund are denoted by C_i. This gives the following payoff for player i in each period:

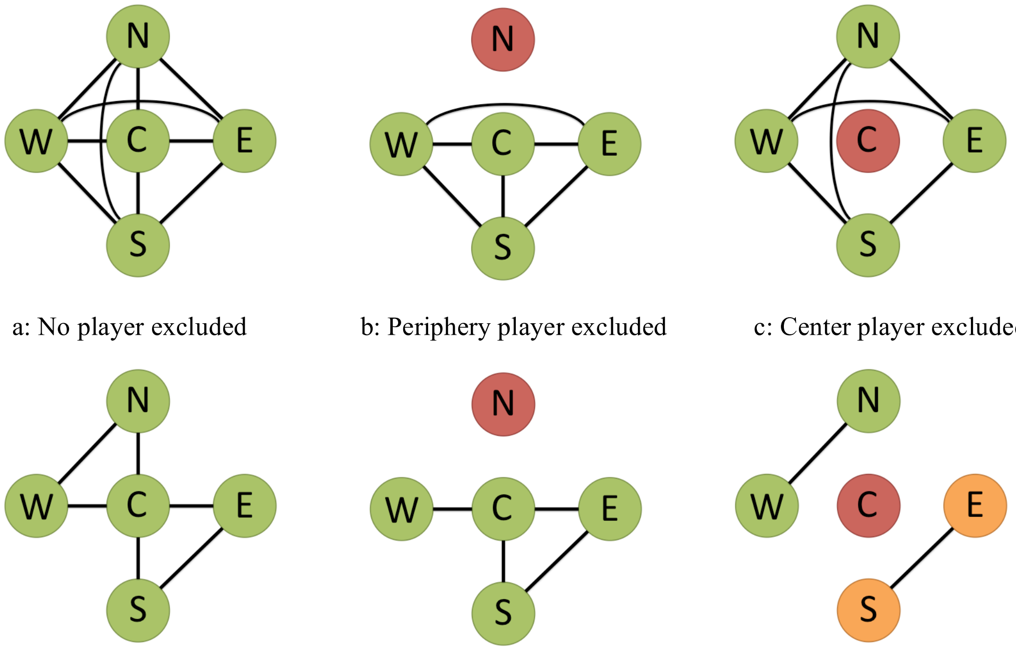
$$\Pi_i = 50 - x_i + C_i.$$

Authority

In all No Authority treatments (*nAnO*, *nAnC* and *nAC*) the returns from the common fund are given by a standard VCM mechanism, with an MPCR of 0.6. This gives:

$$C_i = (0.6)(x_i + \textstyle\sum_{j \in N_i} x_j),$$

Figure 1: Network Structures



a: No player excluded b: Periphery player excluded c: Center player excluded

d: No player excluded e: Periphery player excluded f: Center player excluded

Notes. The top panel shows the network structure for the cases without power from centrality and the lower panel when there is centrality. The three cases per panel distinguish between the network comprising of all players, the network without N and the network without C.

where N_i denotes the set of neighbors of i. Subjects decide simultaneously and the contributions of all players in the group are publicly announced at the end of each round.

In the Authority treatments (*AnO*, *AnC* and *AC*) we depart from the standard setting by allowing the center player to decide how to allocate the returns from the common fund. This player receives all contributions to the common fund from the periphery players. These contributions, and her own are multiplied by a factor $\alpha = 0.6 \cdot (|N_C| + 1)$, i.e. if no player is excluded the contributions are multiplied by a factor $\alpha = 3$. The value of the common fund is then $F = \alpha\,(x_C + \sum_{j \in N_C} x_j)$.[4] The center player determines the division of the common fund within the group. In particular, she decides how much, R, of the common fund, to return to the periphery players, which is then divided equally among them. Thus the returns from the common fund are $C_C = F - R$ for the center and $C_P = R/|N_C|$ for each periphery player. Note that the center player cannot discriminate among periphery players. Subjects thus play a trust game with four senders (the periphery players) and one receiver (the center player, who

[4] The specification of α ensures that the maximal sum of payoffs is the same in all treatments. Note that the VCM without authority is captured by this specification: the group account is split equally and each player receives $\frac{3}{|N_C|+1}\sum_j x_j$, which is $0.6\sum_j x_j$ for full networks, i.e., the MPCR is 0.6.

can also ‘send’). Each period consists of two decisions. First, all players decide simultaneously on their contribution x_i and all contributions are made public information. Second, the center player decides on R, and payoffs are realized and made public.

Centrality

To create power from centrality in this framework (*nAC* and *AC*), we allow players to be excluded from further participation in the network, in a manner explained below. If a player is excluded from the network, the non-excluded players can only continue with players with whom they remain connected. In the complete networks of the type shown in the top panel of Figure 1, those not excluded remain connected in groups of four, irrespective of whether a periphery player (1b) or the center player (1c) has been excluded.

This is different in the centrality case depicted in the lower panel. Power from centrality occurs because one player (C) is necessary to keep others connected to each other. The general idea is that there are two pairs of periphery players (N&W and S&E) that are connected only via the center player C. In a network like 1d, the consequences of removing a player depend crucially on her position. Consider first the situation illustrated in Figure 1e, where player N has been excluded. All of her connections have been removed. Here, the remaining four players remain connected. Figure 1f illustrates the case when the center has been excluded. This creates two separate groups of two players each (N&W and S&E), as all of C’s connections have been removed. This is the way in which power from centrality is operationalized in our design. The center player has more of this power than the periphery players. When C is excluded, the periphery players not only lose her future contributions (this also occurs when a periphery player is excluded), but also of two other periphery players. This makes it more costly to exclude the center player in our centrality treatments.

Ostracism

Our third treatment variable concerns the possibility for players to exclude others from further interaction with the group, i.e., to create the networks in Figures 1b or 1c (1e or 1f) from the five-player networks in 1a (1d). When there is no ostracism, all five players participate in each of the five repetitions of the stage game. With ostracism, each of the first four rounds ends with a voting stage that can lead to one or more players being banned from the group.[5] As explained above, ostracism allows us to create power from centrality.

After having observed others' decisions in the social dilemma played in the current round, each player in a group may cast votes to exclude any number of players she is still linked with. She may cast at most one vote for each player.[6] If a player receives half or more of the possible number of votes, she is excluded until the end of the five rounds and can no longer participate in the remaining rounds. Hence, in a group of five, a player is excluded if she receives three or more votes, and in a group of four two votes suffice to be excluded. Voting is costless and anonymous in the sense that only the number of votes cast by each subject is announced but not at whom the votes were directed. Moreover, it is announced which players have been excluded (if any).

Excluded players receive a fixed sum equal to the endowment in the stage games for each of the remaining rounds, but can no longer interact with the group in any way. Hence an excluded player can no longer contribute nor will she receive any benefits from the common fund. Also, excluded players are not allowed to vote for exclusion. In case a player has been excluded, non-excluded players can only vote to exclude one or more of their remaining neighbors. For the remaining players, the consequences of a player being excluded depend both on the treatment and on the player concerned. In networks without centrality or authority the remaining players play the (four-person) VCM.[7] With authority, but without centrality, the remaining players play a (four-person) trust game if a periphery player has been excluded and a four-person VCM if the central player has been excluded. When there is power from centrality but not authority, players play a four-player VCM after exclusion of a periphery player (Figure 1e) and two-player VCMs if the central player has been excluded (Figure 1f). Finally, with centrality and authority, exclusion of a periphery player (1e) yields a four-person trust game and exclusion of the central player (1f) gives two-player VCM games.

Procedures

The computerized experiment was run in the CREED laboratory of the University of Amsterdam. In total 475 student subjects from a variety of academic disciplines participated, each in one session. For each treatment, we ran 3 sessions, each with 20, 25 or 30 subjects.

[5] There is no voting stage in the final period. Note that the fact that more than one player can be excluded means that the networks in Figure 1 do not cover all possible cases. Other cases are straightforward.
[6] We allow players the option to vote to exclude themselves as well. This was observed in only 16 out of 6,100 instances of voting.
[7] We discuss here the situation after the first exclusion. The cases for subsequent exclusions follow straightforwardly.

Upon entering the laboratory, communication between subjects was restricted and subjects were randomly assigned separate cubicles. Subjects received on-screen instructions, which they could read at their own pace. After reading the instructions, all participants were requested to answer several test questions.[8] The experiment started only after all subjects had answered the test questions correctly. Each session lasted approximately one hour.

Within a session, subjects played five blocks of one of the five-round games described above. To reduce noise and avoid behavioral spill-overs, subjects' roles were kept fixed throughout a session; a subject was randomly assigned either the center or a periphery role at the beginning of a session. In all treatments, subjects' contributions in a round were identified by their position in the network, i.e., North, South, etc. However, to minimize the possibility of long-term reputation formation and thus revenge motives, subjects were randomly re-matched into different groups after each block of five rounds. For this re-matching, we used two matching groups with either 10 or 15 subjects each, depending on the number of participants in the session. After the experiment, subjects were requested to fill out a short questionnaire.

At the end of each session, one block was randomly selected and subjects were paid their earnings from all rounds within this block. Earnings in points were converted to cash using an exchange rate of 60 points to one euro. Subject earned between 10.60 and 30.70 euro, with an average of 15.90 euro, including a show-up fee of 7 euro.

3 Hypotheses

We make the 'standard' assumption that players' preferences are fixed; in this case this means that preferences are not affected by having power. The stage-game equilibria of the social dilemma games are then straightforward for the case where individuals are self-interested. In the VCM, nobody will contribute to the common fund. In the trust game, the only player that will contribute to the common fund is C, who (as long as she is connected to some other player) will contribute her entire endowment and keep all the proceeds for herself. As a consequence, power is not predicted to have any effect; the only treatment effect predicted in the self-interested case is that a player with authority will contribute more to the common fund than players without, though she will not share it. This case of self-interested preferences serves as a benchmark yielding the null hypotheses for our statistical tests.

[8] Summaries of the experimental instructions are provided in Appendix A. Full instructions and the test questions are available in an online appendix (www.borisvanleeuwen.nl/ACinstructionsfull.pdf).

In the experiment, we implement a (finitely) repeated game and we add ostracism in some treatments. If players are assumed to be self-interested and believe that all others are too, there are no repeated game equilibria with positive contributions in the VCM games.[9] This is different in the Authority treatments with ostracism. Here, the threat of exclusion does affect players with authority as the stage-game Nash payoff is strictly higher than the exclusion payoff for players with authority. This opens the door for repeated game equilibria where the center player shares some of her surplus and where periphery players make positive contributions. In any case, repeated play of the stage-game equilibrium is still an equilibrium of the repeated game.

If we assume that (some) players have social preferences or if (some) players believe that some fraction of the population is willing to ostracize free riders, cooperative repeated game equilibria may exist in all treatments. Moreover, in this case, treatment effects may be predicted in equilibrium. However, this leads to a plethora of equilibria, depending on the specific assumptions. Instead of deriving all equilibria and searching for refinements, we will employ a simple setting with cooperative types (similar to Kreps et al. 1982) to derive comparative statics for the effects of ostracism, centrality and authority.

Before doing so, we define what we mean by 'cooperation' in our settings. To measure cooperation at the individual level, we consider how a player's choice affects the payoffs of others. For a periphery player and for a central player without authority, this simply means that the measure should be perfectly (linearly) correlated with the contribution to the common fund. For a player with authority, cooperation entails more than just contributing to the common fund. Since she will contribute her complete endowment even if selfish, she will create more surplus than the periphery players whenever the latter contribute less than their endowments. In addition, a cooperative type must also distribute the surplus raised in a 'cooperative' manner.

We denote 'individual cooperativeness' by γ_i defined as:

(1) $$\gamma_i \equiv \rho_i x_i/50\ , i = C, P$$

where x_i denotes i's contribution level. For players without authority, we adopt the convention that $\rho_i = 1$. For a player with authority, we define a fully cooperative distribution, R_{equal}, as one that divides the surplus generated by the group equally among *all*

[9] In some settings, costless voting could lead to repeated game equilibria with positive contributions (Hirshleifer & Rasmusen, 1989). However, in our study, excluded players still earn their endowment (which equals the Nash stage-game payoff). For this reason, there exists no subgame perfect equilibrium with positive contributions if all agents are payoff-maximizers. More information is available upon request.

players. Our benchmark for cooperative behavior is thus the VCM where this happens by default. Part of our measure of cooperation for players with authority, then, is the fraction of this fully cooperative distribution they implement. To quantify, we define:

$$\rho_C \equiv R/R_{equal}.$$

Note from (1) that – like any player – a player with authority acts uncooperatively if she contributes less than 100% of her endowment (which in her case also does not serve her self-interest).[10] She also acts uncooperatively if she returns an 'uncooperative' amount. In our experiments, almost all subjects with authority contributed their entire endowment. Their level of cooperation is then determined almost exclusively by ρ_C, the extent to which they share the gains from cooperation with the periphery.

We now turn to a formulation of our alternative hypotheses. As mentioned above we consider a simple setting where players are either self-interested or of a 'cooperative type'. We assume that all players act rationally, conditional on their preferences. Cooperative types unconditionally play in a way that achieves $\gamma = 1$ in all periods.[11] If there is the possibility of exclusion, they will vote to exclude any player that does not act as a cooperative type in this way (i.e., any player j with $\gamma_j < 1$). Let p_{ij} be the prior probability that player i assigns to j being a cooperative type and $1 - p_{ij}$ the prior probability that that player i assigns to j being self-interested. Given that interactions are anonymous and positions are randomly allocated, we will simply assume $p_{ij} = p, \forall\ i, j, i\ \neq j$. Note that the model nests the standard case where all players are rational and self interested by setting $p = 0$.

For each of the treatment variables, we will use this setup to derive alternative hypotheses. We do so by examining whether players would prefer to (i) exclude others who were observed to free-ride ($\gamma = 0$) or cooperate ($\gamma = 1$) in the penultimate cooperation stage, and using this, whether (ii) self-interested players will mimic cooperative types in the penultimate cooperation stage. This allows us to derive the relevant comparative statics. We base the direction of the comparative statics on the level of p that is needed to sustain cooperation; we expect higher levels of cooperation in situations where a lower p is required to sustain cooperation.

---

[10] An exception is the case where she returns more than R_{equal}. This yields ρ_C>1 and the possibility that γ_i >1 even if she does not contribute her complete endowment. This scenario where ρ_C>1 was only observed in 4% of the cases.

[11] The assumption of unconditional cooperation for this type is, admittedly, limiting. But, we see our approach as applying a 'toy model' aimed at deriving plausible comparative static predictions.

In all treatments, in the final round, self-interested players will simply play according to the stage-game Nash equilibrium: players without authority will not contribute anything and players with authority will contribute their complete endowment and keep the entire group account. In both cases $\gamma = 0$.

Before considering the effects of power, we consider groups where no one has power ($nAnO$ and $nAnC$). Without the possibility of ostracism, self-interested players will not cooperate in the penultimate round, as there is no mechanism through which cooperative behavior can be enforced. This is independent of the proportion of cooperative types p. This changes with ostracism. Consider a strategy profile where players that do not fully cooperate, i.e., players with $\gamma < 1$, are excluded and others are not. A self-interested player is willing to sustain such a strategy profile as any player with $\gamma < 1$ cannot be a cooperative type and will thus be uncooperative ($\gamma = 0$) in the final period. Excluding such a player will not affect i's payoffs. On the other hand, excluding a player who acted cooperatively ($\gamma = 1$) is costly in expectation if $p > 0$. Similarly, being excluded from the final period is costly when $p > 0$. For complete groups, payoffs for a self-interested player are $50 + 0.6(4 \cdot 50p) = 50 + 120p$ if she has not been excluded and simply 50 if she has been excluded. Given that all other players are either a cooperative type or mimicking a cooperative type, for a self-interested player, mimicking a cooperative type in the penultimate round is a best response if:

$$\big(0.6(5 \cdot 50)\big) + (50 + 120p) > \big(50 + 0.6(4 \cdot 50)\big) + 50,$$

i.e. when $p > \frac{1}{6}$. Hence, a small proportion of cooperative types suffices to sustain cooperation under the threat of exclusion. This leads to our first hypothesis:

Hypothesis 1 (positive effect of ostracism):

γ_C and γ_P are both higher in nAnC than in nAnO.

This is in line with previous evidence that the ability to exclude players from the group raises cooperation (Cinyabuguma et al., 2005). As there are no power asymmetries in *nAnO* and *nAnC*, we expect that both roles (center and periphery players) will be equally (un)cooperative.

Now consider the effects of introducing authority. In the absence of ostracism (*AnO*) cooperation cannot be sustained in equilibrium, regardless of the value of p. Hence, we expect that the introduction of authority will not affect cooperativeness in the absence of ostracism. This changes with ostracism. First, consider the center player. Excluding an uncooperative center player does not come at a cost, but will in fact lead to *higher* payoffs in the final period when $p > 0$. This is because a self-interested center player will keep all the proceeds of the common fund to herself. If the center player acts uncooperatively in the penultimate period, then this reveals that she is self-interested. Excluding the center will then be beneficial, because by doing so, a 4-player VCM will be played in the final period and one might benefit from the presence of cooperative types amongst the other periphery players. If the center player acts cooperatively in the penultimate period, she could be a cooperative type or self-interested. Periphery players may wish to exclude her, again yielding a 4-player VCM in the final period. The expected payoffs for a self-interested player in this 4-player game are $50 + 0.6(3 \cdot 50p) = 50 + 90p$. Leaving the center in the group means that the self-interested player will only benefit from (others') contributions to the group fund if the center player is cooperative (which occurs with probability p). This gives an expected payoff in the final period of $50 + p\big(0.6(50 + 3 \cdot 50p)\big) = 50 + 30p + 90p^2$. Hence, self-interested periphery players will prefer to exclude a center player with authority who cooperated in the penultimate period when $p < \frac{2}{3}$ (which we assume from here onwards). Anticipating that she will be excluded anyway, a self-interested center player will be uncooperative in the penultimate period. Hence, in the presence of ostracism, authority will have a negative effect on the average cooperativeness of center players.

Now consider the periphery players. Excluding an uncooperative periphery player does not come at a cost, while excluding a cooperative periphery player is costly if $p > 0$. Hence, uncooperative periphery players may be excluded in equilibrium but cooperative periphery players may not. Given that a self-interested center will be uncooperative in the penultimate round (and that the center will be excluded in any case), a periphery player will mimic a cooperative type when:

$$p\big(0.6(5 \cdot 50)\big) + (50 + 90p) > \Big(50 + p\big(0.6(4 \cdot 50)\big)\Big) + 50,$$

which holds for $p > \frac{5}{12}$. Recall that without authority, mimicking the cooperative type is a best response for $p > \frac{1}{6}$. Hence, under the threat of exclusion periphery players require a greater proportion of cooperative types in order to sustain cooperation when authority is introduced.

Hypothesis 2 (effects of authority):

(a) *In the absence of ostracism, authority does not affect cooperativeness, i.e.* γ_C *and* γ_P *are the same in AnO and nAnO.*

(b) *In the presence of ostracism, authority lowers cooperativeness of both center and periphery players, i.e.* γ_C *and* γ_P *are lower in AnC than in nAnC, but center players are more likely to be affected, i.e.* $\gamma_C < \gamma_P$ *in AnC.*

We now move to the effects of centrality on cooperativeness within groups. Recall that ostracism is a necessary condition for center players to derive power from centrality. Thus, we compare power from centrality to the situation where groups can ostracize members but there is no power from centrality (*nAnC*). Excluding an uncooperative player without centrality (any player in *nAnC* or a periphery player in *nAC*) will not affect expected payoffs in the final round. Excluding an uncooperative central player comes at a cost though: one can no longer benefit from the contributions by cooperative types who were connected via C. Hence, self-interested periphery players will not be willing to exclude an uncooperative player with centrality for *any* $p > 0$. This reduces the chances that a central player acting uncooperatively will be excluded, allowing her to 'abuse' her power. Note that centrality does not change the effect a threat of exclusion has on periphery players.

Hypothesis 3 (effects of centrality):

(a) $\gamma_C < \gamma_P$ *in nAC.*

(b) γ_C *is lower in nAC than in nAnC.*

Finally, we consider the situation when a player derives power from two sources, i.e., from both authority *and* centrality (*AC*). Since ostracism is necessary for centrality, we focus on situations that allow ostracism. As described above, with only one source of power we expect powerful players to take advantage of their position, both for power from authority and power from centrality. In *AnC*, periphery players may even exclude center players when they act cooperatively. This is because they have an attractive alternative: a 4-player VCM. In *AC*, this alternative is a 2-player VCM, which is less attractive than the 4-player version. In *AC*, periphery players will only exclude a center player acting cooperatively in the penultimate round when

$$50 + 0.6(1 \cdot 50p) > 50 + p\big(0.6(50 + 3 \cdot 50p)\big),$$

which never holds in the presence of cooperative types ($p > 0$). Excluding an uncooperative center does lead to higher expected payoffs for periphery players (if $p > 0$). Now, a self-interested center player will mimic a cooperative type if:

$$0.6(5 \cdot 50) + 3(50 + 4 \cdot 50p) > 3(5 \cdot 50) + 50,$$

which holds if $p > \frac{5}{6}$. This is the same constraint as in *AnC*. Hence, we do not expect that the cooperativeness of center players is affected by the addition of centrality to authority. The other way around, we expect that adding authority to centrality has a positive effect on the cooperativeness of center players. This is because (as argued above) uncooperative center players face no threat of exclusion when there is no authority. When adding authority, the threat of being ostracized provides a reason to be more cooperative.

Hypothesis 4 (effects of authority and centrality):

(a) γ_C *is larger in AC than in nAC.*

(b) γ_C *is not different in AC and in AnC.*

The key to increased cooperation is thus punishment (here, votes to ostracize) specifically targeted at those with low cooperation levels.[12] In our ostracism treatments, we expect lower cooperativeness to be associated with higher rates of ostracism. In addition, we explore whether the distinct roles (center or periphery) are targeted for ostracism with different intensities. If there is no discrimination by voters between roles, targeting would imply (negative) correlations between cooperation and exclusion rates that are similar across network positions. Conditional on cooperativeness, if players in a particular role face a higher (lower) ostracism rate, we say that the role faces negative (positive) discrimination. We hypothesized above that players with power from centrality will abuse their power. In light of the high cost of ostracizing such players, we expect that they will face positive discrimination. In contrast, center players in *AnC* are expected to be ostracized even when they are cooperative.

Hypothesis 5 (exclusion):

(a) *Players with higher cooperativeness receive fewer votes for exclusion.*

[12] Such threats of retaliation have been seen to work in other disciplining mechanisms as well. For instance, see Fehr and Gächter (2000) and Sefton et al. (2007) on peer punishment in public goods games.

(b) *Conditional on cooperativeness, center players face positive discrimination when they have power from centrality, i.e., they receive fewer votes for exclusion than periphery players with similar levels of cooperativeness.*

(c) *Conditional on cooperativeness, center players face negative discrimination when they have power from authority, i.e., they receive more votes for exclusion than periphery players with similar levels of cooperativeness.*

These hypotheses will be tested against a null of no effect of cooperation on votes to exclude and no discrimination, respectively.

The above hypotheses focus on the individual and take into account the impact of a player's choices on the distribution of surplus, especially in the trust game. To examine the impact of power on group-level performance, we construct a measure of cooperation at the group level, which we call the '*net efficiency gain*', ϕ^n:

(2) $$\phi^n \equiv \frac{(realized\ surplus - Nash\ surplus)}{(750 - Nash\ surplus)},$$

where 'Nash surplus' is the surplus in the Nash equilibrium for self-interested players and 750 is the maximum aggregate surplus in a round. Note that for the VCM, ϕ^ncorrelates perfectly with aggregate contributions. For the Trust game, it corrects for the fact that even self-interested players with authority contribute in the Nash equilibrium. Finally (2) only measures the *gain* in efficiency and says nothing about how the surplus is distributed (which is especially relevant for the Trust game). As with individual cooperation, we expect that ostracism will raise group-level efficiency. Excluding players (especially those with authority) will lower efficiency, however. Based on hypothesis 2, we then expect, ceteris paribus, authority to reduce efficiency in the presence of ostracism. Similarly, based on hypothesis 3, we expect centrality to reduce efficiency.

Hypothesis 6 (efficiency):

(a) ϕ^n*is higher when ostracism is available; i.e. higher in nAnC than in nAnO* and higher in *AnC than in AnO.*

(b) *In the presence of ostracism, the addition of power from authority decreases* ϕ^n*; i.e.,* ϕ^n*is lower in AnC than in nAnC.*

(c) *The addition of power from centrality decreases* ϕ^n*; i.e.,* ϕ^n *is lower in nAC than in nAnC.*

We will test hypotheses 1-6 in the next section.

4 Results

We start with a general overview of our results. This is followed by separate discussions of the effects of the two power types, starting with the absence of power and finishing with the combination of the two types. We then provide discussions of voting (to ostracize) behavior and our efficiency results. First, Figure 2 presents the mean cooperativeness (γ) in our various treatments.[13] For this overview we combine data from all blocks.[14] The figure shows that cooperation is higher with ostracism than without and that differences in cooperation between the center and the periphery occur only when there is power of either type (but not when both types of power appear together). We will investigate the differences across treatments in more detail in the following subsections and test the hypotheses derived above. Unless stated otherwise, all reported statistics come from two-sided tests using matching groups as the unit of analysis, and averages over all rounds and blocks.[15] The number of observations is thus six in each comparison sample.

4.1 No Power - The effects of ostracism

Figure 2 shows the usual pattern of declining cooperation over time in the standard public goods game (*nAnO*) in the absence of any power and ostracism (see, for instance, Fehr and Gächter, 2000). In the absence of power, the only difference between player positions is one of framing, where the center player is presented as being in the 'middle' of the group. As predicted, this has no effect; there is no difference in cooperativeness between the center and periphery players. A Wilcoxon signed ranks test (henceforth, W) shows no significant difference in γ between the center and periphery players in *nAnO* (W, $p = 0.463$).

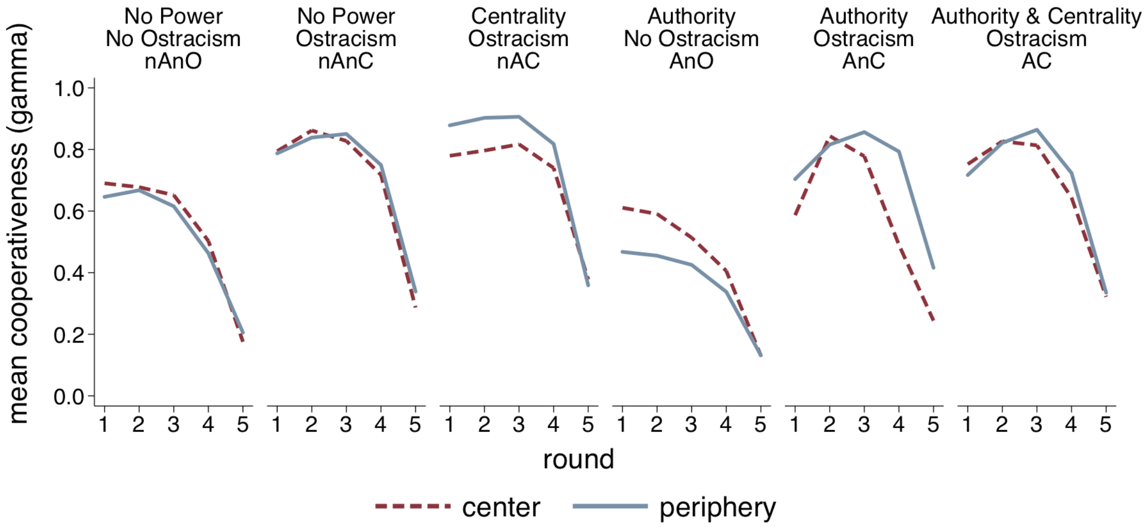
Once we allow for ostracism (*nAnC*), we observe that cooperation increases for both center and periphery players. This pattern replicates findings in Cinyabuguma et al. (2005),

[13] For completeness, the corresponding figure for contribution levels (as opposed to γ) is presented in Appendix B. As expected, players with authority contribute almost 100% of their endowments in all rounds. Appendix B also presents the return rates ρ for players with authority power and a table with the mean cooperation rates by treatment and position.

[14] We do not observe any differences in cooperation between blocks, except in *nAnO*, where we observe a downward trend over blocks.

[15] When we exclude the last round of each block, all results remain qualitatively the same

Figure 2. Mean cooperativeness (γ) by player position



Notes. The three panels on the left denote the treatments without authority and the three panels on the right show the cases with authority. The horizontal axis shows the five rounds. For each round, cooperativeness is averaged across five blocks. The graphs are based on decisions by individuals who have not been excluded from their group and who are not isolated.

an early work documenting the finding that ostracism raises cooperativeness in groups.[16] As expected, there are again no differences between the roles (W, p = 0.463). Mann-Whitney tests (MW) confirm that cooperativeness increases with ostracism for both center (MW, p = 0.078) and periphery (MW, p = 0.016) players. As a consequence, average cooperativeness in the group as a whole is higher (MW, p = 0.010). This gives our first result summarized below.

Result 1 (effects of ostracism): *In the absence of power, the opportunity to ostracize group members raises the cooperativeness of both center and periphery players.*

This supports hypothesis 1.

4.2 The effects of power from authority

The three panels on the right half of Figure 2 show, first, that in the absence of ostracism (*AnO*) the cooperativeness of periphery players is reduced relative to when no one has power (MW, p = 0.055) while that of center players remains unaffected (MW, p = 0.150). Taken

[16] Note the end effect exists even with ostracism; there is a sharp drop in cooperativeness in round 5. This can be attributed to the fact that there are no opportunities to ostracize group members in the final round.

together, average cooperativeness in the group is lower as well (MW, p = 0.037). As before, we observe a steady decline over time. Further, authority drives a wedge between the cooperation levels of the center and periphery players – center players are significantly more cooperative than are periphery players (W, p = 0.028).[17]

When ostracism is allowed in the presence of authority (*AnC*), cooperativeness increases unambiguously for periphery players compared to the case of no ostracism (MW, p = 0.004). While cooperativeness initially increases for center players, the overall effect across five rounds is not significant (MW, p = 1.000).[18] Overall, ostracism does raise average cooperativeness in the group as a whole (MW, p = 0.016). Hence, the threat of exclusion promotes cooperation in the presence of authority as well. However, across all rounds center players are *less* cooperative than periphery players (W, p = 0.046). A comparison of the left and right panels of Figure 2 shows that, when ostracism is available, the addition of authority (*nAnC vs. AnC*) does not affect the cooperativeness of periphery players while that of powerful players is reduced (MW, p = 0.078 for center, p = 0.749 for periphery). The latter effect seems to be due mostly to lower cooperativeness of players with authority in later rounds.

Result 2 (effects of authority):

(a) *In the absence of ostracism, authority does not affect the cooperativeness of center players. However, it decreases that of periphery players. Players with authority are more cooperative than those without power.*

(b) *In the presence of ostracism, authority lowers the cooperativeness of center players. The cooperativeness of periphery players is unaffected. Players with authority are less cooperative than those without.*

Result 2(a) contradicts hypothesis 2(a), which predicts no effect of cooperativeness with authority in the absence of ostracism. Result 2(b) partly supports hypothesis 2(b), which predicts that cooperativeness decreases for both roles. All in all, we find mixed support for hypothesis 2. We will discuss possible reasons in the general discussion of our results in section 5.

[17] The periphery's reaction (reduced cooperation) is reminiscent of the Falk and Kosfeld (2006) finding that subjects dislike being controlled.

[18] Results on the cooperativeness of players with authority in the presence of ostracism may be affected by the fact that there are few left in the group after the first round. We document this later when we discuss ostracism patterns.

4.3 The effects of power from centrality

Figure 2 shows that, in the presence of centrality alone (*nAC*), there is a difference between cooperation levels of the center and periphery players. Center players seem to 'abuse' their power; center players are 7.7 percentage points less cooperative than periphery players and this is statistically significant (W, $p = 0.028$). Compared to a situation in which no player has any power (*nAnC*), Figure 2 suggests that the cooperativeness of players with power from centrality is not very different while that of periphery players is higher. Tests, however, show that cooperativeness is not significantly different for center (MW, $p = 0.873$) or for periphery (MW, $p = 0.262$) players.[19] Hence, though the partial effects in either player type are not significant, the effects of introducing power through centrality are large enough to make the difference between the types statistically significant.[20]

Result 3 (effects of centrality):

(a) *Players who derive power from centrality are less cooperative than players with no power.*

(b) *Cooperation levels of both center and periphery players are not affected by power from centrality.*

Result 3(a) supports hypothesis 3(a). Result 3(b) implies that, for the center player, this does not allow us to reject the null in favor of the alternative in hypothesis 3(b).

4.4 The effects of power from authority *and* centrality

For environments with ostracism, we compare cooperativeness in *AC* with situations where there is only one source of power. When the powerful player is only central (*nAC*), Figure 2 suggests that central players are as cooperative as in *AC* while periphery players are more cooperative than in *AC*. This is confirmed by statistical tests (MW, $p = 0.423$ for center and $p = 0.078$ for periphery players). On the other hand, when the powerful player has only authority (*AnC*), Figure 2 suggests that central players are less cooperative than in *AC* while periphery players are as cooperative as in *AC*. However, tests show no significant differences

[19] Compared to a situation of no power *and* no ostracism (*nAnO*), cooperation levels are higher for both center (MW, $p = 0.025$) and periphery (MW, $p = 0.004$) players. This reiterates Result 1 and the finding in Cinyabuguma et al. (2005) that the ability to ostracize group members successfully raises cooperation in groups.

[20] The differences in cooperativeness between periphery and center players are also significantly larger in *nAC* than in *nAnC* (MW, $p = 0.016$).

(MW, p = 0.150 for center and p = 1.000 for periphery players). Finally, Figure 2 shows no discernible difference in cooperativeness between center and periphery players in *AC*. This observation seems statistically supported (W, p = 0.753). Thus, we do not find that powerful players take advantage of their power when power is derived from both sources.

Result 4 (effects of authority and centrality):

(a) *Adding authority to centrality does not change the cooperativeness of already powerful players. However, it decreases the cooperativeness of players without power.*

(b) *Adding centrality to authority does not change the cooperativeness of either powerful players or periphery players.*

Result 4(a) contradicts hypothesis 4(a), which predicts a positive effect for the center player. Results 4(b) supports hypothesis 4(b).

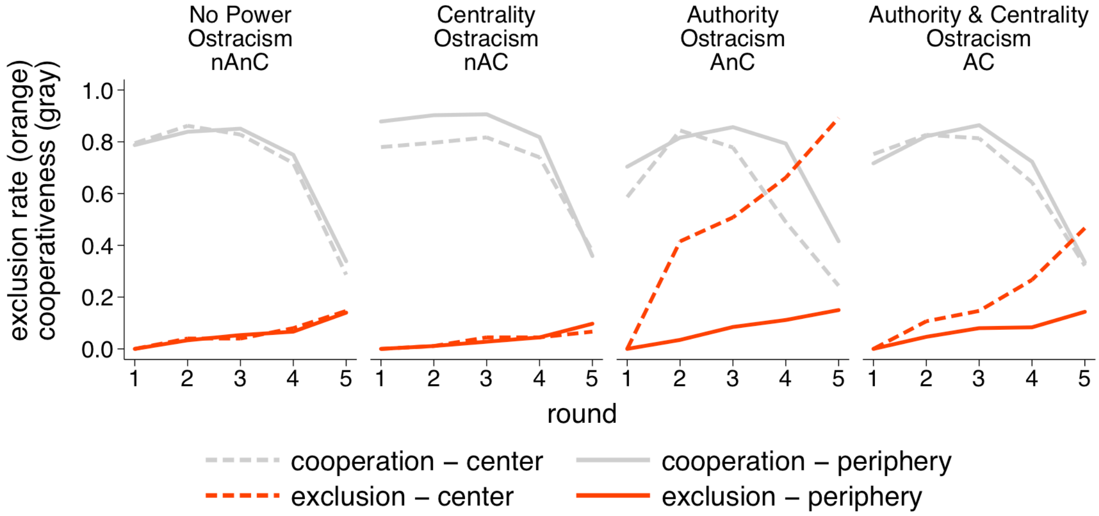
Our setup also allows us to directly compare the two types of power. Compared to when powerful players have only authority (*AnC*), the cooperativeness of both center and periphery players appear somewhat higher when powerful players are only central (*nAC*). However, the difference is only (weakly) significant for center players (MW, p = 0.078) and not significant for periphery players (MW, p = 0.109).

4.5 Voting behavior and ostracism

Figure 3 presents the *cumulative* proportion of center and periphery players excluded over rounds in each of our four ostracism treatments. To facilitate seeing the correlation between cooperativeness and ostracism, the figure also shows cooperativeness. Observe first that when no one has any power (*nAnC*) ostracism rates rise with time for both player roles and remain under 20% to the end. Further, there is no discrimination; there is no discernible difference in cooperativeness or in the rate of ostracism between center and periphery players. When powerful players are central (*nAC*), they are less cooperative than are periphery players. While the ostracism rate rises over time for both roles, it appears that central players face positive discrimination; though their cooperation level is lower, central players are not more likely to be ostracized than are periphery players.

When center players have only authority (*AnC*), the picture shows a stark difference. While cumulative ostracism rates do increase over time for both roles, they differ wildly between the roles. In the first 2-3 rounds, there is hardly any difference in cooperativeness

Figure 3. Exclusion and cooperativeness by player position



Notes: Cumulative exclusion rates and cooperativeness for center and periphery players. Exclusion rates reflect the proportion of players that are on average excluded in the respective round. Means are taken across all blocks.

between the roles. However, powerful players are *much* more likely to be ostracized than are periphery players. At the end of round 1, slightly over 40% of center players are excluded while less than 10% of periphery players are excluded. This difference is only magnified over time; by the end of round 4, over 90% of center players have been excluded while less than 20% of periphery players have been excluded. This difference seems too large to be explained by distinct levels of cooperativeness. It appears that center players with power from authority face severe negative discrimination from periphery players.

When powerful players derive power from both authority and centrality (*AC*), there is no significant difference in cooperativeness between roles. Nevertheless, as in *AnC*, center players face negative discrimination. They are more likely to be ostracized than are periphery players. Once again, this is evident from the very beginning and magnified over time. By the last round about 50% of the center players have been excluded while this holds for less than 20% of the periphery players.

We test for the patterns noted above using individual-level regressions where we explain the number of votes for exclusion received by individuals (in rounds 1 - 4). For each treatment, we estimate a separate Poisson regression where the dependent variable is the number of votes for exclusion received (0–5) in a round. The independent variables include the individual's cooperativeness (γ) in the round, a dummy for being the center player, the interaction between the two, the number of members in the individual's subgroup in the

Table 2. Votes received for exclusion

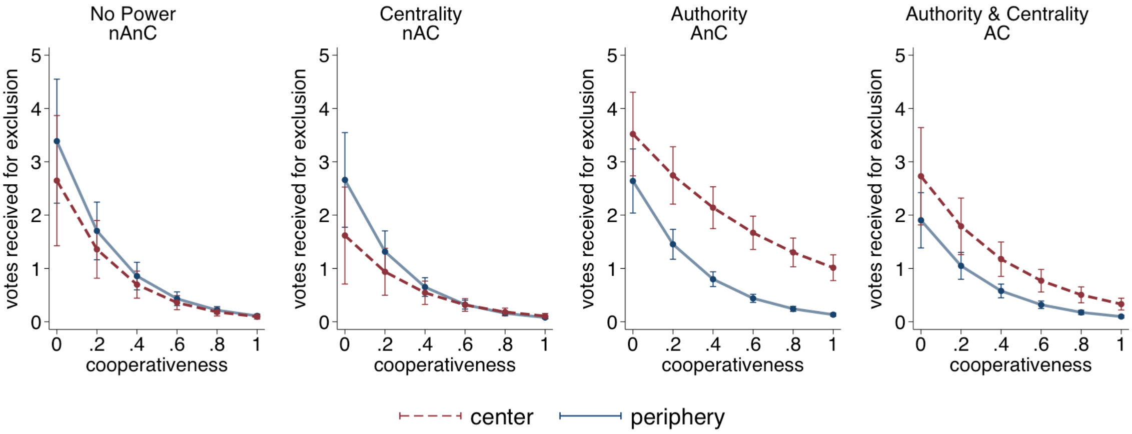
	No Power (*nAnC*)	Centrality (*nAC*)	Authority (*AnC*)	Authority and Centrality (*AC*)
Cooperativeness	-3.436***	-3.522***	-2.989***	-2.980***
	(0.161)	(0.171)	(0.156)	(0.164)
Center	-0.247	-0.498*	0.288**	0.361**
	(0.201)	(0.297)	(0.128)	(0.179)
Center × Cooperativeness	0.0987	0.797**	1.744***	0.870***
	(0.326)	(0.386)	(0.208)	(0.252)
Number of players in subgroup	0.165	0.526***	0.288***	0.543***
	(0.133)	(0.149)	(0.0860)	(0.0751)
Round	-0.166***	0.114**	0.0276	0.0444
	(0.0427)	(0.0447)	(0.0399)	(0.0403)
Block	0.0945***	0.0324	0.0478*	0.00941
	(0.0322)	(0.0347)	(0.0281)	(0.0311)
Constant	0.515	-2.000**	-0.556	-2.088***
	(0.720)	(0.778)	(0.480)	(0.416)
Observations	1442	1758	1136	1385

Notes: Regression estimates from (mixed) random effects poisson regressions, where we allow for random effects at the matching group and the subject level. Dep. variable: Number of votes for exclusion received in the current round. Only non-excluded and non-isolated subjects are included in the regressions. Standard errors clustered on matching groups in parentheses. *, ** and *** indicate significance at 10%, 5% and 1% respectively.

round, a time trend for rounds and a time trend for blocks. For each individual, we only include rounds in which he/she had not yet been excluded and where she was not isolated. The regression estimates are presented in Table 2. Since the estimated coefficients of Poisson regressions are not readily interpretable, we plot the estimated votes received as a function of cooperativeness in Figure 4. Together, Table 2 and Figure 4 allow us to understand the determinants of exclusion votes and their (statistical) importance, and to compare these results across treatments.

The regressions provide statistical underpinning for the patterns described above. The first regression – for the case where no one has any power (*nAnC*) – shows that those with higher cooperation levels receive fewer votes; the coefficient of cooperativeness is negative and significant at the 1% level. Also, as expected, there is no discrimination between roles. The number of exclusion votes received decreases across rounds in a block and increases across blocks. Whereas the latter effect could indicate learning, the former is likely attributable to the fact that exclusion has less effect as the number of rounds remaining decreases. These time trends mostly disappear when we add power. In contrast, the strong

Figure 4. Estimated votes received for exclusion



Notes: Estimated number of votes received as function of cooperativeness for center and periphery players. Estimations based on the regressions reported in Table 2. Error bars indicate 95% confidence intervals.

negative effect of cooperativeness on the number of votes received is observed in all power configurations.

The second regression shows that players with centrality power receive fewer votes for exclusion, even after controlling for cooperativeness. Moreover, the interaction term is positive and significant. Figure 4 illustrates the net effect: for low levels of cooperativeness, players with centrality receive fewer votes for exclusion than periphery players. This result provides empirical support for the tolerated theft hypothesis referred to in the introduction. Note that it is not the case that center players are at an advantage irrespective of their choices. For relatively high levels of cooperativeness, center players receive similar number of votes as her periphery counterparts. The positive discrimination that we observe is completely attributable to tolerated theft.

In both treatments where players have authority (third and fourth regression), center players receive significantly *more* votes than periphery players. However, the interaction term is positive and significant. Figure 4 shows that players with power from authority (but without centrality, *AnC*), receive more votes for exclusion than periphery players, at *any* level of cooperativeness. Even when players with authority are fully cooperative, they can expect to receive a positive number of votes. This negative discrimination explains the divergence in ostracism rates between center and periphery players in the authority treatments evident in Figure 3. In *AC*, center players derive power from both centrality and authority. In this case, again center players receive more votes than periphery players, but the effect is smaller than in *AnC*. This suggests that centrality mitigates the negative discrimination center players face and explains why the divergence between ostracism rates for the two roles is smaller in *AC* than in *AnC*.

Result 5:

(a) *In all ostracism treatments, there is a negative relation between votes received and cooperativeness.*

(b) *Controlling for cooperativeness, players with power from centrality are ostracized less frequently than periphery players.*

(c) *Controlling for cooperativeness, players with power from authority are ostracized more frequently than periphery players.*

Results 5(a), 5(b) and 5(c) directly support hypotheses 5(a), 5(b) and 5(c) respectively.

4.6 Efficiency Comparisons

Table 3 presents summary statistics on group level net efficiency gain (NEG) in each treatment.[21] Note the close relationship with cooperativeness. There are two main differences: (i) cooperativeness by a player with power from authority includes distributional issues, which are absent in our efficiency measure and (ii) excluding players will lower efficiency.

In the absence of any kind of power, efficiency is higher when ostracism is available as a disciplining mechanism (*nAnC vs nAnO*; MW, p = 0.078). This is in spite of the fact that efficiency may be lost if subjects are excluded. The increased cooperation observed with ostracism thus dominates the cost of excluding players. This is different when players have power from authority: then, ostracism does not increase NEG despite the observed increase in cooperativeness (*AnO vs AnC*; MW, p = 1.000). This is because almost all center players are excluded at some point.

Table 3 also shows that NEG is lower when center players are given power from authority. When ostracism is unavailable, giving center players only authority reduces NEG (*AnO vs nAnO*; MW, p = 0.025). This is because, as observed above, periphery players cooperate significantly less in this environment. With ostracism, we again observe that adding authority reduces NEG, both with and without centrality (*AnC vs nAnC*; MW, p = 0.007 and *AC vs nAC*; MW, p = 0.055). Thus, authority is always detrimental to efficiency. As seen above, this arises through two channels. First, while authority does not affect the

[21] As defined in Section 3, ϕ^n measures the increase in earnings over and above the Nash equilibrium earnings. Summary statistics on absolute earnings themselves are presented in Appendix B.

Table 3. Net Efficiency Gain

		No Authority	Authority
No Ostracism (⇒ no Centrality)		*nAnO* 0.521 (0.075)	*AnO* 0.358 (0.112)
Ostracism	No Centrality	*nAnC* 0.636 (0.140)	*AnC* 0.378 (0.171)
	Centrality	*nAC* 0.705 (0.052)	*AC* 0.487 (0.195)

Notes. Cells give mean net efficiency gain with standard deviation in parentheses. The unit of analysis is the matching group.

cooperativeness of the center players, it reduces the cooperativeness of the periphery players. Second, when center players have power from authority, periphery players often ostracize them, and early in the game. Both factors reduce the surplus that groups (can) generate, thus reducing efficiency.

In contrast, centrality is not harmful for efficiency. Compared to a situation with no power (*nAnC*), the addition of centrality (*nAC*) somewhat increases NEG, though this increase is not statistically significant (MW, p = 0.150). Adding centrality to authority also raises NEG, but again not significantly so (*AC vs AnC*; MW, p = 0.262). Finally, *replacing* power from authority with power from centrality significantly increases NEG (*nAC vs AnC*; MW, p = 0.007).

Result 6:

(a) *In the absence of power, ostracism raises efficiency. In the presence of power from authority, ostracism does not affect efficiency.*

(b) *Power from authority lowers efficiency*.

(c) *Power from centrality does not affect efficiency.*

Result 6 provides mixed support for hypothesis 6. Result 6(a) partly supports hypothesis 6(a); ostracism raises efficiency in the absence of power (as hypothesized) but not in the presence of power from authority (where an increase was expected). Result 6(b) supports hypothesis 6(b), while result 6(c) provides no support for the corresponding hypothesis.

5 Discussion and conclusions

We study the impact of power asymmetry on cooperation in classic social dilemmas in a controlled laboratory experiment. We systematically vary power along two dimensions:

centrality and authority. We replicate standard results for social dilemmas without power heterogeneity, including the positive effect that ostracism has on cooperativeness. We add to this literature the general finding that both sources of power matter, albeit in different ways.

We expected players with authority to take advantage of their position. In the presence of (some) cooperative players, those with power are likely to be excluded, *even* when they behave cooperatively. In theory, players with authority expect this. They therefore take advantage of their position for as long as they can. We find support for only part of this hypothesis: the analysis of the voting data reveals that players with authority receive votes for exclusion. This occurs *even* when they behave cooperatively. In fact, in the absence of ostracism, players without authority cooperate even less than those with this power and in the presence of ostracism, they all cooperate equally in early rounds. Hence, it is not so much that theft of surplus is not tolerated for those with power from authority, but more that this kind of power is not tolerated at all.

As for power from centrality, in the treatments without authority, most of our hypotheses are supported. In the presence of (some) cooperative players, excluding central players is more costly than excluding periphery players. Indeed, we observe that players with power from centrality cooperate significantly less than those without power. It seems that our subjects realize that when one player has power from centrality, she enables higher returns for the rest of the group simply by being there. Excluding the central player immediately reduces the group to at most two players and thus drastically reduces the surplus that can possibly be generated. Our results on positive discrimination in terms of exclusion rates provide direct evidence that this 'theft' is 'tolerated' by the others. This is akin to the finding that in a democracy "... having the opportunity to throw the rascals out does not necessarily mean that such capacity will be exerted" (de Sousa and Moriconi, 2013, p. 474). Interestingly, this tolerated theft does no harm to the group as a whole: efficiency is not significantly different from the case without power.

Even more remarkable, perhaps, is the case when both types of power co-exist (*AC*). Here, it is more costly to ostracize the center player than when there is no power from centrality.[22] While center players are ostracized less frequently in *AC* than in *AnC*, they are still ostracized much more frequently than periphery players. Moreover, they are ostracized more frequently despite the fact that they are as cooperative as periphery players. This again suggests that center players with authority might simply be reviled in their groups and are

[22] In *AnC*, exclusion of the center player leaves the periphery in groups of four, where they participate in a VCM. In *AC*, it results in pairs of periphery players in a VCM.

thus discriminated against. This is the *reverse* of tolerated theft; even not taking advantage of the position is punished. Periphery players seem to be so averse to players with authority that they are willing to bear the cost of a dramatically reduced surplus. In sum, we find that power from authority has negative effects on both cooperation and efficiency.

Our results for the case with power from authority are of particular interest in relation to the literature on leadership in social dilemmas, which generally finds a positive effect of leadership on cooperation within groups (e.g., Güth et al. 2007; van der Heijden et al. 2009; Stoddard et al. 2014). Our results point to the possibility that the *frame* of leadership might be a crucial factor in determining the effectiveness of 'leadership' in groups. Unlike the works mentioned above, the leaders with authority in our groups decide on their own share first. The *remainder* is then shared out among the rest of the players. While the final division may not be different (indeed, we find that players with authority are often quite cooperative), this move order might be perceived as exploitative by the rest of the group. This alone could cause resentment among those without power, thus leading to lower cooperation.[23,24] While this move order seems natural, our results suggest that legitimacy of the authority might be crucial in determining success in groups.[25] Our results on ostracism indicate that this is indeed the case. We find overwhelmingly that our subjects *do not want* an exogenously appointed leader with authority, even when she is at least as cooperative as those who vote her out. By ostracizing the authority almost immediately in many cases, players without power do not even allow leadership a sporting chance of success. In future research we aim to further explore this intolerance towards authority by varying the perceived legitimacy of this kind of power.

Similarly, our results also shed new light on the positive effects of ostracism on cooperation in social dilemmas. Cinyabuguma et al. (2005) clearly established that the threat of exclusion can enhance efficiency. Our results show that this may not necessarily hold when there is power asymmetry. When there is power from authority, efficiency is not affected by the introduction of ostracism.

Our experiments are the first to show in a controlled setting that power can provide a license to free ride. They also show, however, that this does not hold for all kinds of power. We observe tolerated theft when power stems from centrality, but not when power is derived

[23] This is also observed in the King treatment in Cox et al. (2013).

[24] In addition, our 'leader' cannot discriminate among the periphery players in the group. Whether this may affect perceived cooperativeness is an issue we leave for future research.

[25] For a discussion on the legitimacy of a centralized *sanctioning* authority in groups, see Baldassari and Grossman (2011); and Markussen et al. (2014). They find that a democratically instituted mechanism is more successful in raising contributions than one that is exogenously imposed.

from authority. The behavioral effect that power from centrality leads to free riding in our experiments suggests that the truism that power corrupts has some ground. However, in the world outside of the laboratory both power from centrality and power from authority exist. Our results show that validity of the truism may depend on the type of power concerned. In this study, we looked at one mechanism that might limit the negative effects of power. However, various mechanisms may have evolved, ranging from elections that allow for the dismissal of politicians to anti-corruption laws. Another interesting avenue for future research would be to explore further such mechanisms in the laboratory.

References

Anderson, L.R., Mellor, J.M. & Milyo, J. (2006). Induced heterogeneity in trust experiments. *Experimental Economics*, 223-235

Baldassari, D. & Grossman, G. (2011) Centralized sanctioning and legitimate authority promote cooperation in humans. *Proceedings of the National Academy of Sciences*, 108(27), 11023-11027.

Blurton Jones, N. G. (1984). A Selfish Origin for Human Food Sharing: Tolerated Theft. *Ethology and Sociobiology*, 5, 1-3.

Bonacich, P. (1987). Power and Centrality: A Family of Measures. *American Journal of Sociology*, 92(5), 1170-1182

Buskens, V., Raub, W. & Veer, J. van der (2010). Trust in Triads: An Experimental Study. *Social Networks*, 32, 301-312.

Cassar, A. & Rigdon, M. (2011). Trust and Trustworthiness in Networked Exchange. *Games and Economic Behavior*, 71, 282-303.

Carpenter, J., Kariv, S. & Schotter, A. (2012). Network Architecture, Cooperation and Punishment in Public Good Experiments. *Review of Economic Design*, 16, 93-118.

Charness, G., Feri, F., Meléndez-Jiménez, M. A. & Sutter, M. (2014). Experimental games on networks: underpinnings of behavior and equilibrium selection. *Econometrica, 82*(5), 1615-1670.

Cherry, T. L., Kroll, S. & Shogren, J.F. (2005). The impact of endowment heterogeneity and origin on public good contributions: evidence from the lab. *Journal of Economic Behavior & Organization*, 57(3), 357-365.

Cinyabuguma, M., Page, T. & Putterman, L. (2005). Cooperation Under the Threat of Expulsion in a Public Goods Experiment. *Journal of Public Economics*, 89, 1421-1435.

Cox, J.C., Ostrom, E., Sadiraj, V. & Walker, J.M. (2013) Provision versus Appropriation in Symmetric and Asymmetric Social Dilemmas. *Southern Economic Journal*, 79(3), 496-512.

DeCelles, K. A., DeRue, D. S., Margolis, J. D., & Ceranic, T. L. (2012). Does power corrupt or enable? When and why power facilitates self-interested behavior. *Journal of Applied Psychology*, 97(3), 681-689.

Eckel, C., Fatas, E., & Wilson, R. (2010). Cooperation and Status in Organizations. *Journal of Public Economic Theory*, 12(4), 737-762.

Falk, A. & M. Kosfeld (2006). The Hidden Costs of Control. *American Economic Review*, 96(5), 1611-1630.

Fatas, E., Meléndez-Jiménez, M. A. & Solaz, H. (2010). An experimental analysis of team production in networks. *Experimental Economics, 13,* 399-411.

Fehr, E. & Gächter, S. (2000) Cooperation and Punishment in Public Goods Experiments. *American Economic Review*, 90(4), 980-994.

Fisher, J.R., Isaac, M., Schatzberg, J.W. & Walker, J.M. (1995). Heterogenous Demand for Public Goods: Behavior in the Voluntary Contributions Mechanism. *Public Choice,* 85(3/4), 249-266.

Freeman, L.C. (1979). Centrality in Social Networks: Conceptual Clarification. *Social Networks,* 215-39.

Güth, W., Levati, M.V., Sutter, M. and Heijden, E. van der. (2007). Leading by Example With and Without Exclusion Power in Voluntary Contribution Experiments. *Journal of Public Economics*, 91, 1023-1435.

van der Heijden, E., Potters, J. & Sefton, M. (2009). Hierarchy and Opportunism in Teams, *Journal of Economic Behavior and Organization*, 69(1), 39-50.

Hirshleifer, D. & Rasmusen, E. (1989). Cooperation in a repeated prisoners' dilemma with ostracism. *Journal of Economic Behavior and Organization*, 12, 87-106.

Jackson, M. O. (2008). Social and economic networks, Princeton University Press.

Kreps, D. M., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory*, 27(2), 245-252.

Leibbrandt, A., Ramalingam, A., Sääksvuori, L. & Walker, J.M. (2014). Incomplete Punishment Networks in Public Good Games: Experimental Evidence. *Experimental Economics*, forthcoming.

Markussen, T., Putterman, L., & Tyran, J.R. (2014). Self-Organization for Collective Action: An Experimental Study of Voting on Sanction Regimes. *Review of Economic Studies*, 81, 301-324.

Ostrom, E., Gardner, R. & Walker, J. (1994). Rules, Games and Common-Pool Resources. University of Michigan Press, Ann Arbor.

Persson, T., Tabellini, G & Trebbi, F. (2003). Electoral Rules and Corruption. *Journal of the European Economic Association*, 1(4), 958 –989.

Pitt, E., 1848: The speeches of the right honourable the earl of Chatham in the houses of Lords and Commons. London, Aylott and Jones.

Reuben, E. & Riedl, A. (2013). Enforcement of Contribution Norms in Public Good Games with Heterogeneous Populations. *Games and Economic Behavior*, 77(1), 122-137.

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

Sefton, M., Shupp, R., & Walker, J. M. (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry*, 45(4), 671-690.

de Sousa, L. & M. Moriconi (2013). Why voters do not throw the rascals out? A conceptual framework for analysing electoral punishment of corruption. *Crime, Law and Social Change,* 60(5), 471-502.

Stoddard, B., Walker, J.M. & Williams, A. (2014). *Journal of Economic Behavior and Organization*, 101, 141-155.

Tan, F. (2008). Punishment in a Linear Public Good Game with Productivity Heterogeneity, *De Economist*, 156(3), 269-293,

Tan, F. & Noussair, C. (2011). Voting on punishment systems within a heterogeneous group, *Journal of Public Economic Theory,* 13(5), 661–693.

Zimbardo, P. G., Maslach, C., & Haney, C. 2000. Reflections on the Stanford Prison Experiment: Genesis, transformations, consequences. In T. Blass (Ed.). Obedience to authority: Current Perspectives on the Milgram paradigm (pp.193-237). Mahwah, N.J.: Erlbaum.

Appendix A: Summary of the experimental instructions

Below are the summaries of the instructions as it was handed out to the participants in the experiment. Full instructions for all treatments are available on www.borisvanleeuwen.nl/ACinstructionsfull.pdf. Each paragraph that was included only in some treatments starts with <*treatment acronym*>.

Summary of instructions

Welcome to this experiment on decision-making. You will be paid € 7 for your participation plus whatever you earn in the experiment.

During the experiment you are **not allowed to communicate**. If you have any questions at any time, please raise your hand. An experimenter will assist you privately. You will record your decisions privately and **anonymously** at your computer terminal. Other participants will never be able to link you with your personal decisions or earnings from the experiment.

During the experiment, all **earnings are denoted in points**. At the end of the experiment, your earnings will be converted to euros at the rate: **60 points = € 1**.

The experiment consists of **5 blocks. Each block consists of 5 rounds**. At the end of the experiment, **one block will be randomly selected** and everyone will be paid for their decisions in that block.

The composition of the groups will remain the same for the 5 rounds in a block. At the end of a block, participants will randomly be divided into new groups of five.

At the beginning of the experiment, each participant will randomly be assigned a position - North (N), East (E), South (S), West (W) or Center (C). These **positions will remain fixed throughout the experiment**. For example, if you are assigned the North position, you will be in the north position in each round in each block of the experiment.

At the beginning of each round, each participant receives an **endowment of 50 points**. You decide on how much of this endowment **to invest in each of two accounts**. These are called a "private account" and a "group account". You may invest everything in the private account, everything in the group account, or any combination of the two, as long as you invest 50 points in total.

Your earnings include earnings both from your private account and the group account:

- Earnings from your **private account**: You will earn **1 point for each point invested in your private account**.
- <nAnO, nAnC, nAC> Earnings from the **group account**: Your earnings from the group account are based on the total number of points invested in the group account by all members in your group. Each group member will earn **0.6 points for each point in the group account** regardless of who made the investment.
- <AnO, AnC, AC> Earnings from the **group account**: First, the **total number of points invested in the group account** by all group members is determined. Then, this is **multiplied by 3**. Finally, the **group member in the center (C)** determines how to divide the total amount. S/he determines **how much to keep** for her- or himself. **The remainder is equally divided** among the other group members.

<nAnC, nAC, AnC, AC> Exclusion of a member means that this player can **no longer invest in the group account and will not receive any earnings from the group account in the remaining rounds.** The excluded participant will receive an endowment of 50 points in each of the remaining rounds in the block. All 50 points will automatically be invested in the private account.

<AnC> If someone is excluded, your **earnings from the group account** will depend on the total points invested in the group account by participants in your **sub-group alone**. The amount in the group account will now be multiplied by a factor of 0.6 x (number of sub-group members). If the Center is not excluded, she decides on how much to keep for her/himself and how much to divide equally among the other sub-group members. If the Center is excluded, the amount in the group account is equally divided between the remaining sub-group members.

<AC> If someone is excluded, your **earnings from the group account** will depend on the total points invested in the group account by participants in your **sub-group alone**. The amount in the group account will now be multiplied by a factor of 0.6 x (number of sub-group members). If the Center is not excluded, she decides on how much to keep for her/himself and how much to divide equally among the other sub-group members. If the Center is excluded, the amount in the group account is equally divided between the remaining sub-group members. You will **not earn anything from the group account investments of participants in other sub-groups.**

<nAnC, nAC, AnC, AC> To decide on who will be excluded, the group members will **select candidates for exclusion.**

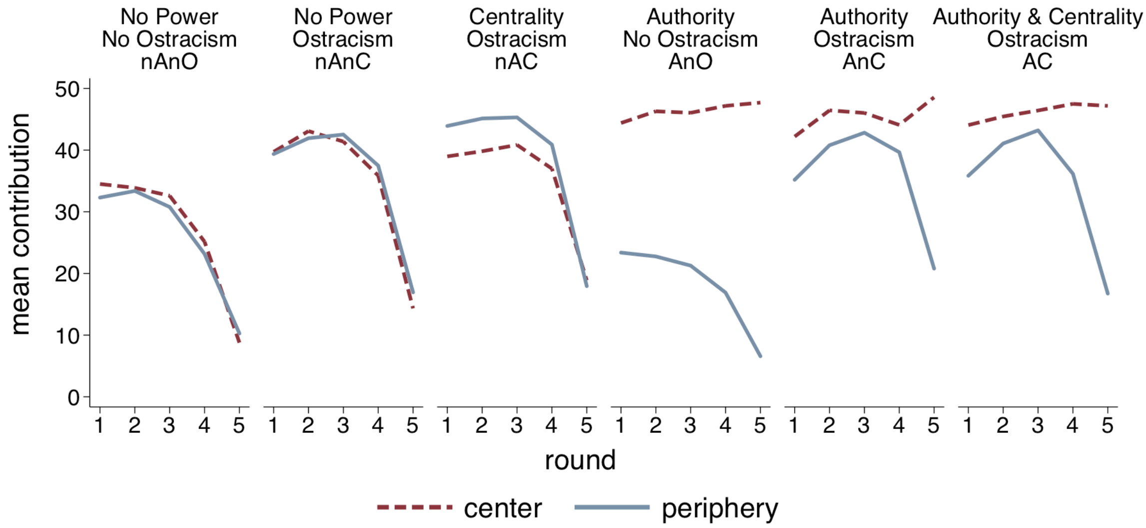
<nAnC, nAC, AnC, AC> You can **indicate for each member of your (sub-)group whether or not you think that s/he should be excluded** from the group in future rounds in the current block. You can vote for as many or as few participants as you want.

<nAnC, nAC, AnC, AC> If some member(s) of the group have previously been excluded, you can only vote on excluding members of the sub-group you are in. Participants who previously have been excluded cannot vote for the exclusion of others.

<nAnC, nAC, AnC, AC> **If half or more members of the (sub-)group vote to exclude a participant, that participant will be excluded in future rounds in the current block.**

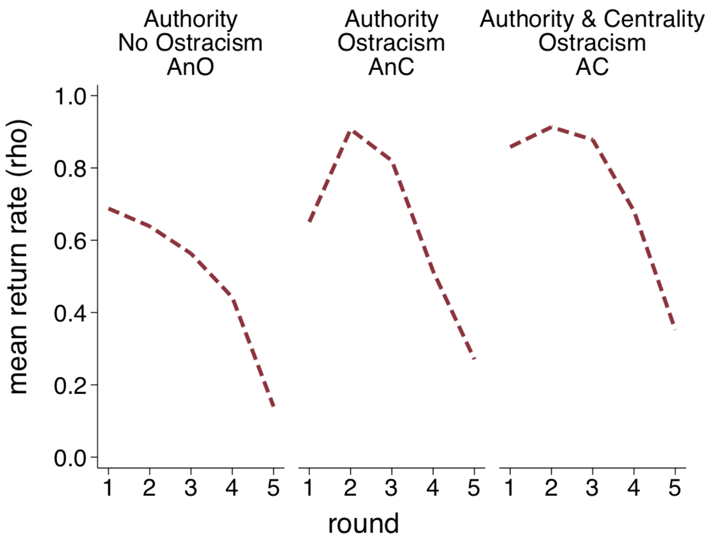
Appendix B: Additional figures and tables

Figure B1. Mean contributions by position



Notes. The three panels on the left denote the treatments without authority and the three panels on the right show the cases with authority. For each round, contributions are averaged across five blocks. The graphs are based on decisions by individuals who are not excluded from their group and who are not isolated.

Figure B2. Mean return rates



Notes. The three panels denote the treatments with authority. For each round, return rates are averaged across five blocks. The graphs are based on decisions by individuals who are not excluded from their group and who are not isolated.

Table B1. Mean cooperation

			No Authority	Authority
No Ostracism (⇒ no Centrality)			*nAnO*	*AnO*
		All players	0.521 (0.075)	0.392 (0.105)
		Center	0.533 (0.112)	0.464 (0.124)
		Periphery	0.518 (0.076)	0.375 (0.103)
Ostracism	**No Centrality**		*nAnC*	*AnC*
		All players	0.699 (0.115)	0.651 (0.099)
		Center	0.692 (0.136)	0.497 (0.169)
		Periphery	0.701 (0.110)	0.689 (0.101)
	Centrality		*nAC*	*AC*
		All players	0.756 (0.057)	0.670 (0.105)
		Center	0.694 (0.091)	0.645 (0.161)
		Periphery	0.771 (0.054)	0.677 (0.098)

Notes. Cells give mean cooperation (γ) with standard deviations in parentheses. Entries based on all rounds and blocks, and players that were not excluded or isolated. The unit of analysis is the matching group.

Table B2. Mean earnings

			No Authority	Authority
No Ostracism (⇒ no Centrality)			*nAnO*	*AnO*
		All players	102.1 (7.5)	98.7 (8.9)
		Center	101.5 (8.2)	182.6 (17.3)
		Periphery	102.3 (7.6)	77.7 (12.5)
Ostracism	**No Centrality**		*nAnC*	*AnC*
		All players	113.6 (14.0)	100.3 (13.7)
		Center	114.0 (14.4)	137.9 (29.7)
		Periphery	113.5 (14.1)	90.8 (10.3)
	Centrality		*nAC*	*AC*
		All players	120.5 (5.2)	109.0 (15.6)
		Center	124.0 (5.6)	163.4 (26.4)
		Periphery	119.6 (5.4)	95.4 (17.5)

Notes. Cells give mean earnings per round in points, with standard deviations in parentheses. Entries based on all rounds and blocks, and all players, regardless of being excluded or isolated. The unit of analysis is the matching group.