Peers or Police?

Detection and Sanctions in the Provision of Public Goods*

Gregory DeAngelo[1] and Laura K. Gee[2]

[1]West Virginia University, Economics (gregory.deangelo@mail.wvu.edu)

[2]Tufts University, Economics (laura.gee@tufts.edu)

February 22, 2017

(Keywords: Public Goods, Punishment, Detection, Deterrence)

(JEL: C72, C91, C92, D7, H41)

*A huge thanks to Michael Kuhn for his comments on the experimental design. We would also like to thank for their comments participants at the Social Dilemmas Workshop, the New England Experimental Economics Workshop, the ESA conference, and the Appalachian State seminar, as well as David Huffman and Michael Makowsky. Thanks also to Abdul Haleem Kidwai for his excellent research assistance. This research was approved by the Tufts IRB (1408019).

Abstract

Promoting good or discouraging bad behavior is a feature of society, as seen in punishments for breaking laws, shaming in social environments, or online reviews of products or services. While there has been much focus on the incentives that encourage compliance, one's actions need to be detected to be rewarded or sanctioned. In this work, we examine both detection and sanctioning in public goods games in order to determine how detection and sanctioning interact to encourage prosocial behavior. We compare peer detection of one's closest connection with central detection, which broadcasts actions to the whole group. We also compare sanctioning of one's closest connection with concentrated sanctioning, which focuses sanctions on the largest free-rider. We find that the choice to endogenize detection leads to a large decline in detection when central detection is in place but a smaller decline when peer detection is employed. This detection behavior arises because free-riders are willing to detect their neighbors but unwilling to implement the centralized detection regime that will catch their own bad acts. In general, we find that peer detection coupled with concentrated sanctioning encourages the highest level of contributions to the public good, which is consistent with the mechanisms that are often observed in practice.

1 Introduction

Promoting good behavior or, alternatively, discouraging bad behavior is a feature of society at all levels. Whether discussing law enforcement issuing citations or arresting citizens for behavior that is not compliant with the legal code, citizens shaming other members of society for behaving in a disorderly manner, or reviewers criticizing their experience with a vendor online, observing others and offering feedback are both critical to altering behavior. While research on the feedback mechanisms that can be utilized to discourage bad behavior or encourage good behavior is plentiful, we can only offer feedback if we observe others' actions.

While it seems straightforward that one can only be rewarded/sanctioned if they have been detected, it can be a matter of practical difficulty. Detecting someone that is behaving antisocially can be accomplished without any group coordination when, for example, reporting shoddy goods from a seller or inappropriate social media posts of peers (e.g., on eBay or Facebook). In these settings, one detects only the peers they are directly interacting with and may remain uninformed about the actions of the population as a whole. In contrast, when detection is done by a centralized agency, that agency may inform you about the actions of many members of society and, in turn, may broadcast your actions widely. However, when detection is assigned to a centralized agency, implementing detection can become complicated. Centralization of detection requires the group to collectively decide on how to implement detection and opens up the opportunity for a single individual or a subset of the group to veto implementation. For example, a single country in the UN Security Council can block a resolution to detect nuclear arms.[1] Or in the case of the Transportation Security Administration, a single individual has made it increasingly difficult for whistle-blowers to detect and expose overspending.[2]

[1] See https://www.un.org/en/sc/meetings/voting.shtml.

[2] In a 2011 interview discussing the Whistleblower Protection Act, Tom Devine, the legal director of the Government Accountability Project, noted that the Merit Systems Protection Board (the governing body that determines whether an individual is protected under the WPA) noted that agencies such as the TSA have the right to cancel the WPA through internal secrecy regulations. The decision to suspend the WPA in matters of secrecy was enacted by a single individual, Neal McFee, who was the chairman of the MSPB. Moreover, under the direction of McFee, the MSPB has only ruled in favor of a whistle-blower one time in 45 hearings. As a result, whistle-blowers are unlikely to come forward, thereby decreasing the likelihood that wasteful spending, misuse of authority, etc., will be detected. This unfortunate outcome is essentially due to the authority of a single decision maker's prerogative.

This paper provides an experimental comparison of two types of detection: peer detection, where bad acts are *individually* detected by one's peers (or individuals that they frequently interact with), and centralized detection, where a *coordinated* agency monitors everyone's actions and broadcasts them to the group. Both types of detection are observed in practice, but a thorough investigation of different regimes has not been conducted experimentally.

In addition to comparing peer and centralized detection, we also vary the type of sanctioning mechanism that is coupled with each detection mechanism. Specifically, we vary whether subjects can sanction through a neighbor sanctioning mechanism or a concentrated sanctioning mechanism. Our neighbor sanctioning mechanism differs from the traditional peer-to-peer sanctioning by allowing a subject to only sanction a single person, while in our concentrated sanctioning mechanism subjects can collectively punish the lowest detected contributor to the public good.[3] In both sanctioning mechanisms, subjects can only be sanctioned if they are detected. The four resulting detection and sanction combinations (PeerDetect-NeighborSanction, PeerDetect-ConcentratedSanction, CentralDetect-NeighborSanction, and CentralDetect-ConcentratedSanction) may seem very different, but by design the average costs of detection, expected number of detected people, and possible earnings are the same across all four.

These detection and sanctioning regimes are examples of the type of combinations that can arise. The peer detection–neighbor sanctioning combination (PeerD-NeighborS) is a stylized version of detecting if your neighbor did not shovel their snow and, if they did not, writing them a nasty note in response. Peer detection and concentrated sanctioning (PeerD-ConcentratedS) could represent how eBay buyers monitor the sellers they interact with, and, if reported, the worst offenders are removed from the website. Centralized detection and neighbor sanctioning (CentralD-NeighborS) can be thought of as a simplification of websites like charitynavigator.com, which provides ratings of charities and then allows individuals to withhold contributions from poorly rated charities. Last, the centralized detection and concentrated sanctioning (CentralD-ConcentratedS) is a stylized version of the police, which detect wrongdoing, and the courts, which sanction individuals. The specific

[3]Our concentrated sanctioning mechanism can be thought of as a directed star network and is closely related to the sanctioning mechanisms of Yamagishi [1986], Andreoni and Gee [2012], and Andreoni and Gee [2015].

formulations used in our study are meant to be one of any number of examples of detection regimes and sanctioning mechanisms.

Exploring these questions with observational data is difficult due to the endogenous adoption of detection/punishment regimes. While optimal sanctioning has received considerable attention within the experimental economics literature, especially with regards to the public goods game, detection is still understudied. These studies find that a common way to increase contributions to the public good is by adding a sanctioning mechanism (see Chaudhuri [2011] for a review). Yet there has been some debate about whether groups are best served by an informal peer-to-peer mechanism, where anyone can punish any other person, or a formal unified mechanism (e.g., Andreoni and Gee [2012], Markussen et al. [2014], Kamei et al. [2015]).

In contrast to the debate about the type of sanctioning, most previous experimental work exogenously imposes a detection regime that automatically observes all (or some fraction of) the contributions and broadcasts information to the whole group. In reality, information is not always widely available, either because a person can only monitor those in close proximity or because people do not exert the effort to monitor each other.

Experimental designs that utilize automatic detection assume that individuals will always be willing to pay the costs to either individually investigate others' actions or to create a centralized monitoring agency. In reality, the optimal type of detection mechanism might be highly dependent on the individual or group's willingness to pay for detection. It could be the case that the willingness to pay may be particularly low for the groups with the most free-riding. It is difficult to imagine criminals paying their taxes so that there will be better funding for centralized policing, or individual drag racers calling the police to report speeding by a peer. So, it is an open question whether peer or centralized detection yields the highest overall well-being, and if the same detection regime is best when we do not assume everyone is willing to pay for detection.

In our experiments, when there is an endogenous choice rather than an exogenous mandate to detect, endogenous peer detection is identical to exogenous peer detection because there is no coordination issue when peers simply detect other peers. However, when groups must endogenously decide whether to implement a centralized detection regime in our experiment, this requires unanimous approval. Endogenizing this framework permits a single individual to prevent group-level detection from occurring, similar to the way a single country can block a UN resolution. We also report experiments where a high percentage of subjects (but not all) must approve centralized detection in order to test the robustness of our results.

We find that contributions to the public good are lower under centralized detection than peer detection if subjects have to coordinate to implement a detection regime. The coordination failure stems from the largest free-riders being unwilling to support a centralized agency (e.g., the IRS) whose sole purpose is to expose their free-riding and enable sanctioning. We also find contributions to the public good are lower under concentrated sanctioning than peer sanctioning when only a few people are detected (either centrally or by peers). In these instances, even hefty fines do not deter bad behavior when someone knows they are unlikely to be caught.

Centralizing detections or sanctions when the community might let free-riding members go undiscovered is not welfare improving for the group. When the choice to establish a centralized detection regime (e.g., organize a police force) is endogenous, it leads to lower contributions and less detection, as the largest free-riders are unwilling to fund the monitoring agent that might detect them. Also, concentrated sanctions proves fruitful only when the group will detect some antisocial behavior. In our experiment, concentrated sanctioning deters antisocial behavior when there is a reasonable amount of detection (either by peer or central detection) but fails when there were very few detections. When no one is caught, then there is no one for the concentrated sanctioning mechanism to punish. Even the highest fines are nondeterrent when you know you will not be detected by either a peer or the police.

In the next section, we review the related literature, with special attention to the experimental literature on public goods games. In Section 3, we present a model of our experimental design as well as theoretical predictions regarding detection and sanctioning behavior. In Section 4, we present the results of our experiment, and in Section 5 we offer concluding remarks.

2 Related Literature

There is a long-standing debate about whether groups are better off with a centralized authority or with self-governance [Hobbes, 1969, Ostrom, 2015]. Although centralization (e.g., the police or a homeowner association) is common in the real world, some observational research has noted that peers can be just as effective. For example, Karstedt-Henke [1991] notes that for every juvenile crime detected by police, parents detect at least four, teachers detect two, and friends detect more than five similar crimes. Moreover, informal sanctioning mechanisms (e.g., shaming) have proven quite useful in deterring antisocial acts in some settings (e.g., sex offender registrations).[4]

While observational research has examined the effect of different types of detection and sanctioning, these analyses struggle with endogeneity issues. Namely, locations with higher willingness to pay for enforcement might invest in higher levels of centralized enforcement or supplement centralized enforcement with peer enforcement. In either situation, the level and type of enforcement is endogenously determined. Moreover, these environments can usually only measure sanctions that are centrally administered (e.g., fines collected by the government, jail time, etc.). In light of these shortcomings, controlled lab experiments provide a fruitful environment for examining how different types of detection and sanctions interact.

Many previous experiments have concentrated on the optimal sanctioning mechanism to increase public goods provision. Both peer-to-peer sanctioning[5] and unified sanctioning[6] have been shown to increase public contributions relative to no sanctioning.

Unified sanctions might be especially effective because they allow higher sanctions to be focused on the largest free-riders. The magnitude of the sanction can increase prosocial actions [Beccaria, 1764, Becker, 1968, Grogger, 1991]. Nikiforakis and Normann [2008] find higher peer-to-peer sanc-

[4]See DeAngelo and Smith [2016], DeAngelo and Reimers [2016], Ellickson (1991), Greif [1993, 1994], Harel and Klement [2005], and Teichman [2005] for empirical work on the effectiveness of peer detection and sanctioning. Levitt [1997], Klick and Tabarrok [2005], and DeAngelo and Hansen [2014] examine the effectiveness of centralized enforcement and sanctioning on antisocial behavior. Acemoglu and Wolitzky [2015], Acemoglu and Jackson [2014], and Deb and González-Dıaz [2014] offer theoretical contributions on the relationship and effectiveness of both community and peer enforcement, both independently and interdependently.

[5]See Egas and Riedl [2008], Gächter et al. [2008], Fehr and Gächter [2002], Fehr and Gachter [2000], and Ostrom et al. [1992].

[6]See Baldassarri and Grossman [2011], Dickinson and Villeval [2008], Falkinger et al. [2000], and Yamagishi [1986].

tions result in higher public contributions. And beyond the public goods setting, the evidence is mixed, with some finding the optimal fine is the maximum fine [Polinsky and Shavell, 1984] and others finding the opposite (see Andreoni [1991] and Polinsky and Rubinfeld [1991]).[7] However, a comparison of whether subjects make larger contributions when there are concentrated rather than peer sanctions depends on the exact setting and costs. For example, Andreoni and Gee [2012] find that public contributions are similar when comparing peer-to-peer sanctioning alone relative to concentrated sanctioning alone. But when both peer-to-peer and concentrated sanctions are available simultaneously, contributions are higher than the levels that are observed when peer-to-peer sanctioning only is permitted.

In many of the aforementioned experimental public goods papers, the detection of unwanted behavior has been exogenously imposed, costless, and centralized such that information is broadcast to the whole group with 100% certainty. So there has been relatively little investigation into the optimal type of detection. One line of work related to optimal detection finds that adding uncertainty about whether the actions detected have actually taken place generally lowers cooperation.[8] Also, outside the public goods setting, research has shown that marginal increases in the probability of detection are more effective than marginal increases in sanctions (see Grogger [2007], Bar-Ilan and Sacerdote [2001], DeAngelo and Charness [2012], Friesen [2012]). More closely related to this paper is previous work that explores how detection of only a subset of the group affects cooperation.

Fatas et al. [2010] find that centralized detection does not always result in the highest contributions when there is no sanctioning mechanism. They find that public contributions are highest in an undirected star network where a single central subject's contribution is known to all and that central subject knows everyone else's contributions. The public contributions from the standard complete network (the setting where everyone is able to detect everyone else) are the next highest, although the two are not always statistically significantly different from each other. These two detection regimes

[7] Anderson et al. [2017] test whether decreasing or increasing sanctions are optimal, finding support for decreasing penalty structures.

[8] See Aoyagi and Frechette [2009], Grechenig et al. [2010], Fudenberg et al. [2012], Ambrus and Greiner [2012], Fischer et al. [2013], and Dal Bó and Fréchette [2014].

outperform a bidirectional circle or bidirectional line network.[9] So even without any sanctioning, it is not clear whether the standard centralized detection regime results in the greatest public good provision.

Adding in the option to sanction further complicates the main findings in this literature. Papers that explore the optimal number of subjects to detect and sanction find mixed results. Boosey and Isaac [2014] and Carpenter et al. [2012] find no difference in contributions between the standard complete network setting where everyone is able to detect/sanction everyone else and a setting where each subject can only detect/sanction their two neighbors in a bidirectional circle. Similarly, O'Gorman et al. [2009] find no difference in contributions but increased net earnings when comparing the standard complete network setting and a setting where a single subject can punish everyone. Faillo et al. [2013] find that when subjects must concentrate their punishments on the largest free-riders, if they detect all contributions (similar to our centralized detection and concentrated sanction treatment) this can increase contributions and earnings, but if the group only detects the lowest contributors there is a decline in contributions and earnings. Carpenter [2007] uses four different detection/sanctioning treatments and finds that the lowest contributions come from a directed circle network where subjects can only detect/sanction a single subject on a directed circle (similar to our peer detection neighbor sanction treatment). In a follow-up paper, Carpenter et al. [2012] find that the level of contributions in this same directed circle network is actually higher than that from the standard complete network. Thus, there is not currently a consensus on the optimal choice of joint detection/sanctions. Furthermore, all previous research has made the ability to detect/sanction exogenously automatic.

However, there is evidence that some subjects are willing to endogenously choose to detect/sanction others. For example, Page et al. [2008] find that about 20% of subjects will pay a $1 cost to detect/sanction the actions of the group, and Andreoni and Gee [2012] find that 72% to

[9]In an undirected circle network, a person can detect the actions of their left- and right-hand-side neighbors; this differs from our setup, which is a directed circle where a person can only detect their right-hand-side neighbor. A line is similar, but the endpoints can only detect the actions of one neighbor. In an undirected star, the center person can detect everyone's contributions and can themselves be detected by everyone, while the arms of the star can only be detected by the center person.

85% of subjects will pay an average of $0.50 each to implement a centralized sanctioning mechanism. Yet, Botelho et al. [2007] find that subjects do not vote for a peer-to-peer sanctioning mechanism. And Ertan et al. [2009] find that initially subjects don't vote for a peer-to-peer sanctioning mechanism but that eventually they will support such a mechanism if it only punishes the largest free-rider.[10] Although the exact level of opting into a sanctioning mechanism varies in the previous work, it is clear that when given a choice to detect/sanction we do not see the same results as when detection/sanctions are exogenously imposed.

3 Model and Experiment

The experiment is a $2 \times 2 \times 2$ design run between subjects. The first dimension we vary is whether detection of others' actions is carried out by a peer or by a centralized detection regime. The second dimension we vary is whether the sanctioning mechanism allows sanctions to be imposed by a single neighbor or by many group members and concentrated on the largest free-rider. The last dimension we vary is whether the choice to detect is exogenously automatic or endogenously determined. Note that we do not allow subjects to choose between detection regimes, but, rather, subjects decide whether to pay a cost to use the detection regime they have been assigned to. That leaves us with the following eight treatments:

1. EXG PeerD-NeighborS: Exogenous Peer Detect-Neighbor Sanction
2. EXG PeerD-ConcentratedS: Exogenous Peer Detect-Concentrated Sanction
3. EXG CentralD-ConcentratedS: Exogenous Central Detect-Concentrated Sanction
4. EXG CentralD-NeighborS: Exogenous Central Detect-Neighbor Sanction
5. ENDG PeerD-NeighborS: Endogenous Peer Detect-Neighbor Sanction
6. ENDG PeerD-ConcentratedS: Endogenous Peer Detect-Concentrated Sanction
7. ENDG CentralD-ConcentratedS: Endogenous Central Detect-Concentrated Sanction

[10] There are also papers that explore whether subjects prefer a peer-to-peer sanctioning mechanism to a unified sanctioning mechanism (See Traulsen et al. [2012], Markussen et al. [2014], and Kamei et al. [2015]). We only allow the choice to use the assigned type of detection in our study, rather than a choice of type of detection.

8. ENDG CentralD-NeighborS: Endogenous Central Detect-Neighbor Sanction

The experiment lasts a total of 10 periods with three stages in each period. Subjects are paid for one randomly selected period to minimize income effects. We chose to have subjects remain in the same group of four for all 10 periods, with fixed identities to approximate the repeated interactions and reputations that arise in small groups providing public goods.

In Stage 1, subjects play a linear public goods game. In Stage 2, the actions of other subjects are probabilistically revealed, depending on whether the subject purchased the ability to detect, which can be exogenously or endogenously determined. In Stage 3, subjects choose whether to sanction other group members, depending on whether the subject has detected the behavior of other group members.

Below, we describe the actions in all three stages and then the equilibrium predictions.

3.1 Stage 1: Contribution

In Stage 1 in all treatments, subjects play a linear public goods (LPG) game. Each person is endowed with $\omega_1 = 5$ tokens that they can allocate between a private account that pays $\beta = \$2$ to only that person, or to a public account that pays $\alpha = \$1$ to every person in their four-person group, which makes group returns to the public account $N*\alpha = \$4$. Each person chooses an amount g_i to contribute to the public account. Stage 1 payoffs can be represented by the following:

$$\pi_i^1 = \beta(\omega_1 - g_i) + \alpha \sum_{j=1}^{N} g_j \tag{1}$$

$$\pi_i^1 = 2(5 - g_i) + \sum_{j=1}^{4} g_j \tag{2}$$

3.2 Stage 2: Detection

In Stage 2, subjects can detect the actions of other group members. We examine two different detection environments: exogenous and endogenous detection. In both environments, the subjects are provided a Stage 2 supplement of \$1. In the exogenous detection treatments (EXG), the supplement is automatically paid to an information fund that makes the subject eligible to observe the contributions of other group members.[11] In the endogenous detection treatments (ENDG), the subjects are given a choice of whether they would like to keep the \$1 Stage 2 supplement or use it to become eligible to detect others. We will discuss this choice further in the following sections. Finally, subjects are randomly assigned to a treatment that either has peer detection or centralized detection.

3.2.1 Peer Detection

In the peer detection treatments (ENDG/EXG PeerD-NeighborS and PeerD-ConcentratedS), each subject is assigned to detect their right-side group member's contribution, g_r. That means that Person A is assigned to detect the actions of Person B. It follows that B detects C, C detects D, and D detects A, so that information flows as shown in Figure 1. This is meant to approximate the idea of detecting only the people one interacts with closely (e.g., your immediate neighbors, your friends on Facebook, specific buyers/sellers on eBay). In the exogenous environment, subjects automatically pay their Stage 2 supplement and are eligible to detect their right-side group member's contributions. In the endogenous treatment, subjects have a choice between keeping their \$1 Stage 2 supplement or instead using the supplement to be eligible to observe their right-side group member's contributions. Each subject is told that they will *only* be able to sanction another group member in Stage 3 if that person is detected in Stage 2.

If the Stage 2 supplement is paid, then the subject's right-side group member is detected with probability $P = \frac{2}{3}$. We set $P < 1$ because we wanted a range of zero to four subjects to be detected in both the exogenous and endogenous environments. In the special case where the Stage 2 supplement is paid and the right-side group member contributed their whole Stage 1 endowment to the public good, $g_i = \omega_1 = 5$, then that prosocial person is detected with probability $P = 1$. This is meant to capture the idea that when a neighbor behaves prosocially, they have no reason to hide their behavior, whereas a free-riding neighbor may try to hide this action from others, so there is some

[11] For brevity and ease of comprehension, we do not inform subjects in the exogenous enviroment that they have been given a \$1 supplement that will automatically be paid.

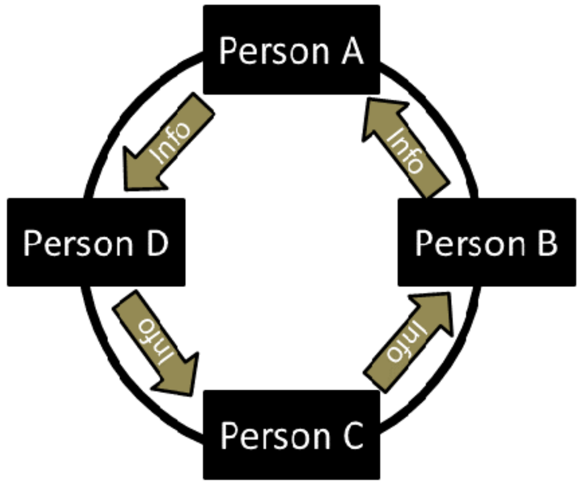


Figure 1: Peer Detection Information Flow

probability they will go undetected. Stage 2 payoffs can be summarized as $\pi_i^{2,PeerD} = \pi_i^1$ if the Stage 2 supplement is paid, or $\pi_i^{2,PeerD} = \pi_i^1 + 1$ if the Stage 2 supplement is not paid.

For each player, a random number is drawn from a uniform distribution between zero and one, and if that number is less than or equal to two-thirds, then that person is detected if the Stage 2 supplement was paid (either automatically in the exogenous environment or by choice in the endogenous environment). So, the probability of detecting the actions of the right-side group member can be summarized as:

$$P^{PeerD} = \begin{cases} \frac{2}{3} & \text{if supplement is paid and} \quad g_r < \omega_1 = 5 \\ 1 & \text{if supplement is paid and} \quad g_r = \omega_1 = 5 \\ 0 & \text{if supplement is not paid} \end{cases}$$

At the end of Stage 2, subjects are shown the actions of their right-side group member if they were detected.

3.2.2 Central Detection

In the central detection treatments (ENDG/EXG CentralD-NeighborS and CentralD-ConcentratedS), subjects must collectively fund a centralized detection regime. When the detection regime is in place, it will report information about all detected members to everyone in the group. This is meant to approximate the idea that centralized detection requires some form of group coordination.

In the exogenous environment, subjects automatically pay their Stage 2 supplement and are eligible to detect everyone in the group. In the endogenous treatment, the subjects have a choice between keeping their Stage 2 supplement or, instead, pledging it toward a centralized detection regime. This is meant to approximate a costly vote or costly fundraising to hire an investigator. Each subject is told that they will *only* be able to sanction another group member in Stage 3 if that person is detected in Stage 2.

If everyone in the group pledges their Stage 2 supplement such that the sum of pledges is $4, then the centralized detection regime is implemented, and everyone pays the $1 supplement. If the sum of pledges is lower than $4, then no one is detected, and no one pays the $1 supplement.

We require unanimity for the endogenous centralized detection treatments to keep the costs of detection the same across peer versus centralized detection. In peer detection, it costs $1 per two-thirds chance of seeing the actions of any single person in the group. In centralized detection, it costs $4 per two-thirds chance of seeing the actions of all four persons in the group; i.e., an average of $1 per chance of seeing the actions of any single person in the group.

If we were to use another voting rule than unanimity—for example, if we were to require three of the four group members to implement centralized detection—then the costs would fall to $3 per two-thirds chance of seeing the actions of all four persons in the group; i.e., an average of $0.75 per chance of seeing the actions of any single person in the group.[12] Thus, we would be changing both the price of detection and the type of detection simultaneously, which would not enable us to isolate the effect of peer detection versus centralized detection from the effect of the price change.

Although unanimity is a common way for groups to make decisions, it certainly is not the only way. It is not immediately obvious what would happen if we were to lower the requirement from unanimity to three out of four in the group. Lowering the threshold to three out of four could

[12] We could also change the price of the pledge from $1 per person to $1.33 per person and require only three out of four, but then we would be changing the Stage 2 supplement of our subjects.

increase the likelihood of centralized detection being implemented by both lowering the costs and the number who need to choose it. On the other hand, lowering the threshold to three out of four could decrease the likelihood of implementation by creating coordination issues and making individuals less pivotal. It is therefore an empirical question whether unanimity decreases the chances that centralized detection will be implemented.

However, we have some data from pilot sessions that can address this question. In pilot sessions, we found that lowering the threshold from unanimity to three out of four nearly halves the proportion who pledged the information fee.[13] This implies that were we to loosen the unanimity requirement, we would be even less likely to observe centralized detection implemented than we are in our current setup.

When the centralized detection regime is implemented, the actions of each group member are detected with probability $P = \frac{2}{3}$. It is made clear to the subjects that these are independent draws, so that when the centralized detection mechanism is created they may see the Stage 1 contributions of zero, one, two, or three of their other group members. It is similarly known that when the centralized detection regime is implemented, their own actions may or may not be revealed to the group. However, in the special case when the centralized detection regime is implemented and a group member contributed their whole Stage 1 endowment to the public good ($g_i = \omega_1 = 5$), then that person is detected with probability $P = 1$. So Stage 2 payoffs can be summarized as:

$$\pi_i^{2,CentralD} = \begin{cases} \pi_i^1 & \text{if the supplement is pledged, and \$4 is raised} \\ \pi_i^1 + 1 & \text{if the supplement is pledged, but less than \$4 is raised} \\ \pi_i^1 + 1 & \text{if the supplement is not pledged} \end{cases}$$

And the probability of detecting the actions of each group member can be summarized as:

[13]See footnote 21 for details.

$$P_i^{CentralD} = \begin{cases} \frac{2}{3} & \text{for each } i \text{ if centralized regime implemented and} \quad g_i < \omega_1 = 5 \\ 1 & \text{for each } i \text{ if centralized regime implemented created and} \quad g_i = \omega_1 = 5 \\ 0 & \text{if centralized regime not implemented} \end{cases}$$

At the end of Stage 2, subjects are shown the actions of all detected group members. After Stage 2 has concluded, subjects move to Stage 3.

3.3 Stage 3: Sanctions

In Stage 3, subjects can assign sanctions to their other detected group members. They are randomly assigned to a treatment that either has neighbor sanctions or concentrated sanctions.

3.3.1 Neighbor Sanctions

In the neighbor sanction treatments (ENDG/EXG PeerD-NeighborS and CentralD-NeighborS), each subject is assigned to sanction their right-side group member if that person was detected in Stage 2. That means that Person A is assigned to sanction Person B. It follows that B sanctions C, C sanctions D, and D sanctions A, so that sanctions flow as shown in Figure 2. This is meant to approximate a situation where you can only sanction the person you interact with the most—i.e., your closest neighbor—and you don't require any group consensus to do so.

Subjects decide how many sanctioning points they would like to assign to their right-side group member. Note that in the special case where their right-side group member has given the maximum to the public good, a subject cannot assign any sanctioning points,[14] so that $p_{i,r}$ is the number of sanctioning points from subject i to their right-side group member r. Each sanctioning point costs person i \$1 and reduces the payoff of the right-side group member r by \$3. We denote sanctioning

[14]We prohibit this type of antisocial sanction because such sanctions are not available in our centralized sanctioning mechanism. Antisocial sanctions are one of the reasons that peer sanctioning may result in lower contributions than centralized sanctioning, so by precluding this extreme antisocial sanctioning, we bias ourselves toward finding fewer difference between our peer and central sanctioning mechanisms.

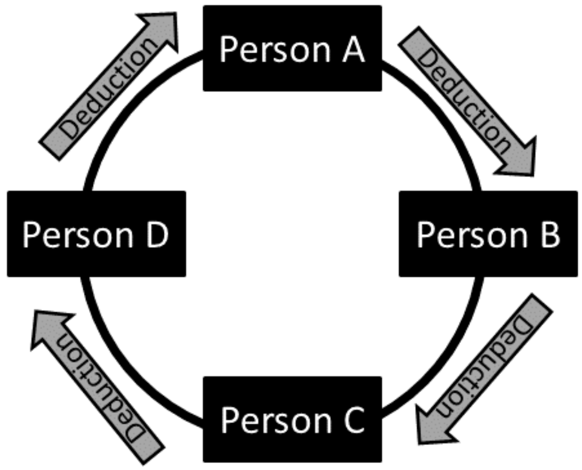


Figure 2: Peer Sanction Flow

points assigned to player i from the person to their left as $p_{l,i}$. So Stage 3 payoffs can be summarized as:

$$\pi_i^{3,NeighborS} = \begin{cases} \pi_i^2 - p_{i,r} - 3p_{l,i} & \text{if detect right-member and if person } i \text{ detected} \\ \pi_i^2 - p_{i,r} & \text{if detect right-member and if person } i \text{ not detected} \\ \pi_i^2 - 3p_{l,i} & \text{if do not detect right-member and if person } i \text{ detected} \\ \pi_i^2 & \text{if neither right-member nor } i \text{ detected} \end{cases}$$

Under the neighbor sanctioning mechanism, each subject may sanction exactly one other subject, and each subject may be sanctioned by exactly one other subject.

3.3.2 Concentrated Sanctions

In the concentrated sanctions treatments (ENDG/EXG PeerD-ConcentratedS and CentralD-ConcentratedS), subjects can assign sanctions that will reduce the payoff of a single group member who is chosen by the concentrated sanctioning mechanism. This is meant to approximate a situation where a group hires a single organization to punish the largest offender.

The cost structure is the same as the neighbor sanction treatments. Each sanctioning point costs person i \$1 and reduces the payoff of the sanctioned group member by \$3. The concentrated sanctioning mechanism examines the Stage 1 contributions (g_i) of all the subjects who were detected in Stage 2 and locates the subject with the lowest *detected* contribution in the group. If multiple players tie for the lowest contribution, then the sanction is split evenly among them. If the lowest detected contributor gave their whole endowment $g_i = 5$, then no punishment is levied and payments for punishment points are refunded to subjects. This is meant to approximate the idea that enforcement would not sanction a person obeying the law. If the lowest detected contributor did not contribute the full amount, however, then they will have their payoff reduced by three times the sum of the group's sanctioning points, $3\sum_j^N p_j$. Stage 3 payoffs can be summarized as:

$$\pi_i^{3,ConcentratedS} = \begin{cases} \pi_i^2 - p_i - 3\sum_j^N p_j & \text{if lowest detected member} \\ \pi_i^2 - p_i & \text{if not lowest detected member} \\ \pi_i^2 & \text{if no one detected} \end{cases}$$

Under the concentrated sanctioning mechanism, each subject may sanction exactly one other subject, but a single subject may be punished by multiple other subjects.[15] At the end of Stage 3, subjects are shown their own and others' Stage 3 payoffs.

3.4 Equilibrium

In treatments with either neighbor or concentrated sanctioning, in the last period there is no incentive for subjects to invest in costly sanctioning, so we expect $p_i = 0$ for all i. So, under both sanctioning mechanisms, we expect no sanctioning to take place in Stage 3 in the final period.[16]

In treatments with peer or centralized detection in Stage 2, since no sanctions will take place in Stage 3, there is no reason to pay a fee to detect players that will not be sanctioned either by an individual or by a centralized mechanism. As we expect no one to pay the \$1 supplement, there is a

[15] With the exception of ties for lowest detected contributions, then all those tied split the punishment equally between them.

[16] If subjects could make some sort of contingent contract, like "I will punish any person who is detected enough that they wish that had not engaged in free-riding," then there would be other possible equilibria since a best response would be to not free-ride in this setting [Andreoni and Gee, 2012, 2015]. However, such contracts are not allowed in our setup.

$P = 0$ probability of detection in the endogenous treatments. In the exogenous treatments, subjects do not take any action in Stage 2 because detection is automatic. However, since subjects will not sanction in Stage 3 in the final period, subjects will essentially ignore any information provided by an exogenous detection regime.

Since no sanctions are levied in Stage 3, the Stage 1 game is the simple linear public goods game with the following payoffs:

$$\pi_i^1 = 2(5 - g_i) + \sum_{j=i}^{4} g_j \tag{3}$$

Since the payoff to keeping a token is \$2, while the payoff to contributing a token is only \$1, the subgame equilibrium is for each person to choose $g_i = 0$, which would result in $\pi_i^1 = \$10$. However, the group utility maximizing choice would instead set $g_i = \omega = 5$, which would result in $\pi_i^1 = \$20$. Thus, in all treatments, the Nash equilibrium prediction is zero contributions to the public good, with no one paying for detection and no sanctions being levied. Subjects play for a finite number of periods, so, using backward induction, we have this same subgame equilibrium in each period.

4 Results

We recruited 272 students to participate in the laboratory at the University of Massachusetts Amherst, using the Online Recruitment System for Experimental Economics software [Greiner, 2004]. Subjects participated in one of the eight treatments in fixed groups of four for 10 periods. Subjects only interacted during the experiment via a computer program that was created using z-Tree [Fischbacher, 2007]. The number of subjects per treatment varied from 28 to 40 subjects, and each session always had at least 12 subjects to preserve anonymity. On average, subjects earned a payoff of \$19.20 (inclusive of a \$5.00 show-up payment).

As we previously noted, almost all research related to public goods games has automatic detection and reporting of all actions to all group members. We relax this assumption by allowing subjects in our endogenous treatments to choose whether they invest in a regime that either allows peer or centralized detection. In focusing on these changes to the detection of subjects, we contribute to the research on public goods by determining how/whether the detection mechanism alters punishment and contribution behavior relative to the benchmark where subjects are automatically detected and observed by others.

Two findings stand out from this exercise. First, when we endogenize detection choices, we find that centralized detection environments yield high willingness to pay for detection from those making substantial contributions to the public good but low willingness to pay from free-riders. Alternatively, in the peer detection environment, we find no systematic relationship between the level of public goods contributions and willingness to pay to detect contributions of others. Intuitively, free-riders are not willing to fund the centralized enforcer that could detect their behavior, but they are willing to detect other free-riders in a decentralized system. Second, despite differences in the detection and sanctioning environments, as well as the level of contributions to public goods, net payoffs are similar. Thus, net earnings are not improved from the automation of detection.

4.1 Detection

Given that this research focuses on modifications to the detection mechanism in the standard public goods game, we start with a discussion of the effect of the environment and detection regime on behavior. As we report in Table 1, the number of subjects detected in the exogenous environment is around 2.47, by design, for all detection-sanction combinations.[17] Exogenous automatic detection has been standard in the public goods literature. Implicitly, however, this research design is analogous to assuming that all residents in a community either help fund centralized detection (e.g., police) or individually detect the actions of their peers. However, citizens sort into communities with a higher willingness to pay for detection services [Tiebout, 1956, Fehr and Williams, 2013].

As has been noted in Hadfield and Weingast [2012, 2016], community detection is a critical

[17] In the standard public goods game, four subjects are always detected because detection is automatic, with 100% probability. In our setup we set the probability of detection at two-thirds, so we expect two-thirds of the four-person group to be detected (approximately 2.66 subjects) in the exogenous treatments. The actual average number of subjects detected was 2.47. This could be a problem if the average number detected differed by treatment. But pair-wise, t-tests of the average number detected by group are statistically insignificant between the four exogenous treatments (t-test values are always less than 1.16).

Table 1: **Summary Statistics by Treatment**

Treatment	**Public Contribution**	**Number Detected**	**WTP to Detect**	**Total Sanctions**	**Net Earnings**	**Obs.**
EXG CentralD-ConcentratedS	2.94	2.38	1.00	2.09	13.78	400
	(1.54)	(1.10)	(0.00)	(6.31)	(6.21)	
EXG CentralD-NeighborS	2.49	2.50	1.00	1.05	13.93	320
	(1.69)	(1.21)	(0.00)	(2.97)	(4.64)	
EXG PeerD-ConcentratedS	2.80	2.50	1.00	1.63	13.97	360
	(1.90)	(0.67)	(0.00)	(3.30)	(4.45)	
EXG PeerD-NeighborS	2.48	2.50	1.00	1.61	13.34	320
	(1.92)	(1.03)	(0.00)	(3.84)	(4.96)	
ENDG CentralD-ConcentratedS	1.73	0.04	0.34	0.04	14.41	360
	(1.82)	(0.42)	(0.47)	(0.69)	(3.71)	
ENDG CentralD-NeighborS	1.71	0.00	0.34	0.00	14.41	280
	(1.92)	(0.00)	(0.47)	(0.00)	(3.84)	
ENDG PeerD-ConcentratedS	2.71	0.86	0.25	1.32	14.85	320
	(1.86)	(0.80)	(0.43)	(3.36)	(4.67)	
ENDG PeerD-NeighborS	2.66	0.98	0.28	1.12	14.91	360
	(1.99)	(0.96)	(0.45)	(3.18)	(4.82)	

Note: This table reports the mean and standard deviation in parentheses. There were a total of 272 subjects in fixed groups of four who each played for 10 periods. Contributions could be between zero and five tokens. Number Detected could be between zero and four subjects per group. WTP to Detect is the proportion of subjects (0%–100%) who were willing to pay to detect. In the EXG environment, subjects automatically had to pay, so 100% of subjects paid. In the ENDG environment, subjects had a $1 Stage 2 supplement. In peer detection treatments, WTP to Detect is the proportion of subjects who paid $1 to detect their right-side group member. In centralized detection treatments, WTP to Detect is the proportion of subjects who pledged $1 toward a centralized detection regime, but if the group did not collect $4 this amount was refunded back. Total Sanctions is the cost of assigned sanctions added to received sanctions. Net Earnings represents the total earnings minus the costs of sanctions both assigned and received.

component in generating public goods (e.g., public safety). But a community may not have residents who are willing to incur the cost to detect others. Since we can only sanction those we detect, we focus on changes in detection when the choice to engage in detection is endogenous. Indeed, we see an 80% decline in the number of detected subjects when the choice to detect is made endogenous versus the more standard exogenous, automatic detection.[18] First, in the endogenous peer detection, regime subjects only pay the cost to detect their right-side group member 26% of the time, resulting in an average of 0.92 subjects being detected in each group. Second, when the choice to detect is endogenous, the central detection regime is implemented less than 1% of the time, resulting in an average of 0.02 people detected per group.[19] Thus, when detection is endogenous, we observe 63% and 98% decreases in centralized and peer detection, respectively.[20]

By introducing a centralized detection regime that requires unanimous commitment to detection, we open up the opportunity for a single person or small minority to veto implementation (e.g., a single juror insisting on innocence, a single faculty member refusing to reach consensus on a vote, or a permanent member of the UN Security Council exercising veto power). But, as aforementioned, data from pilot sessions show that lowering the need for support from unanimity to three out of four members nearly halves the number of people willing to pay for detection.[21] This implies that

[18]The average number detected is 2.47 in the exogenous treatments and 0.49 in the endogenous treatments. So, there is an 80% decline in the number of detected subjects ($\frac{2.47-0.49}{2.47} = 0.80$). This difference is statistically significant if we use one observation per group with 33 groups in exogenous treatments and 35 groups in the endogenous treatments. The two-sample t-test with equal variances is $t = 17.9082$ $Pr(|T| > |t|) = 0.0000$.

[19]There are a total of 16 groups in the endogenous central detection regime (nine in ENDG CentralD-ConcentratedS and seven in ENDG CentralD-NeighborS), and each plays 10 periods. The central detection regime was successfully implemented once out of 160 possible times.

[20]These differences are statistically significant if we use one observation per group. There are 17 groups in exogenous peer detection treatments (average detected, 2.5) and 17 groups in the endogenous peer detection treatments (average detected, 0.92). The two-sample t-test with equal variances is $t = 10.6050$ $Pr(|T| > |t|) = 0.0000)$. There are 18 groups in exogenous central detect treatments (average detected, 2.4) and 16 groups in the endogenous central detect treatments (average detected, 0.025) The two-sample t-test with equal variances is $t = 40.2590$ $Pr(|T| > |t|) = 0.0000)$.

[21]We ran 24 subjects in the CentralDetect-NeighborSanction treatment, with a threshold of three and 26 subjects in the CentralDetect-NeighborSanction treatment, with a threshold of four. These pilot sessions are not directly comparable to the data reported in this paper because we used randomly rematched groups in these pilots sessions but fixed groups in this paper. In our pilot sessions only 14% of subjects in the CentralDetect-NeighborSanction treatment, with a threshold of three, pledged to the information fund, while in the pilot sessions, with a threshold of four, 27% pledged to the information fund. This difference is statistically significant at the 12% level, using a random effects regression (results available from the authors upon request). Using randomly rematched groups, fewer people (27%, with $N = 360$ subject-period observations) pledged to the information fund than using fixed groups (34%, with $N = 280$ subject-period observations); but this difference is not statistically significant using a random effects regression ($z = -0.75$ $P > |z| = 0.451$).

if we were to loosen the unanimity requirement we would be even less likely to observe centralized detection implemented relative to our current setup.

4.1.1 Who Pays to Detect?

When the choice to detect is endogenous, we might suspect that subjects engaged in the most free-riding would be the least willing to pay for detection. Intuitively, those who do not contribute to the public good would seem unlikely to pay to hire agents to identify free-riders. This is precisely what we observe in the right column of Figure 3, as subjects who make smaller contributions to the public good are also less likely to pay for a centralized detection regime.[22]

However, when detection is decentralized such that each subject only detects their neighbor, no clear relationship between contributions to the public good and payment for detection exists. Those making large contributions to the public good might want to detect others to discern who is not contributing to encourage higher contributions, since everyone benefits. By the same logic, though, a free-rider would want to detect for the same reason—that is, because detecting other free-riders might encourage them (via sanctions) to contribute more to the public good. In addition to this reason, a free-rider might also want to detect other free-riders because detecting other free-riders may lower one's own received sanctions (see Freeman et al. [1996]). We see evidence that free-riders are willing to pay to detect in the left-hand column of Figure 3, which shows willingness to pay for detection by public contributions. In the left-hand column, there is not the simple upward sloping relationship, but, rather, both free-riders and large contributors are willing to pay for detection.[23]

[22] We confirm that a positive, linear relationship exists between contributions to the information fund and public goods contributions in the central detection treatments in appendix Table **??**. The average willingness to pay for detection is not statistically significantly different across the endogenous treatments if we use one observation per group, as our pair-wise t-tests are always less than 1.3425.

[23] In the ENDG PeerD-ConcentratedS treatment, we cannot observe significance in the linear or quadratic relationship between public goods contributions and information fund contributions, whereas in the ENDG PeerD-NeighborS treatment, both the linear and quadratic terms are significant. In ENDG PeerD-NeighborS, the relationship between public goods contributions and willingness to pay for detection is negative for those contributing less than \$4 to the public good but positive for individuals contributing \$4 or \$5. See Appendix Table **??**.

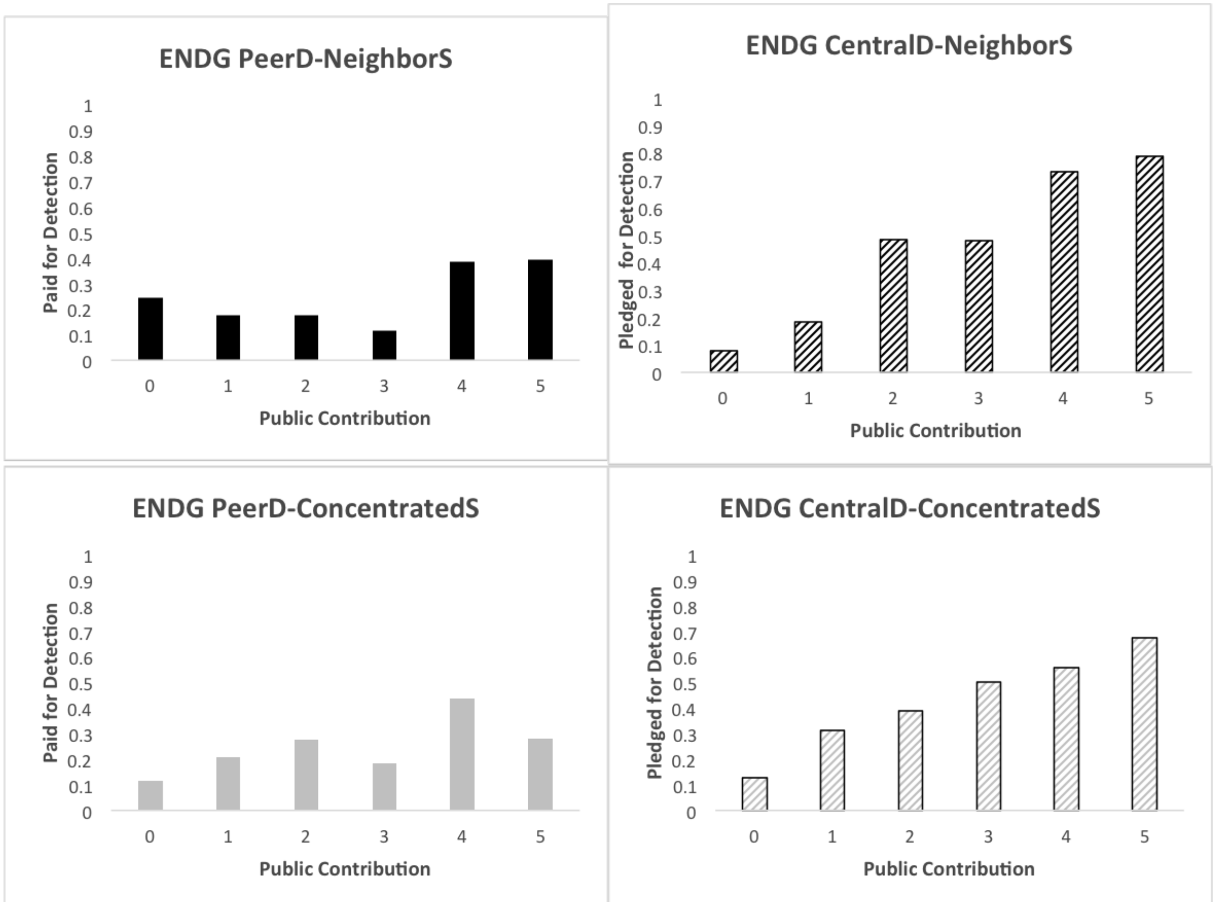


Figure 3: Average Proportion Who Paid/Pledged for Detection by Treatment (ENDG Environment only)

4.2 How Do Differences in Detection Change Outcomes?

Given that the environment (ENDG/EXG) as well as detection regime (centralized/decentralized) and sanction mechanism (concentrated/neighbor) lead to different choices about who is willing to pay to detect, we might expect sanctioning, contributions, and, therefore, net earnings to respond to these differences. We now explore these possibilities.

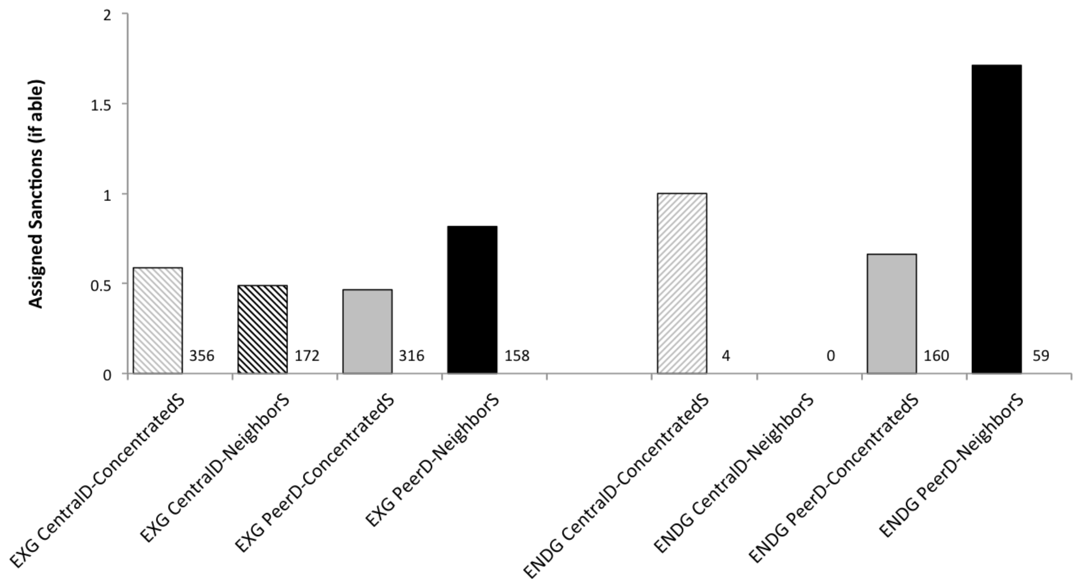
4.2.1 Sanctioning

As we saw in the previous section, the number of detected subjects declines precipitously when we endogenize the decision to detect. However, the economics of deterrence literature [Becker, 1968] points out that decreases in the likelihood of detection and sanctions have an interdependent relationship since deterrence depends on the expected sanction, not just the likelihood of detection or size of the sanction alone. Having examined the likelihood of detection, we now turn our attention to the level of sanctioning.

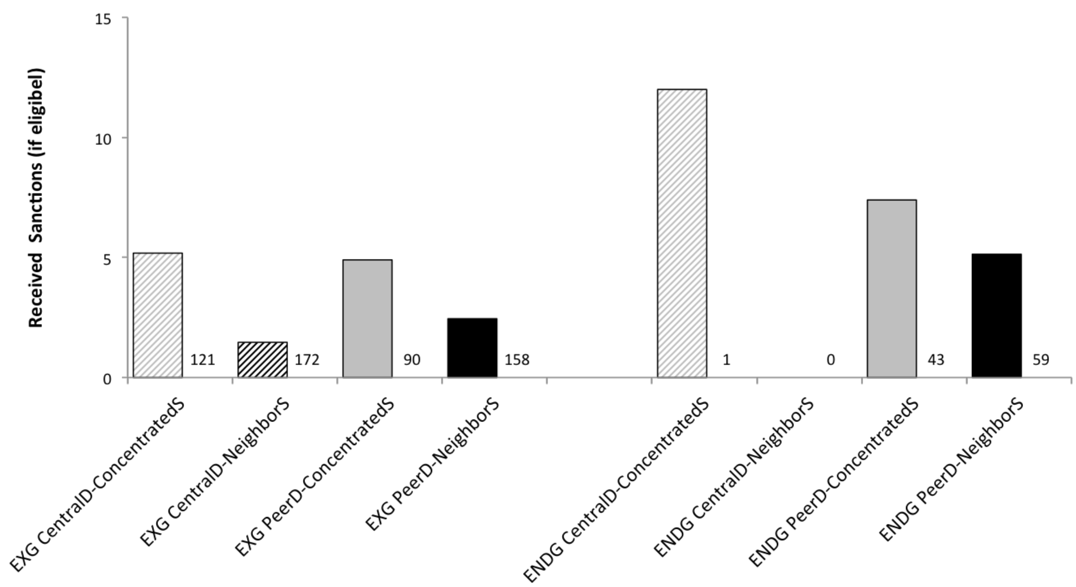
Figure 4 presents both the assigned and received sanctions across all of the treatments of the experiment. In the bottom panel of Figure 4, we find that received sanctions are always higher when penalties are concentrated (gray bars), compared to penalizing one's neighbor (black bars).[24] Interestingly, in the ENDG environment, the higher received sanctions conform to the economics of deterrence literature. Stated differently, since the probability of being detected is significantly lower in the ENDG treatments, we should expect higher sanctions in the ENDG rather than the EXG environment. Specifically, in the EXG environment, the likelihood of being eligible to sanction is 38.6%, while it is 7.8% in ENDG environment. The average sanctions received when eligible are 3.155 and 6.145, respectively, in the EXG and ENDG environments. Thus, the expected sanction in the EXG environment is 1.22, whereas the expected sanction in the ENDG environment is 0.48.[25]

[24]Two-way t-tests reported in Appendix Table 5, from the random effects regression reported in Appendix Table 4, show that these differences are statistically significant.

[25]This difference is statistically significant if we use one observation per group, with 35 groups in exogenous treatments and 33 groups in the endogenous treatments. The two-sample t-test with equal variances is $t = 3.4636$ $Pr(|T| > |t|) = 0.0005$.



(a) Assigned Sanctions (if Able to Assign)



(b) Received Sanctions (if Eligible to Receive)

Figure 4: Assigned and Received Sanctions by Treatment (With Null Values)

Note: Panel (a) shows the average per subject per period sanctions assigned when there are null values if a subject was not able to assign a sanction either because there was no detection or the recipient gave their full endowment to the public good. Panel (b) shows the average per subject per period sanctions received when there are null values if a subject was not eligible to receive a sanction either because there was no detection or the recipient gave their full endowment to the public good. The numbers to the right of each bar are the total observations of this type. In both Panel (a) and Panel (b), there are instances of the ability to assign/receive sanctions where a value of 0 is assigned/received.

4.2.2 Public Goods Contributions

Given that detection and sanctioning levels show considerable variation across treatments, we now examine public goods contributions across these contexts. Figure 5 shows the average public contributions by treatment (see Figure 6 in the Appendix for average public contributions by period). The right side of Figure 5 displays the endogenous detection treatments, while the left side examines the exogenous detection treatments, which are more typical in the public goods literature.

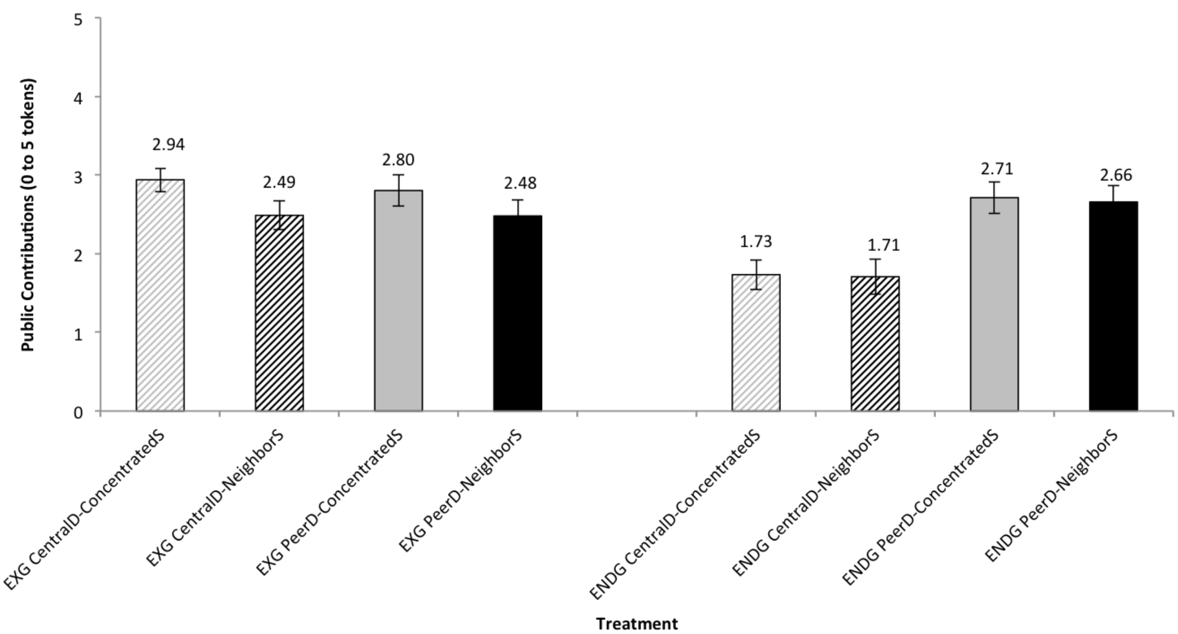


Figure 5: Average Contributions by Treatment and Environment

Note: In this figure we show the average per person contributions to the public good by treatment out of a possible five tokens and the 95% confidence interval. The four bars on the left represent treatments where detection only took place if subjects endogenously chose it by paying a $1 supplement. The four bars on the right represent treatments where detection was exogenously made automatic. A bar has a striped pattern if there was a centralized detection regime. A bar has gray shading if there is a centralized sanctioning mechanism.

Examining the right side of Figure 5, where detection is a costly endogenous choice, centralized detection leads to lower contributions than those with peer detection (the striped bars in Figure 5 are shorter than the solid bars).[26] This appears to occur because endogenous centralized detection leads to a very low probability of detection, as explained above.

When we examine the treatments with exogenously assigned detection, public goods contributions

[26]This difference is statistically significant if we use one observation per group with 17 groups in ENDG PeerD-NeighborS/PeerD-ConcentratedS treatments and 16 groups in ENDG CentralD-NeighborS/CentralD-ConcentratedS treatments. The two-sample t-test with equal variance is $t = 2.9690$ $Pr(|T| > |t|) = 0.0057$.

are not statistically different whether subjects experience the central or peer detection regime.[27] But when detection is exogenously automatic, public contributions are higher when there are centralized sanctions instead of neighbor sanctions (looking at the left four bars, the gray bars are higher than the black bars).[28] This is because fines received are higher with a concentrated sanctioning mechanism.

Overall, we find the public goods provision is similar under all peer detection regimes. Additionally, public goods provision is lowest when there is endogenous centralized detection.

4.2.3 Net Earnings

Thus far, we have identified significant differences in the likelihood of detection, sanctions, and public good provisions. While the mechanics (i.e., detection, sanctions, and contributions) of each of these environments and treatments are considerably different, subjects will adjust their behavior to account for these differences. Importantly, subjects adjust the amount that they sanction to account for a lower likelihood of being detected in the ENDG environment such that the expected sanctions are not statistically significantly different across environments. Additionally, the net earnings of our subjects are nearly identical.[29]

5 Conclusion and Discussion

Research on the optimal provision of public goods has focused on whether informal peer-to-peer sanctions or formal centralized sanctions should be utilized to discourage free-riding. While this is an important research endeavor, it makes strong assumptions concerning the likelihood of detection. Since a community can only sanction those persons that they catch behaving badly, attention to the detection mechanism is warranted. Most previous research has assumed an automatic and central

[27]This difference is statistically insignificant if we use one observation per group with 17 groups in EXG PeerD-NeighborS/PeerD-ConcentratedS treatments and 18 groups in EXG CentralD-NeighborS/CentralD-ConcentratedS treatments. The two-sample t-test with equal variances is $t = -0.3104$ $Pr(|T| > |t|) = 0.6209$.

[28]This difference is statistically significant at the 16% level if we use one observation per group with 16 groups in EXG PeerD-NeighborS/CentralD-NeighborS treatments and 19 groups in EXG PeerD-NeighborS/CentralD-NeighborS treatments. The two-sample t-test with equal variances is $t = -1.4213$ $Pr(|T| > |t|) = 0.1646$.

[29]Only ENDG PeerD-ConcentratedS ($14.85) and ENDG PeerD-NeighborS ($14.91) both have net earnings that are statistically significantly higher than those in EXG PeerD-NeighborS ($13.34). All other comparisons are not statistically significantly different. See Appendix Table 2 and Appendix Table 3.

detection technology that broadcasts information to the whole group. In this paper, we show that whether one can detect everyone's actions or just a close neighbor significantly effects contributions to the public good. Furthermore, the optimal detection-sanction pair varies considerably if the decision to detect is a costly endogenous choice, rather than being automatic.

When detection is exogenously automatic, then contributions to the public good are similar whether a person can detect only their closest neighbor or instead can see information about the whole group's actions through centralized detection. However, the story is very different when the choice to detect is made endogenous. Peer detection leads to increases in public contributions in the endogenous environment as subjects now have more autonomy over the use of their funds, leading only those subjects that desire detection and sanctioning to participate in the institution. However, contributions decrease in the centralized detection environment since the largest free-riders are unwilling to fund the centralized detection regime that might catch them misbehaving.[30]

Overall, we observe an 80% reduction in detection when the choice to detect is made endogenous versus the more standard automatic detection. This decline in detection may not lead to a decline in deterrence if sanctions are high enough, though. We find that concentrating sanctions leads to higher received sanctions. But those higher sanctions only consistently translate into higher contributions if a community can be counted on to detect free-riding since even the highest threatened sanction is nondeterrent if a free-rider knows that they will not be caught.

It is not advisable to centralize detection or concentrate sanctions when the community might let free-riding members go undiscovered. We have explored this question with only a few specific formulations of detection regimes and sanctioning mechanisms. We view these as stylized versions of any number of examples of these detection-sanctioning pairs. Further research into what specific attributes of a detection regime or sanctioning mechanism lead to higher provision of the public good is necessary.

An interesting ordering of what led to the highest provision of the public good emerged in our analysis. When exogenously imposing automatic detection, then centralized detection with centralized

[30] An alternative explanation for our findings would be that imposing centralized regulatory detection does not necessarily improve the welfare of the group, as predicted in Peltzman [1975].

sanctioning (EXG CentralD-ConcentratedS) generates higher contributions to the public good than peer detection and peer sanctioning (EXG PeerD-NeighorS), although the difference is not statistically significant. However, when we endogenize the detection decision, we find that ordering switches so that the centralized detection and concentrated sanctioning (ENDG CentralD-ConcentratedS) yields lower public goods contributions than peer detection and neighborhood sanctioning (ENDG PeerD-NeighborS). Stated differently, when the residents of a community always engage in detection, then a homeowner association that can detect and fine residents has better results than detecting your neighbor and writing nasty notes. But if that choice to detect is endogenous, then the ordering and effectiveness flips.

This flipping of ordering and effectiveness could make the decision of the social planner difficult. However, in either the exogenous or endogenous environment the peer detection with concentrated sanctioning combination (EXG/ENDG PeerD-ConcentratedS) results in high public contributions that are not statistically significantly different from the other successful combinations (EXG CentralD-ConcentratedS and ENDG PeerD-NeighborS).[31]

Contextual examples from citizen interactions provide evidence of this exact combination: Facebook, where users report the misconduct of other users to the central complaints department; universities, where individual students report on poor teachers to the administration; cities, where one can report a noisy neighborhood party to the police. Peer detection and concentrated sanctioning may be so popular because the organizer of a group (e.g., Facebook, universities, or cities) does not always know if their population will be willing to pay a cost to detect one another, and this combination of detection and sanctioning is robust to either exogenous or endogenous choice of detection. Thus, when a social planner is unsure if detection can be automated or will be by choice, peer detection coupled with centralized sanctioning is a relatively safe way to ensure public good provision. Work exploring the optimal combination of detection and sanctioning under uncertainty about the environment is an important area for further research.

[31] Contributions in the ENDG PeerD-ConcentratedS treatment are statistically significantly higher than those in the CentralD-ConcentratedS. Contributions in the EXG PeerD-ConcentratedS are not statistically significantly different from the other EXG treatments.

This research highlights that the common assumption of automatic detection of each group member's actions in the public goods game can lead to very different conclusions about what is optimal for public good provision relative to the scenario where we allow people to shirk the responsibility of detecting others. Centralizing detection or sanctions when the community might let free-riding members go undiscovered results in the underprovision of public goods, especially when compared to an environment where peer detection and central sanctioning is imposed.

References

Daron Acemoglu and Matthew O Jackson. History, expectations, and leadership in the evolution of social norms*. *The Review of Economic Studies*, page rdu039, 2014.

Daron Acemoglu and Alexander Wolitzky. Sustaining cooperation: Community enforcement vs. specialized enforcement. 2015.

Attila Ambrus and Ben Greiner. Imperfect public monitoring with costly punishment: An experimental study. *The American Economic Review*, 102(7):3317–3332, 2012.

Lisa Anderson, Gregory DeAngelo, Winand Emons, Beth Freeborn, and Hannes Lang. Penalty structures and deterrence in a two-stage model: Experimental evidence. *Economic Inquiry*, Forthcoming, 2017.

James Andreoni. Reasonable doubt and the optimal magnitude of fines: should the penalty fit the crime? *The RAND Journal of Economics*, pages 385–395, 1991.

James Andreoni and Laura K Gee. Gun for hire: delegated enforcement and peer punishment in public goods provision. *Journal of Public Economics*, 96(11):1036–1046, 2012.

James Andreoni and Laura K Gee. Gunning for efficiency with third party enforcement in threshold public goods. *Experimental Economics*, 18(1):154–171, 2015.

Masaki Aoyagi and Guillaume Frechette. Collusion as public monitoring becomes noisy: Experimental evidence. *Journal of Economic theory*, 144(3):1135–1165, 2009.

Delia Baldassarri and Guy Grossman. Centralized sanctioning and legitimate authority promote cooperation in humans. *Proceedings of the National Academy of Sciences*, 108(27):11023–11027, 2011.

Avner Bar-Ilan and Bruce Sacerdote. The response to fines and probability of detection in a series of experiments. *NBER Working Paper*, 2001.

Cesare Beccaria. 1986. *On crimes and punishments*, 1764.

Gary Becker. Crime and punishment: An economic approach. *Journal of Political Economy*, 76, 1968. URL http://EconPapers.repec.org/RePEc:ucp:jpolec:v:76:y:1968:p:169.

Luke A Boosey and R Mark Isaac. Network monitoring and punishment in public goods games. *Available at SSRN 2536526*, 2014.

Anabela Botelho, Glenn W Harrison, Lígia M Costa Pinto, and Elisabet E Rutström. Social norms and social choice. 2007.

Jeffrey P Carpenter. Punishing free-riders: How group size affects mutual monitoring and the provision of public goods. *Games and Economic Behavior*, 60(1):31–51, 2007.

Jeffrey P Carpenter, Shachar Kariv, and Andrew Schotter. Network architecture and mutual monitoring in public goods experiments. *The Review of Economic Design*, 2012.

Ananish Chaudhuri. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, 14(1):47–83, 2011.

Pedro Dal Bó and Guillaume R Fréchette. On the determinants of cooperation in infinitely repeated games: A survey. *Available at SSRN 2535963*, 2014.

Gregory DeAngelo and Gary Charness. Deterrence, expected cost, uncertainty and voting: Experimental evidence. *Journal of Risk and Uncertainty*, 44(1), 2012.

Gregory DeAngelo and Benjamin Hansen. Life and death in the fast lane: Police enforcement and roadway safety. *American Economic Journal: Economic Policy*, 6(2), 2014.

Gregory DeAngelo and Taylor Smith. Private security, piracy and the international provision of public goods. *Working Paper*, 2016.

Humphreys Brad DeAngelo, Gregory and Imke Reimers. Are public and private enforcement complements or substitutes evidence from high frequency nhl. *Working Paper*, 2016.

Joyee Deb and Julio González-Dıaz. Enforcing social norms: Trust-building and community enforcement. Technical report, Working paper, 2014.

David Dickinson and Marie-Claire Villeval. Does monitoring decrease work effort?: The complementarity between agency and crowding-out theories. *Games and Economic behavior*, 63(1):56–76, 2008.

Martijn Egas and Arno Riedl. The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society of London B: Biological Sciences*, 275(1637):871–878, 2008.

Arhan Ertan, Talbot Page, and Louis Putterman. Who to punish? individual decisions and majority rule in mitigating the free rider problem. *European Economic Review*, 53(5):495–511, 2009.

Marco Faillo, Daniela Grieco, and Luca Zarri. Legitimate punishment, feedback, and the enforcement of cooperation. *Games and Economic Behavior*, 77(1):271–283, 2013.

Josef Falkinger, Ernst Fehr, Simon Gächter, and Rudolf Winter-Ebmer. A simple mechanism for the efficient provision of public goods: Experimental evidence. *The American Economic Review*, 90 (1):247–264, 2000.

Enrique Fatas, Miguel A Meléndez-Jiménez, and Hector Solaz. An experimental analysis of team production in networks. *Experimental Economics*, 13(4):399–411, 2010.

Ernst Fehr and Simon Gachter. Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994, September 2000. doi: 10.1257/aer.90.4.980. URL http://www.aeaweb.org/articles?id=10.1257/aer.90.4.980.

Ernst Fehr and Simon Gächter. Altruistic punishment in humans. *Nature*, 415(6868):137–140, 2002.

Ernst Fehr and Tony Williams. Endogenous emergence of institutions to sustain cooperation. 2013.

Urs Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178, 2007.

Sven Fischer, Kristoffel R Grechenig, and Nicolas Meier. Cooperation under punishment: Imperfect information destroys it and centralizing punishment does not help. *MPI Collective Goods Preprint*, (2013/06), 2013.

Scott Freeman, Jeffrey Grogger, and Jon Sonstelie. The Spatial Concentration of Crime. *Journal of Urban Economics*, 40(2):216–231, September 1996. URL https://ideas.repec.org/a/eee/juecon/v40y1996i2p216-231.html.

Lana Friesen. Certainty of punishment versus severity of punishment: an experimental investigation. *Southern Economic Journal*, 79(2), 2012.

Drew Fudenberg, David G Rand, and Anna Dreber. Slow to anger and fast to forgive: Cooperation in an uncertain world. *American Economic Review*, 102(2):720–749, 2012.

Simon Gächter, Elke Renner, and Martin Sefton. The long-run benefits of punishment. *Science*, 322 (5907):1510–1510, 2008.

Kristoffel Grechenig, Andreas Nicklisch, and Christian Thöni. Punishment despite reasonable doubta public goods experiment with sanctions under uncertainty. *Journal of Empirical Legal Studies*, 7 (4):847–867, 2010.

Avner Greif. Contract enforceability and economic institutions in early trade: The maghribi traders' coalition. *The American economic review*, pages 525–548, 1993.

Avner Greif. Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies. *Journal of political economy*, pages 912–950, 1994.

Ben Greiner. The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics. Working Paper Series in Economics 10, University of Cologne, Department of Economics, June 2004. URL https://ideas.repec.org/p/kls/series/0010.html.

Jeffrey Grogger. Certainty vs. severity of punishment. *Economic Inquiry*, 29(2):297–309, 1991. ISSN

1465-7295. doi: 10.1111/j.1465-7295.1991.tb01272.x. URL http://dx.doi.org/10.1111/j.1465-7295.1991.tb01272.x.

Jeffrey Grogger. Certainty vs. severity of punishment. *Economic Inquiry*, 29(2), 2007.

Gillian K Hadfield and Barry R Weingast. What is law? a coordination model of the characteristics of legal order. *Journal of Legal Analysis*, 4(2):471–514, 2012.

Gillian K Hadfield and Barry R Weingast. Is rule of law an equilibrium without private ordering? *Available at SSRN*, 2016.

Alon Harel and Alon Klement. The economics of shame: Why more shaming may deter less. *Working Paper*, 2005.

Thomas Hobbes. *Leviathan, 1651.* Scolar Press, 1969.

Kenju Kamei, Louis Putterman, and Jean-Robert Tyran. State or nature? endogenous formal versus informal sanctions in the voluntary provision of public goods. *Experimental Economics*, 18 (1):38–65, 2015.

S. Karstedt-Henke. Attribution theory and deterrence research: A new approach to old problems. *Developments in Crime and Crime Control Research*, pages 22–40, 1991.

Jonathan Klick and Alexander Tabarrok. Using terror alert levels to estimate the effect of police on crime. *Journal of Law and Economics*, XLVIII, 2005.

Steven Levitt. Using electoral cycles in police hiring to estimate the effect of police on crime. *American Economic Review*, 87(3), 1997.

Thomas Markussen, Louis Putterman, and Jean-Robert Tyran. Self-organization for collective action: An experimental study of voting on sanction regimes. *The Review of Economic Studies*, 81(1): 301–324, 2014.

Nikos Nikiforakis and Hans-Theo Normann. A comparative statics analysis of punishment in public-good experiments. *Experimental Economics*, 11(4):358–369, 2008.

Rick O'Gorman, Joseph Henrich, and Mark Van Vugt. Constraining free riding in public goods games: designated solitary punishers can sustain human cooperation. *Proceedings of the Royal Society of London B: Biological Sciences*, 276(1655):323–329, 2009.

Elinor Ostrom. *Governing the commons*. Cambridge university press, 2015.

Elinor Ostrom, James Walker, and Roy Gardner. Covenants with and without a sword: Self-governance is possible. *American political science Review*, 86(02):404–417, 1992.

Talbot Page, Louis Putterman, and Bruno Garcia. Getting punishment right: Do costly monitoring or redistributive punishment help? *Available at SSRN 1084064*, 2008.

Sam Peltzman. The effects of automobile safety regulation. *Journal of Political Economy*, 83(4): 677–725, 1975. ISSN 00223808, 1537534X. URL http://www.jstor.org/stable/1830396.

A Mitchell Polinsky and Daniel L Rubinfeld. A model of optimal fines for repeat offenders. *Journal of Public Economics*, 46(3):291–306, 1991.

A Mitchell Polinsky and Steven Shavell. The optimal use of fines and imprisonment. *Journal of Public Economics*, 24(1):89–99, 1984.

Doron Teichman. Sex, shame, and the law: An economic perspective on megan's law. *Harvard Journal on Legislation*, 42(2), 2005.

Charles M. Tiebout. A Pure Theory of Local Expenditures. *Journal of Political Economy*, 64:416–416, 1956. doi: 10.1086/257839. URL https://ideas.repec.org/a/ucp/jpolec/v64y1956p416.html.

Arne Traulsen, Torsten Röhl, and Manfred Milinski. An economic experiment reveals that humans prefer pool punishment to maintain the commons. *Proceedings of the Royal Society of London B: Biological Sciences*, page rspb20120937, 2012.

Toshio Yamagishi. The provision of a sanctioning system as a public good. *Journal of Personality and social Psychology*, 51(1):110, 1986.

6 Appendix

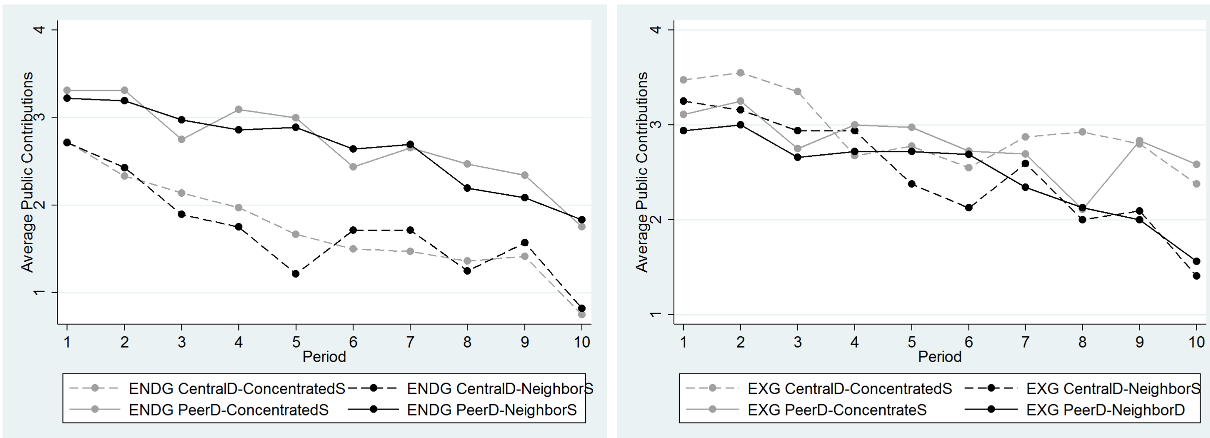


Figure 6: Average Contributions per Period by Treatment

Table 2: Treatment Effects on Contributions and Net Earnings

	(1) Contributions	(2) Net Earnings
ENDG CentralD-ConcentratedS	−0.74+	1.07
	(0.39)	(0.80)
ENDG CentralD-NeighborS	−0.77*	1.08
	(0.37)	(0.75)
ENDG PeerD-ConcentratedS	0.24	1.51+
	(0.37)	(0.79)
ENDG PeerD-NeighborS	0.18	1.58
	(0.41)	(0.96)
EXG CentralD-ConcentratedS	0.46	0.44
	(0.35)	(0.77)
EXG CentralD-NeighborS	0.01	0.59
	(0.28)	(0.81)
EXG PeerD-ConcentratedS	0.33	0.63
	(0.41)	(0.91)
Period	−0.14***	−0.26***
	(0.01)	(0.03)
Intercept	3.24***	14.77***
	(0.23)	(0.51)
overall R^2	0.09	0.03
N	2720	2720

EXG Peer-Peer is the omitted category. Sanctions are assigned sanctions plus received sanctions (note that four multiplied by assigned sanctions equals total sanctions and that three multiplied by assigned sanctions equals received sanctions). Random effects regression with period controls and standard errors clustered at the group level are reported (68 clusters). *** $p < 0.001$, ** $p < 0.001$, * $p < 0.05$, + $p < 0.1$.

Table 3: t-Tests for Treatment Effects on Contributions and Net Earnings

	ENDG CentralD-ConS	ENDG CentralD-NeighborS	ENDG PeerD-ConS	ENDG PeerD-NeighborS	EXG CentralD-ConS	EXG CentralD-NeighborS	EXG PeerD-ConS	EXG PeerD-NeighborS
ENDG CenD-ConS	-, -		c	c	c	c	c	c
ENDG CenD-NeighborS		-, -	c	c	c	c	c	c
ENDG PeerD-ConS	c	c	-, -					n
ENDG PeerD-NeighborS	c	c		-, -				n
EXG CenD-ConS	c	c			-, -			
EXG CenD-NeighborS	c	c				-, -		
EXG PeerD-ConS	c	c					-, -	
EXG PeerD-NeighborS	c	c	n	n				-, -

This table presents the results from two-way t-tests on the coefficients from the random effects regression reported in Table 2. A letter c denotes if the public goods contributions are statistically significantly different at the the 10% level. A letter n denotes if net income is statistically significantly different at the the 10% level. A dash denotes a comparison is not possible.

Table 4: Treatment Effects on Sanctions

	(1) Total Sanctions	(2) Assigned Sanctions	Received Sanctions
ENDG CentralD-ConcentratedS	−1.57***	0.17	9.60***
	(0.24)	(0.15)	(0.49)
ENDG CentralD-NeighborS	−1.61***	.	.
	(0.23)	.	.
ENDG PeerD-ConcentratedS	−0.29	−0.19	5.27***
	(0.32)	(0.20)	(1.59)
ENDG PeerD-NeighborS	−0.49	0.74**	1.96**
	(0.40)	(0.23)	(0.76)
EXG CentralD-ConcentratedS	0.48	-0.24	2.78
	(0.71)	(0.29)	(2.35)
EXG CentralD-NeighborS	−0.56	−0.40+	−1.37+
	(0.44)	(0.24)	(0.76)
EXG PeerD-ConcentratedS	0.02	−0.39*	2.33*
	(0.35)	(0.18)	(0.96)
Period	−0.02	−0.01	−0.13
	(0.02)	(0.01)	(0.08)
Intercept	1.70***	0.94***	3.41***
	(0.26)	(0.19)	(0.81)
overall R^2	0.03	0.08	0.03
N	2720	1225	644

EXG Peer-Peer is the omitted category. Total sanctions are assigned sanctions plus received sanctions. Random effects regression with period controls and standard errors clustered at the group level are reported. (Column 1 has 68 clusters. Columns 2 and 3 have 51 clusters.) *** $p < 0.001$, ** $p < 0.001$, * $p < 0.05$, + $p < 0.1$.

Table 5: t-Tests for Treatment Effects on Sanctions (Total, Assigned, and Received)

	ENDG CentralD-ConS	ENDG CentralD-NeighborS	ENDG PeerD-ConcS	ENDG PeerD-NeighborS	EXG CentralD-ConS	EXG CentralD-NeighborS	EXG PeerD-ConcS	EXG PeerD-NeighborS
ENDG CenD-ConS	-,-,-	-,-,-	t, r, a	t, r, a	t, r, a	t, r, a	t, r, a	t, r
ENDG CenD-NeighborS	-,-,-	-,-,-	t, -, -	t, -, -	t, -, -	t, -, -	t, -, -	t, -, -
ENDG PeerD-ConS	t, r, a	t, -, -	-,-,-	r, a		r	r	r
ENDG PeerD-NeighborS	t, r, a	t, -, -	r, a	-,-,-	a	r, a	a	r, a
EXG CenD-ConS	t, r, a	t, -, -		a	-,-,-	r		
EXG CenD-NeighborS	t, r, a	t, -, -	r	r, a	r	-,-,-	r	r, a
EXG PeerD-ConS	t, r, a	t, -, -	r	a		r	-,-,-	r, a
EXG PeerD-NeighborS	t, r	t, -, -	r	r, a		r, a	r, a	-,-,-

This table presents the results from two-way t-tests on the coefficients from the random effects regression reported in Table 4. A letter t denotes if the total sanctions are statistically significantly different at the the 10% level. A letter a denotes if assigned sanctions is statistically significantly different at the the 10% level. A letter r denotes if received sanctions is statistically significantly different at the the 10% level. A dash denotes a comparison is not possible.