# Peers or Police?: The Effect of Choice and Type of Monitoring in the Provision of Public Goods*

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

[1]Claremont Graduate University, Economics (gregory.deangelo@cgu.edu)
[2]Tufts University, Economics (laura.gee@tufts.edu)

July 9, 2020

(Keywords: Public Goods, Punishment, Monitoring, Feedback, Deterrence)

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

**Abstract**

Punishments are meant to deter bad acts, but we commonly only punish those we catch. At some point in time, a society must choose *how* to catch bad acts. We explore the effect of *how* we catch bad acts on public good provision. Specifically, we contrast monitoring done by peers (e.g., a neighbor reporting illegal dumping of waste) to that done by an organized group (e.g., the police patrolling for crime). We find that when either type of monitoring is exogenously imposed, both peer and group monitoring lead to similar levels of public good provision. However, when monitoring is an endogenous choice, societies fail to implement group monitoring, resulting in a 44% drop in public good provision. In contrast, peer monitoring results in similar levels of public good provision when either endogenously chosen or exogenously imposed. As such, if the willingness of a society to monitor is unknown, it may be safer to rely on peer rather than group monitoring.

*Thanks to Michael Kuhn for his comments on the experimental design. We would also like to thank participants at the Social Dilemmas Workshop, the New England Experimental Economics Workshop, the ESA conference, and the Appalachian State seminar, as well as Kristy Buzard, David Huffman, Sarah Jacobson, Michael Makowsky, Louis Putterman, Abhijit Ramalingam, Rustam Romaniuc, the anonymous referees and editors. Thanks also to Abdul Haleem Kidwai, Emma Grazier, and Irene Mussio for excellent research assistance. This research was approved by the Tufts IRB (1408019).

# 1 Introduction

Punishments are meant to deter bad acts. However, we commonly only punish those who are caught. We issue speeding tickets to drivers caught by the police, and we complain to neighbors when we notice their illegal dumping of waste. When speeding is done while the police are otherwise engaged, or if your neighbor is negligent when you are out of town, these bad acts go unmonitored. And when we do not monitor bad acts, we do not usually punish them. Monitoring is likely a necessary but not a sufficient condition for punishments to deter bad acts.

How monitoring becomes a feature of an institution can vary and is usually an endogenous choice at some point in time. Imagine a neighborhood where each household is responsible for safely disposing of their waste, yet some homes have piles of garbage nearby. The community members could form a patrol to police everyone's actions or instead rely on peers to report noncompliance. Both group and peer monitoring exist in different settings, so what is the right level of monitoring for the specific setting of safe waste disposal in a neighborhood? Rather than concentrating on a specific setting, this paper uses a laboratory experiment to address the question of whether peer monitoring is superior to group monitoring, and if the answer varies by whether monitoring is endogenously chosen or exogenously imposed.

Historically, monitoring has often started at the peer level and then either evolved into a more formal centralized institution or instead remained mostly at the peer level. For example, early policing was entirely conducted by individuals known as "night-watches" who detected gambling or prostitution offenses but evolved into centralized law enforcement agencies that are now publicly funded through taxes.[1] In contrast, in the early days of eBay, users sent complaints directly to the founder, and to this day eBay still relies heavily on the peer feedback forum to surface bad transactions [Cohen, 2003]. Knowing why some groups rely on centralized monitoring, like the police, while others detect bad acts through peer monitoring, like eBay user feedback, can help inform which type of monitoring a group will benefit most from.

Peer monitoring and group monitoring differ in many ways, and those differences may be particularly important when monitoring is an endogenous choice. Peer monitoring does not require group agreement and may be more likely to be implemented when a group's members are heterogeneous, with a mixture of bad actors and good citizens. When a peer reports illegal dumping or an inappropriate transaction on eBay, that peer acts alone. But

[1] For a comprehensive review of the origins of policing, see https://time.com/4779112/police-history-origins/.

when monitoring is assigned to a group agency, it may require a collective decision, which opens up the opportunity for a single individual or subset of society to veto implementation. Under endogenous group monitoring, a subset of citizens could vote against sufficient taxes to adequately staff patrols that identify illegal dumping, just as a single member country of the UN Security Council could exercise a veto to block a resolution.[2] A homogeneous group of all good citizens has little to gain from monitoring, while a heterogeneous group with some bad actors might see larger gains from monitoring. One could imagine that the groups least likely to endogenously form a centralized monitoring agency might also be the groups that are most likely to benefit from monitoring. The reason is intuitive. As groups have more bad actors, these bad actors want to avoid detection and punishments.

Exploring these questions with observational data is hard due to the endogenous adoption of different types of monitoring and the difficulty in measuring all instances of monitoring and free riding behavior. It would be tough to find a data set that includes free riding, like measuring all instances of illegal dumping, when it is not officially monitored by anyone. Controlled laboratory experiments provide a fruitful environment for examining how monitoring affects public goods provision without these measurement issues.

However, much of the previous experimental work in this area has exogenously imposed monitoring that is automatic, accurate, and centralized. Yet, outside of the laboratory, monitoring must at some point arise endogenously. To that end, studies that have exogenously imposed monitoring may fail to capture some of the benefits of peer monitoring, namely that it might be easier to adopt than group monitoring. We build off the current literature in exploring optimal provision of public goods by no longer assuming that a society will always be able to organize and fund an all-seeing monitor that broadcasts feedback to everyone.

To demonstrate the potential for research in this area, we use a two-by-two between-subjects design. In all our games, subjects play a linear public goods game for 10 periods. In each period, the subjects must divide an endowment between a public good that pays returns to all group members and a private good that pays returns only to that individual group member. We begin with an exogenously imposed group monitoring regime because that is closest to the bulk of the previous work about public goods provision. Then we vary whether the cost of the group monitoring is automatically imposed or is an endogenous choice. Next, we do an analogous comparison of exogenous versus endogenous peer monitoring.

We show that when monitoring is exogenously imposed, contributions to the public good are similar under either peer or group monitoring. However, this masks an important dif-

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

ference in the efficacy of the two monitoring regimes when we endogenize monitoring. Peer monitoring results in similar levels of public good provision whether monitoring is endogenous or exogenous. However, public good provision drops by 44% when we endogenize group monitoring.

The reason for this decline is that under group monitoring, a subset of subjects often are unwilling to pay for group monitoring, which leads to almost no one being monitored or punished for bad acts. There is also a strong correlation between contributing to the public good and willingness-to-pay for group monitoring, which implies that free riders do not want to fund a group monitoring regime that will catch their own bad acts. However, under peer monitoring there is no strong correlation between free riding behavior and willingness-to-pay for monitoring. Intuitively, people who illegally dump their waste are not willing to fund a patrol force that will detect their bad acts. Yet, these same people might report a neighbor's bad acts to draw attention away from their own wrongdoing. While our models and games are very specific formulations, we argue that our results are suggestive of a common real-world experience: when the willingness of a society to monitor bad acts is unclear, it may be safer to depend on peer monitoring rather than hope for agreement on group monitoring.

# 2 Background

Monitoring regimes arise endogenously at some point in time, yet much of the work on the effects of monitoring comes from public goods lab experiments where a subject's actions are automatically monitored and broadcast to the whole group. This common setup assumes that monitoring is automatic, done by a group agency, accurate, and immediate. However, relaxing these assumptions about monitoring may lead to different outcomes. For example, relaxing the assumption that monitoring is accurate by making monitoring an imperfect noisy signal of true actions leads to more free riding.[3] The evidence is mixed about the effect of lowering the probability of monitoring below 100%,[4] though relaxing the assumption that

[3]This is true whether there are no punishments [Aoyagi and Frechette, 2009], peer-to-peer punishments [Grechenig et al., 2010, Ambrus and Greiner, 2012], or centralized punishments [Fischer et al., 2016].

[4]If only information about actions is revealed by monitoring, as in our setup, when the probability of monitoring falls below 100%, Andreoni and Petrie [2004] find no effect on public contributions, while Friesen [2012], DeAngelo and Charness [2012], and Anderson et al. [2017] find a decrease in bad acts. However, when the likelihood of revelation of action and identity of the person rises, this generally leads to less free riding [Spraggon et al., 2015, Rege and Telle, 2004, Andreoni and Petrie, 2004], while Noussair and Tucker [2007] finds the opposite relationship. Outside the laboratory, decreased detection leads to more running of red lights [Bar-Ilan and Sacerdote, 2004], speeding on roadways [DeAngelo and Hansen, 2014], and tax evasion [Alm et al., 2012].

monitoring is immediate does not seem to affect outcomes unless communication is also allowed [Cason and Khan, 1999]. Thus it is not clear which of the most commonly used monitoring setup assumptions (accuracy, immediacy, group level, etc.) affect outcomes and which do not.

One could relax these assumptions one by one, but the level of monitoring (peer versus group) seems particularly likely to matter given the well-developed literature about the optimal level for levying punishments.[5] For example, a handful of papers contrast various exogenously imposed monitoring regimes in the lab.[6] Fatas et al. [2010] and Faillo et al. [2013] find that group monitoring leads to greater contributions than some alternative monitoring regimes. Alternatively, Boosey and Isaac [2016] find no difference between a group monitoring/punishment setup and the other monitoring/punishment setups they investigate. In the setup most closely related to our paper, Carpenter et al. [2012] find no difference in contributions between group monitoring paired with peer-to-peer punishments and peer monitoring with slightly constrained punishments. With such mixed results, it may be that when monitoring is exogenously imposed, there is not a systematic effect that depends on whether monitoring is performed at the group or peer level.

All the aforementioned papers exogenously impose a monitoring regime, which is like assuming the members of a society will always put in the effort to implement monitoring. Our paper is one of the first to contrast the effect of endogenizing a specific monitoring regime in a public goods game. Others have shown that subjects will pay to monitor others and that monitoring when coupled with punishments improves public good provision.[7] But these papers do not contrast endogenous monitoring with exogenous monitoring. To get an idea of the effect of endogenizing monitoring, we can look to the rich literature on the effects of endogenizing punishment.

When looking at endogenous punishment, most papers find that societies that endogenously select into a punishment mechanism have higher contributions to the public good

[5]Both peer-to-peer punishments [Egas and Riedl, 2008, Gächter et al., 2008, Fehr and Gächter, 2002, Fehr and Gachter, 2000, Fuster and Meier, 2010, Ostrom et al., 1992], where every person in a group can punish any other person, and centralized/concentrated punishments [Baldassarri and Grossman, 2011, Dickinson and Villeval, 2008, Falkinger et al., 2000, Yamagishi, 1986] have been shown to increase public contributions relative to no punishments. See Andreoni and Gee [2012] for a discussion of peer versus centralized punishments.

[6]We define *monitoring* to mean simply observing the actions of another subject but not necessarily doing anything in reaction to what is observed. Consider the household item of the baby monitor; if one observes the bad act of a baby waking at 2 a.m., one does not always (perhaps never) punish that bad act. Others have used the term *feedback* or *detection* to describe what we call monitoring.

[7]Nicklisch et al. [2019] explore monitoring in a public goods game with punishments, while Moir [2007] and Casari and Plott [2003] instead use a common pool resource game.

versus when that same mechanism is exogenously imposed.[8] But if the societies that did and did not select into endogenous punishment are pooled together, then endogenizing has been found to lead to lower contributions.[9] It may be that societies that do not endogenously select into punishment are the societies with people who most want to avoid being punished. Likewise, this hints that societies that are unable to endogenously select into monitoring may be societies with at least some bad actors.

The current literature suggests that when monitoring is endogenous, only those societies with relatively few bad actors might successfully implement it. These selection effects may be particularly likely when implementing monitoring requires group agreement. As such, when the appetite for endogenous monitoring is unknown, a society may be better off with peer rather than group monitoring. In the next section, we describe the specific games we use to test this line of reasoning.

# 3 The games

The main results come from a two-by-two experiment that varies (i) whether monitoring is exogenously automatic or endogenously chosen and (ii) whether monitoring is done by the group or by peers. Subjects are assigned to one of these four monitoring games. We begin by discussing the exogenous group monitoring game because it bears the most resemblance to previously studied games. All four games consist of a contribution stage, monitoring stage, and punishment stage.

## 3.1 The exogenous group monitoring (EXG GroupMonitor) game

### 3.1.1 EXG GroupMonitor: Contribution Stage

All four of our games start with a contribution stage, where subjects are given an endowment of five tokens that they then allocate between a public good and a private good. Each token

[8]For example, Andreoni and Gee [2012] find that subjects contribute 95% of their endowment when they endogenously select the gun-for-hire punishment mechanism versus 91% when it is exogenously imposed. A similar pattern can be observed in Fehr and Williams [2018], Ramalingam et al. [2016], Dal Bó et al. [2010], and Tyran and Feld [2006]. There are also many papers looking specifically at what type of punishment mechanisms are chosen but without contrasting that with the effects of exogenously imposing those same mechanisms (see Nicklisch et al. [2016], Kamei et al. [2015], Markussen et al. [2014], Traulsen et al. [2012], Ertan et al. [2009]).

[9]For example, Andreoni and Gee [2012] find if they pool those who did and did not select the hired gun, then subjects contribute 83% of their endowment when punishment is endogenously selected versus 91% when it is exogenously imposed.

contributed to the public good pays a return of \$1 to all four group members for an aggregate social return of \$4. Each token invested in the private good pays a return of \$2 to only the individual who made the investment. Let $g_i$ be subject $i$'s contribution to the public good. The earnings for the contribution stage are as follows:

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

In all our games, subjects make contributions to the public good without knowing if they will be monitored or punished because bad acts are generally committed before knowing if the crime will go unobserved and unpunished. At the end of the contribution stage, the subject is not informed about other's actions because that is determined by the monitoring stage.

### 3.1.2 EXG GroupMonitor: Monitoring Stage

In the monitoring stage, all the subjects are *eligible* to be monitored. However, each subject in the group has an independent probability of *actually* being monitored. If a subject is actually monitored, then that subject's contribution to the public good, private good, and contribution stage payoff are broadcast to all subjects, as shown in the "Observed By" column of Panel I of Figure 1. A subject's actions might only be known to that subject or might be broadcast to all group members.

In most of the previously studied games, the *eligibility* for monitoring and probability of *actually* being monitoring are 100%; as such, each subject would observe exactly three other group members' actions. However, to compare exogenous and endogenous variation in the number monitored, we set the probability of *actually* being monitored to 66.6% in our games, so a subject may see zero, one, two, or all three group members' actions.

Because people do not tend to hide their prosocial actions, a subject who contributes the whole endowment to the public good is *actually* monitored with 100% probability. In our exogenous games, we give subjects a \$1 supplement that they are forced to pay toward group monitoring.[10] Since there are no payoff-relevant decisions, monitoring stage payoffs are equal to payoffs from the contribution stage:

$$\pi_i^1 = \pi_i^2.$$

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

### 3.1.3 EXG GroupMonitor: Punishment Stage

In the final stage, any subject can choose an amount $s_i$ to punish the person who is able to be punished.[11] In our games, the person who is able to be punished is the lowest contributor who was actually monitored, which is similar to the punishment mechanism in Yamagishi [1986].[12] Subjects choose $s_i$ in integer increments and pay \$1 per unit to punish the lowest monitored contributor by \$3 per unit. For example, in a group where two subjects were actually monitored with one person contributing zero tokens and the other person contributing one token, all punishments would be levied on the zero contributor.

Subjects may choose any value of $s_i$ that they can afford with their earnings from the earlier stages. If multiple subjects tie for the lowest monitored contribution, then the punishment is split evenly among them (e.g., if the group assigned \$6 of punishment and there are subjects tied for lowest, then a \$3 punishment would be levied on each of the two lowest contributors).[13] If the person who is able to be punished gave the full endowment, there are no punishments levied and punishment costs are refunded. This approximates the idea that enforcement does not punish people obeying the rules. Punishment-stage payoffs can be summarized as follows:

$$\pi_i^3 = \begin{cases} \pi_i^2 - s_i - 3\sum_j^N s_j & \text{if lowest monitored member} \\ \pi_i^2 - s_i & \text{if not lowest monitored member} \\ \pi_i^2 & \text{if no one monitored or all those monitored contribute fully} \end{cases}$$

In all our games after the punishment stage, subjects are shown their own contribution, initial payoff, costs of punishments that they assigned to the lowest monitored contributor,

[11] We allowed subjects to punish themselves, although it is hard to understand what might motivate such self-flagellating behavior. Subjects only punish themselves less than 1.2% of the time.

[12] Andreoni and Gee [2012] and Andreoni and Gee [2015] use the gun-for-hire mechanism, which also punishes the lowest monitored contributor. However, the gun-for-hire mechanism punishes an amount that is specified by the mechanism to change the own-profit-maximizing equilibria of the game toward greater public good provision. Since we are most interested in the effects of changes in monitoring, we wanted to use a punishment mechanism that kept free riding as the own-profit-maximizing equilibrium, and so we modeled our mechanism after Yamagishi [1986]. Furthermore, the gun-for-hire mechanism of Andreoni and Gee [2012] requires information about the two largest free riders in the group, and since we allow people to endogenously choose how many people will be monitored, it is not totally clear how the gun-for-hire mechanism should be adapted when information about only a single free rider is detected.

[13] Alternatively, we could have randomly chosen one of the subjects from all those who tied to be punished. Because subjects have been shown to have issues understanding compound lotteries [Halevy, 2007], we instead opted to evenly split punishments among those who tied with certainty. Ties were not very common, only occurring 11% percent of the time.

and costs of punishments that they received. Subjects are never explicitly informed of group contributions to the public good, although they could figure this out from the information that they are provided.[14]

**I. Example of EXG GroupMonitor**

| Subject | Contribution | WTP Monitor | Monitored | Observed By | Observes | Can Be Punished | Can Punish | Punished By |
|---|---|---|---|---|---|---|---|---|
| A | 0 | forced 1 | No | self | self, C, D | No, not observed | C | |
| B | 1 | forced 1 | No | self | self, C, D | No, not observed | C | |
| C | 1 | forced 1 | Yes | all subjects | self, D | Yes | C | A, B, C, D |
| D | 5 | forced 1 | Yes | all subjects | self, C | No, fully contributed | C | |

**II. Example of ENDG GroupMonitor**

*If monitoring threshold met*

| Subject | Contribution | WTP Monitor | Monitored | Observed By | Observes | Can Be Punished | Can Punish | Punished By |
|---|---|---|---|---|---|---|---|---|
| A | 0 | 1 | No | self | self, C, D | No, not observed | C | |
| B | 1 | 1 | No | self | self, C, D | No, not observed | C | |
| C | 1 | 1 | Yes | all subjects | self, D | Yes | C | A, B, C, D |
| D | 5 | 1 | Yes | all subjects | self, C | No, fully contributed | C | |

*If below monitoring threshold*

| Subject | Contribution | WTP Monitor | Monitored | Observed By | Observes | Can Be Punished | Can Punish | Punished By |
|---|---|---|---|---|---|---|---|---|
| A | 0 | 0 | No | self | self | No, not observed | | |
| B | 1 | 0 | No | self | self | No, not observed | | |
| C | 1 | 1 | No | self | self | No, not observed | | |
| D | 5 | 1 | No | self | self | No, not observed | | |

**III. Example of EXG PeerMonitor**

| Subject | Contribution | WTP Monitor | Monitored | Observed By | Observes | Can Be Punished | Can Punish | Punished By |
|---|---|---|---|---|---|---|---|---|
| A | 0 | forced 1 | No | self | self | No, not observed | C | |
| B | 1 | forced 1 | No | self | self, C | No, not observed | C | |
| C | 1 | forced 1 | Yes | self, B | self, D | Yes | C | A, B, C, D |
| D | 5 | forced 1 | Yes | self, C | self | No, fully contributed | C | |

**IV. Example of ENDG PeerMonitor**

| Subject | Contribution | WTP Monitor | Monitored | Observed By | Observes | Can Be Punished | Can Punish | Punished By |
|---|---|---|---|---|---|---|---|---|
| A | 0 | 0 | No | self | self | No, not observed | C | |
| B | 1 | 1 | No | self | self, C | No, not observed | C | |
| C | 1 | 1 | Yes | self, B | self, D | Yes | C | A, B, C, D |
| D | 5 | 0 | Yes | self, C | self | No, fully contributed | C | |

*Notes:* WTP is Wilingness-To-Pay for monitoring which is forced to 1 in the Exogenous (EXG) games.

Figure 1: Examples of information revealed and punishment by game

### 3.1.4 EXG GroupMonitor: An Example

In Panel I of Figure 1, we show an example of what might happen during the EXG GroupMonitor game. This Figure 1 is meant to illustrate the differences in information revealed and punishment across our main games.

[14]For example, if there is no payoff from the public good, it is clear that no one has contributed to the public good.

## 3.2 The endogenous group monitoring (ENDG GroupMonitor) game

The endogenous group monitoring (ENDG GroupMonitor) game is very similar to the exogenous group monitoring (EXG GroupMonitor) game.

### 3.2.1 ENDG vs. EXG GroupMonitor: What is different?

The only difference is that in the monitoring stage, subjects choose whether to pledge their $1 supplement toward a monitoring regime. Subjects know that they can only punish the group member with the lowest contribution from those members who are *actually* monitored. If the sum of the four-person group's pledges reaches a threshold, then a group monitoring regime is implemented and the group pays a total of $4 to make all four subjects *eligible* to be monitored. If the sum of pledges is below the threshold, then no one is eligible to be monitored and no one pays the $1 (and all who pledged are given a full refund).

### 3.2.2 ENDG GroupMonitor: How many pledges are needed?

In our main games, we require unanimity (four pledges) for the endogenous group monitoring games. This ensures that (i) we do not force anyone to pay for a monitoring regime they did not make a pledge toward and that (ii) we keep the costs of monitoring the same across group monitoring and peer monitoring, which we describe next.

Although unanimity is a common way for groups to make decisions, it certainly is not the only way. In the real world, it is not immediately obvious what would happen if one lowered the requirement from unanimity to, for example, majority rule. On the one hand, decreasing the threshold could increase the likelihood of group monitoring by lowering the number of pledges needed. On the other hand, decreasing the threshold could decrease implementation by creating coordination issues and making individuals feel less pivotal.[15] To ensure our results are not specific to the unanimity condition without introducing coordination issues, we also run a variant of the ENDG GroupMonitor game where we only require three out of four subjects to be willing-to-pay for group monitoring, but we still require all four subjects to pay for monitoring. We compare the unanimous and nonunanimous versions in section 6.1.[16]

[15] Gee and Schreck [2018] present a theoretical model and lab/field experiment results that illustrate how raising the number needed to reach a goal can increase participation rates through increasing feelings of being pivotal. See Agranov et al. [2017] for similar findings in the arena of elections and a more thorough summary of the related literature.

[16] In a previous version of this paper, we reported data from pilot sessions with a nonunanimous voting

### 3.2.3 ENDG GroupMonitor: An Example

In Panel II of Figure 1, we show an example of what might happen during the ENDG GroupMonitor game. Subjects choose whether to make $1 pledges toward a monitoring regime rather than being forced to pay as in the EXG GroupMonitor game. In the section of Figure 1 labeled "if monitoring threshold met," we see that everyone is eligible to be monitored, but only persons C and D are actually monitored. Person C is the lowest contributor who is observed, and all punishments are levied upon Person C. In contrast in the section labeled "if below monitoring threshold," we find that no one is eligible to be monitored, actually monitored, or punished.

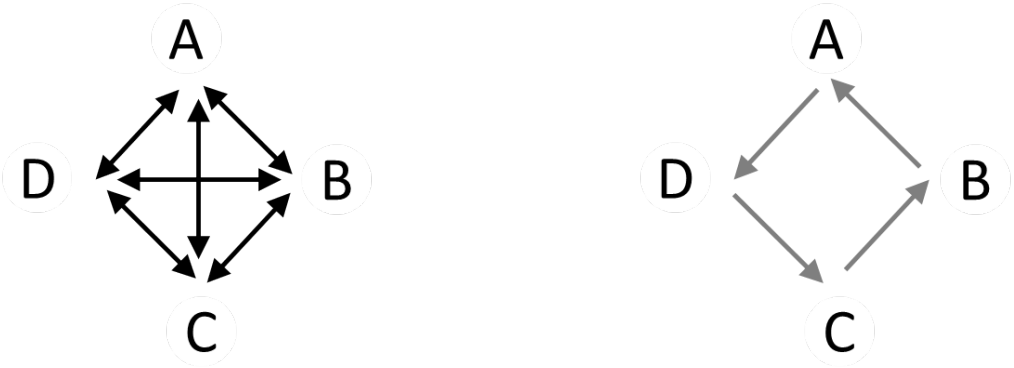
## 3.3 The exogenous peer monitoring (EXG PeerMonitor) game

The exogenous peer monitoring game (EXG PeerMonitor) begins with a linear public goods game that is exactly the same as the one in both the EXG GroupMonitor and ENDG GroupMonitor games. In the monitoring stage, subjects are forced to pay their $1 supplement to make a single other group member eligible to be monitored.

### 3.3.1 PeerMonitor vs. GroupMonitor: What is different?

The main difference between our GroupMonitor and PeerMonitor games is how information flows within the group. In the EXG PeerMonitor game, subjects are forced to pay a fee to make the group member to their right eligible to be monitored. That means that subject A learns the actions of B, B learns the actions of C, C learns the actions of D, and D learns the actions of A, so information flows as shown in panel (b) of Figure 2 when everyone in the group is actually monitored. This type of information flow approximates how people often only monitor those they interact with most (e.g., observing the actions of your immediate neighbors).

rule. The pilot game differs from the game presented in this paper in two main ways: (1) the groups were randomly rematched each period, and (2) we required that only three-of-four pay for group monitoring. In these pilot sessions, we found that requiring only three-of-four to pledge/pay to implement the GroupMonitor regime decreased the number who pledged (results available from authors upon request). In contrast, using the more comparable data reported in section 6.1, we do not find any statistically significant difference in willingness-to-pay for monitoring by unanimity versus nonunanimity.




(a) Group Monitor (b) Peer Monitor

Figure 2: Information flow in group versus peer monitoring if everyone in group is actually monitored

#### 3.3.2 PeerMonitor vs. GroupMonitor: What is the same?

In both the EXG PeerMonitor game and the EXG GroupMonitor game, each group member's actions are *actually* revealed with the same probability—in our specific setup, a 66.6% chance. The cost of monitoring is \$1 per subject in both games, which is achieved by subjects paying \$1 to make one other subject eligible to be monitored in the EXG PeerMonitor game, while groups pay \$4 to make four subjects eligible to be monitored in the EXG GroupMonitor game. In both exogenous games, monitoring payments are forced, so monitoring stage payoffs are simply the payoffs from the linear public goods game.

#### 3.3.3 EXG PeerMonitor: An Example

In Panel III of Figure 1, we show an example of what might happen during the EXG PeerMonitor game, which also highlights the differences in the information observed in the PeerMonitor versus GroupMonitor games. In the EXG PeerMonitor game, actions are only revealed to a single group member, which is similar to when a neighbor observes the bad acts of someone in their area. In Panel III of Figure 1, Person C's actions are only observed by Person B, and Person D's actions are only observed by Person C. In contrast in the EXG GroupMonitor game in Panel I of Figure 1, Person C's and Person D's actions are observed by all subjects.

### 3.3.4 EXG PeerMonitor: Punishments and Lack of Information

Under the PeerMonitor regime, a subject can punish the lowest monitored contributor without seeing the exact amount the lowest contributor put toward the public good. In our example in Panel III of Figure 1, Subject A can punish Subject C, even though A does not observe how much C gave to the public good. We believe this lack of information could make PeerMonitor less effective than GroupMonitor.[17] Given this setup, we may be biased toward finding that the GroupMonitor games are more successful than the PeerMonitor games. To preview our results, we do not find this to be the case.

## 3.4 The endogenous peer monitoring (ENDG PeerMonitor) game

The last of our main games is the ENDG PeerMonitor game.

### 3.4.1 ENDG PeerMonitor: An Example

The endogenous peer monitoring (ENDG PeerMonitor) game is identical to the EXG PeerMonitor game except that subjects choose whether to pay their $1 supplement toward making a peer eligible to be monitored. There are three cases of what can happen in this game. First, if all four subjects pay, then the game proceeds exactly as the EXG PeerMonitor game. Second, if none pays, it is a linear public goods game without information or punishments. Third, since the decision is made at the peer rather than group level, it is also possible for a subset of subjects to pay, and thus a subset of subjects are eligible for monitoring. In this third case, the lowest contributor from the subset that is actually monitored is then able to be punished. This third case is shown in Panel IV of Figure 1.

### 3.4.2 ENDG PeerMonitor vs. ENDG GroupMonitor

In both the ENDG PeerMonitor game and the ENDG GroupMonitor (Unanimous) game, each person is pivotal in the choice to make others eligible for monitoring. In the former, it is because each person is assigned to exactly one other group member; in the latter, it is because we require unanimity for the monitoring regime to be implemented. In both cases, it is impossible to free ride on making others eligible for monitoring, so differences in the

[17] In our data, subjects were about equally willing to punish someone whether they observed the exact contribution or not. In the EXG PeerMonitor game, 25% of subjects assign some punishment (12% who observe and 13% who do not observe the punished person's contribution). In the ENDG PeerMonitor game, 18.7% of subjects assign a punishment (8.4% who observe and 10.4% who do not observe the punished person's contribution).

efficacy of the regimes are not driven by miscoordination or free riding on the monitoring regime but rather by differences in who is willing-to-pay to monitor and how information is shared.

All of the main games are followed by a concentrated punishment mechanism; we will discuss other punishment mechanisms in section 6.

## 3.5 Procedures

We recruited students to participate at the laboratory at the University of Massachusetts Amherst using the Online Recruitment System for Experimental Economics software [Greiner, 2015]. Each session involved at minimum 12 subjects to preserve anonymity. Subjects are informed that they are in fixed groups of four with fixed identities for all 10 periods, so subjects were able to develop reputations. Each game had at least 32 subjects who participated. In this paper we report the results from 340 subjects in total. Each session was conducted using z-tree software [Fischbacher, 2007]. Subjects were paid for a randomly selected period and on average earned a payoff of $20.32 (inclusive of a $5 show-up payment).

Instructions are read aloud that describe the game as beginning with a contribution stage, then a monitoring stage, and ending with a punishment stage. In all games, when making the monitoring eligibility decision, subjects only know their own contribution to the public good. If a specific subject actually monitors someone else in a given period, then the specific subject observes the monitored subject's contribution and initial earnings before making a choice about assigning punishment points (when that option is available).

In the GroupMonitor games, all group members can easily see the total number of group members who have actually been monitored, whereas in the PeerMonitor games, subjects can only tell if they have been actually monitored by a peer. So the total number actually being monitored is unknown in PeerMonitor games. In all games, at the end of each period, subjects receive a report of their own contribution, initial payoff, costs of punishments they assigned to the lowest monitored contributor, and costs of punishments they received. Subjects are not told group contributions, although they could calculate this from the information given.

The instructions were written in neutral language by referring to the public good as the "BLUE investment," the private good as the "RED investment," and punishments as "deductions." Full instructions and screenshots are available from the authors in the online appendix.[18]

[18] Available at https://laurakgee.weebly.com/research.html.

# 4 Predictions

The own-profit-maximizing equilibrium prediction of all four games is the same. If the game is repeated a finite number of periods, in the final period no subject should pay to punish since punishments in the last period cannot influence future behavior. In the endogenous games, subjects will not pay to make anyone eligible for monitoring since no one will punish any other subject after. In the exogenous games, even though the monitoring payment is forced, the subjects still know that no punishments will take place. When the subjects know that no one will be punished, the own-profit-maximizing choice in the standard linear public goods game is to contribute zero to the public good.

If all subjects are own-profit maximizing, then each will earn \$10 in the linear public goods game such that final payments would be \$10 in exogenous games and \$11 in endogenous games.[19] However, the group-welfare-maximizing choice would be to contribute the full five-token endowment to the public good, not pay for detection and not punish, resulting in final payments of \$20 in exogenous games and \$21 in endogenous games.[20]

Although the own-profit-maximizing equilibrium predictions are the same across all four games, many previous works find that subjects often behave differently in games with the same equilibrium predictions, especially when those games include punishments [Chaudhuri, 2011]. With this in mind, we generate out-of-equilibrium predictions about the willingness-to-pay for monitoring for our two endogenous games.

## 4.1 Prediction for the ENDG PeerMonitor game

In the ENDG PeerMonitor game, we can find examples of when it would be own-profit maximizing for a subject to free ride and to pay to make someone eligible for monitoring. At first this may seem counterintuitive, as one might expect free riders not to value the provision of the public good enough to contribute toward monitoring its provision, so we will proceed with an illustrative example.

Consider yourself in the position of subject 4 in a group with three full free riders ($g_1 = g_2 = g_3 = 0$). You can ensure no punishments by being a full contributor ($g_i = 5$), since you

[19] dollars because each person keeps all five tokens, which are worth \$2 each, and earns nothing from the public good. In the exogenous game, this is their total earnings, whereas in the endogenous, they also elect to keep their \$1 stage 2 endowment.

[20] dollars because each person keeps all zero tokens, resulting in a group total of 20 tokens being in the public good. Each token in the public good is worth \$1, and each person earns nothing from the private good. In the exogenous game, this is their total earnings, whereas in the endogenous, they also elect to keep their \$1 stage 2 endowment.

cannot be punished when you have made a full contribution. Then you would earn \$6 if you kept the \$1 supplement.[21] But you may wonder, can I get away with partial free riding? Suppose you set $g_i = 3$ and do not pay to make anyone eligible to be monitored, resulting in a payoff of \$8.[22] Setting $g_i = 3$ instead of $g_i = 5$ results in a \$2 benefit.

Now that you are a partial free rider, should you pay to make someone eligible to be monitored? If you do not pay, you keep the \$1 supplement, but there is a greater chance you will be the lowest monitored contributor. Suppose you are the only one eligible to be monitored, then the expected punishment is the two-thirds chance that you will actually be monitored, multiplied by the \$3 penalty per punishment point, multiplied by the sum of punishments, $(\frac{2}{3})(3)\sum_j^N s_j$. The expected punishment is less than the \$2 benefit if $\sum_j^N s_j < 1$. It is only profitable to deviate from $g_i = 5$ to $g_i = 3$ if you believe less than one punishment point will be assigned.

However, you can decrease your chances of being the lowest monitored contributor by paying the \$1 supplement to make someone else eligible to be monitored, although this lowers the benefits to \$1. In your group everyone else free rode more than you, so if you pay for monitoring, there is a two-thirds chance you'll successfully catch someone who gave less than you. In other words, if you pay, there is a one-third chance you'll be unsuccessful in monitoring a bigger free rider. This makes your expected punishment the two-thirds chance that your actions will actually be revealed, multiplied by the one-third chance that you will be the lowest monitored contributor, multiplied by \$3, multiplied by the sum of punishments, $(\frac{2}{3})(\frac{1}{3})(3)\sum_j^N s_j$. The expected punishment is less than the \$1 benefit if $\sum_j^N s_j < 1.5$. By paying to monitor, we've altered the constraints on when it is profitable to partially free ride. Although this is a specific example, it illustrates the intuition that in the ENDG PeerMonitor game, it can be own-profit maximizing for a subject to partially free ride and pay to make others eligible for monitoring. When only the largest offenders are punished, it makes sense to seek out other offending peers. This leads us to our first prediction:

**Prediction 1.** *We may not observe a strong negative correlation between free riding and willingness-to-pay for monitoring in the ENDG PeerMonitor game if free riders believe there are larger free riders in their group who may be worth the expense of monitoring.*

[21] The sum of contributions to the public good is $0 + 0 + 0 + 5 = 5$, so payoff from the linear public goods game would be $0 \times 2 + 5 = 5$, and then you keep the \$1 supplement to earn \$6.

[22] The sum of contributions to the public good is $0 + 0 + 0 + 3 = 3$, so payoff from the linear public goods game would be $2 \times 2 + 3 = 7$, and the additional dollar supplement makes \$8.

## 4.2 Prediction for the ENDG GroupMonitor game

In the ENDG GroupMonitor (Unanimous) game, it is never own-profit maximizing for a subject to free ride and to pay to make the group eligible for monitoring. If a subject sets $g_i < 5$ and group monitoring is implemented, there is a chance the subject will incur the expected punishment. That punishment will be greater than zero if both punishments, $\sum_j^N s_j$, and the chance the subject is the lowest contributor are nonzero. Because group monitoring requires unanimity, a free riding subject can deterministically set the expected punishment to zero by not paying for monitoring. Even if we did not require unanimity, it would still be the case that free riding subjects would not pay for group monitoring because it would still raise the chance that they would be monitored/punished. When free riders can control whether they will be eligible to be monitored, they will not want to fund the group-wide auditor that may catch their bad acts. This leads us to our second prediction:

**Prediction 2.** *We will not observe free riders paying for monitoring in the ENDG GroupMonitor game because nonpayment reduces the likelihood of monitoring and punishment. In the special case requiring unanimous support for implementation, free riders can single-handedly reduce the likelihood of monitoring their own bad acts to zero.*

We concentrate on predictions about what happens when monitoring is endogenous because, although many examples of group and peer monitoring exist in the real world, monitoring is at some point in time a costly endogenous choice.[23] Consider the costs of reporting inappropriate content on social media or the costs of a task force monitoring illegal dumping. Monitoring, whether at the peer or group level, is not always automatic, yet this has been the starting point for the majority of the research on public good provision.

# 5 Results

The EXG GroupMonitor game is our baseline since it is closest to the previously studied games. We first test whether endogenizing group monitoring in the ENDG GroupMonitor (Unanimous) game changes outcomes. If endogenizing monitoring has no effect on outcomes,

[23] A last thing to note is that our two predictions are about how willingness-to-pay might vary by whether a subject is in the ENDG GroupMonitor instead of ENDG PeerMonitor game. However, were we to assume a simple decision-making model where each subject has an independent probability of paying for monitoring that does not vary by peer versus group monitoring, we would mechanically find fewer people monitored in the ENDG GroupMonitor game than the ENDG PeerMonitor game. Our main outcome variable is public contribution rather than number monitored, so we direct the reader to footnote 31 for details.

then it is benign to assume groups are always willing and able to centrally monitor one another. However, we find that endogenizing group monitoring decreases the number of bad acts detected and lowers public good provision. Knowing this, we then test whether the alternative of peer monitoring suffers from these same decreases when we switch from exogenous to endogenous monitoring. Finally, we delve into why peer monitoring survives endogenization while group monitoring does not.

## 5.1 Endogenizing group monitoring: EXG GroupMonitor versus ENDG GroupMonitor (Unanimous)

We first test whether endogenizing group monitoring changes outcomes. Rows three and four of Table 1 show that endogenizing group monitoring lowers public good contributions by 44% as tokens contributed fall from 2.94 in the EXG GroupMonitor game to 1.64 in the ENDG GroupMonitor (Unanimous) game.[24] This fall is immediate and persistent, as seen by comparing the black lines in Figure 3. This is coupled with almost no one being monitored or punished in the ENDG GroupMonitor (Unanimous) game.

Table 2 shows the results from a difference-in-differences model that includes two indicator variables for whether monitoring is done at the group level and whether monitoring is endogenous, as well as the interaction of these two indicators. The omitted category is the EXG PeerMonitor game. This difference-in-differences model allows us to look at the individual effects of endogenizing monitoring and of moving monitoring to the group level, in addition to the combined effect of both those things together.[25]

[24]This change is statistically significant, as can be seen from the coefficient on Endog*GroupM (Unan.) in the regression results presented in Table 2. The same significance results can be shown using a Wilcoxon rank-sum test with one observation per group for public contributions ($Prob > |z| = 0.002$), percent gave endowment ($Prob > |z| = 0.22$), number monitored ($Prob > |z| = 0.00$), WTP for monitoring ($Prob > |z| = 0.00$), total punishments ($Prob > |z| = 0.00$), and net earnings ($Prob > |z| = 0.63$). There are 14 groups that play the game 10 times in ENDG GroupMonitor (Unanimous), and the GroupMonitor is only implemented once of the possible 140 times. One may worry that requiring unanimity guarantees that group monitoring will not be implemented. We show these results are not specific to the unanimity requirement in section 6.1, and they are summarized in the Extensions panel of Table 1.

Since almost no one is monitored or punished in our ENDG GroupMonitor game, one could argue that this is similar to a game where subjects have been exogenously assigned to see no information about others' actions and have no punishment opportunities. However, we believe it is fundamentally different when a group has been offered the opportunity to monitor but fails to implement it versus when that opportunity is exogenously not allowed (and perhaps not even mentioned to subjects).

[25]The specific estimating equation used in Table 2 is

$$y_{it} = \beta_1 * GroupM(Unan.)_i + \beta_2 * EndogM_i + \beta_3 * EndogM_i * GroupM(Unan.)_i + c + \epsilon_{it},$$

where $y_{it}$ is an outcome for subject $i$ in period $t$. The variable $GroupM(Unan.)_i$ takes the value 1 if person $i$

Table 1: **Summary Statistics by Game**

| | Game | Public Cont. | % Gave All | % Gave 0 | Number Monitored^ | WTP to Monitor | Total Punishments | Net Earnings | Obs. |
|---|---|---|---|---|---|---|---|---|---|
| Main Games | EXG PeerM - ConcenP | 2.80 | 0.30 | 0.19 | 2.50 | 1.00 | 1.63 | 13.97 | 360 |
| | ENDG PeerM - ConcenP | 2.71 | 0.25 | 0.19 | 0.86 | 0.25 | 1.32 | 14.85 | 320 |
| | EXG GroupM - ConcenP | 2.94 | 0.18 | 0.09 | 2.38 | 1.00 | 2.09 | 13.78 | 400 |
| | ENDG GroupM (Unan.) -ConcenP | 1.64 | 0.10 | 0.43 | 0.03 | 0.32 | 0.03 | 14.25 | 560 |
| | ENDG GroupM (Non-Unan.) -ConcenP | 2.15 | 0.14 | 0.27 | 0.37 | 0.37 | 0.35 | 14.81 | 480 |
| Extensions | EXG PeerM - NeighborP | 2.48 | 0.23 | 0.26 | 2.50 | 1.00 | 1.61 | 13.34 | 320 |
| | ENDG PeerM - NeighborP | 2.66 | 0.29 | 0.25 | 0.98 | 0.28 | 1.12 | 14.91 | 360 |
| | EXG GroupM - NeighborP | 2.49 | 0.15 | 0.19 | 2.50 | 1.00 | 1.05 | 13.93 | 320 |
| | ENDG GroupM (Unan.) - NeighborP | 1.71 | 0.15 | 0.46 | 0.00 | 0.34 | 0.00 | 14.41 | 280 |

*Notes:* ^ excludes those detected because they gave their full endowment. In all models, standard errors are clustered at the group level. There were a total of 340 subjects in 85 fixed groups of four who each played for 10 periods. Contributions could be between zero and five tokens. Number Monitored could be between zero and four subjects per group. WTP to Monitor is the proportion of subjects (0%–100%) who were willing-to-pay to monitor. 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 PeerMonitor games, WTP to Monitor is the proportion of subjects who paid $1 to monitor their right-side group member. In GroupMonitor (Unanimous) games, WTP to Monitor is the proportion of subjects who pledged $1 toward a group monitoring regime, but if the group did not collect $4, this amount was refunded back. Total Punishments is the cost of assigned punishments added to received punishments. Net Earnings represents the total earnings minus the costs of punishments both assigned and received.

Table 2 shows that there are clear differences in many of the outcomes of interest, in particular the amount of the public good provided. This should give us pause since the lack of a group monitoring leads to an underprovision of the public good. Interestingly, average net earnings are not harmed by this lack of monitoring. Net earnings are slightly higher in the ENDG GroupMonitor (Unanimous) game than in the EXG GroupMonitor one, but this difference is neither statistically nor economically significant. Groups that fail to implement group monitoring provide less of the public good but also spend less on punishments so that on net they have similar earnings.

**Result 1.** *When we no longer assume that everyone automatically opts to implement group monitoring, this affects many outcome variables. In particular, we see a 44% drop in provision of the public good in the ENDG GroupMonitor (Unanimous) game.*

## 5.2 Endogenizing peer monitoring: EXG PeerMonitor versus ENDG PeerMonitor

Endogenizing our group monitoring regime leads to very different outcomes than exogenously imposing it. We next compare whether endogenizing peer monitoring leads to differences. Looking at the top two rows of Table 1, we see that the average public good provision is not statistically significantly different between the EXG PeerMonitor game (2.80 tokens) and the ENDG PeerMonitor one (2.71 tokens).[26] This is in spite of a large drop in the

is in a game with Group Monitoring and 0zero otherwise. The variable $EndogM_i$ takes the value 1 if person $i$ is in a game with Endogenous Monitoring and zero otherwise.

[26]This can be seen from the coefficient on EndogM in Table 2. The same significance of the results from Table 2 can be shown using a Wilcoxon rank-sum test with one observation per group for public contributions

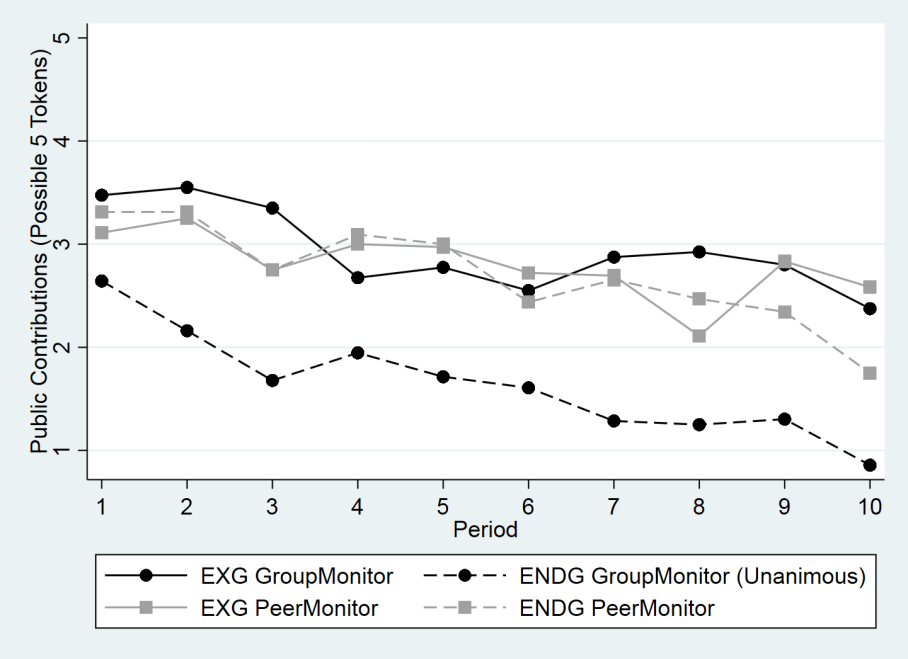



Figure 3: Public Contributions

average number monitored, from 2.50 to 0.86 per four-person group and a less dramatic decline in average total punishments from \$1.63 to \$1.32 per person. The combination of slightly lower public good contributions, lower punishments, and cost savings from less monitoring leads to average net earnings that are similar or even slightly higher when peer monitoring is endogenous instead of exogenous, although the difference is neither statistically nor economically significant.

Peer monitoring is quite robust to endogenization, which may not be so surprising as many real-world examples of peer monitoring exist. Consider calling out negative transactions on eBay or reporting your neighbor who has illegally dumped their waste. Although these bad acts could be detected by a centralized all-seeing panopticon, they tend to be detected by peers.

Additionally, the ENDG PeerMonitor game results in similar levels of public good provision to EXG GroupMonitor, the game most similar to the previous literature. This is particularly clear when looking at Figure 3, where public contributions for all but the ENDG GroupMonitor (Unanimous) game overlap. We find this comforting because it means there

($Prob > |z| = 0.80$), percent gave endowment ($Prob > |z| = 0.62$), number monitored ($Prob > |z| = 0.00$), WTP for monitoring ($Prob > |z| = 0.00$), total punishments ($Prob > |z| = 0.41$), and net earnings ($Prob > |z| = 0.11$).

Table 2: **Difference-In-Difference Regressions**

| Variable | 1<br>Public<br>Cont. | 2<br>% Gave<br>All | 3<br>% Gave<br>0 | 4<br>Number<br>Monitored^ | 5<br>WTP<br>to Monitor | 6<br>Total<br>Punishments | 7<br>Net<br>Earnings |
|---|---|---|---|---|---|---|---|
| GroupM (Unan.) | 0.132 | -0.118 | -0.099+ | -0.120 | 0.000*** | 0.457 | -0.192 |
| | (0.439) | (0.109) | (0.053) | (0.088) | (0.000) | (0.726) | (0.989) |
| EndogM | -0.090 | -0.056 | -0.004 | -1.637*** | -0.750*** | -0.308 | 0.878 |
| | (0.454) | (0.122) | (0.069) | (0.116) | (0.036) | (0.348) | (1.007) |
| Endog*GroupM (Unan.) | -1.200* | -0.027 | 0.342** | -0.714*** | 0.068 | -1.753* | -0.404 |
| | (0.576) | (0.138) | (0.097) | (0.148) | (0.053) | (0.761) | (1.260) |
| Constant | 2.803*** | 0.303** | 0.192*** | 2.500*** | 1.000*** | 1.633*** | 13.972*** |
| | (0.344) | (0.093) | (0.043) | (0.000) | (0.000) | (0.265) | (0.784) |
| Adj. R2 | 0.091 | 0.037 | 0.098 | 0.680 | 0.529 | 0.045 | 0.004 |
| Obs. | 1640 | 1640 | 1640 | 1640 | 1640 | 1640 | 1640 |
| EndogM | -1.290*** | -0.083 | 0.338*** | -2.351*** | -0.682*** | -2.061** | 0.474 |
| +EndogGroupM (Unan.) | (0.356) | (0.063) | (0.068) | (0.093) | (0.039) | (0.678) | (0.759) |

*Notes:* ^ excludes those detected because they gave their full endowment. + significant at 10%; * significant at 5%; ** significant at 1%, *** significant at .1%. In all models, standard errors are clustered at the group level. There were a total of 164 subjects in 41 fixed groups of four, who each played for 10 periods. Contributions could be between zero and five tokens. Number Monitored could be between zero and four subjects per group. WTP to Monitor is the proportion of subjects (0%–100%) who were willing-to-pay to monitor. 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 PeerMonitor games, WTP to Monitor is the proportion of subjects who paid $1 to monitor their right-side group member. In GroupMonitor (Unanimous) games, WTP to Monitor is the proportion of subjects who pledged $1 toward a group monitoring regime, but if the group did not collect $4, this amount was refunded back. Total Punishments is the cost of assigned punishments added to received punishments. Net Earnings represents the total earnings minus the costs of punishments both assigned and received.

are endogenous ways to achieve similar results to the more standard EXG GroupMonitor setting. We need not be worried that previous results assuming exogenous peer monitoring are not robust to alternative monitoring environments. The results from this experiment imply that similar levels of public good provision can be obtained through endogenous monitoring if that monitoring is done by peers.

**Result 2.** *Peer monitoring results in almost identical levels of public good provision regardless of whether we assume that everyone will automatically monitor or allow endogenous choices. Furthermore, public good provision under endogenous peer monitoring is similar to that found under the more traditional exogenous group monitoring setup.*

## 5.3 Why does peer monitoring survive being endogenous?

In this section, we delve into why peer monitoring is robust to endogenization.

### 5.3.1 Who is willing-to-pay for monitoring?

The Main Results panel of Table 1 shows willingness-to-pay to monitor is similar across the ENDG GroupMonitor (Unaninmous) game, 32%, and the ENDG PeerMonitor one, 25%.[27]

[27]The difference is not statistically significant, as can be seen in the coefficient on $EndogGroupMonitor$ (Unanimous) in Table 2 or by using a Wilcoxon rank-sum test with one observation per group $Prob > |z| = 0.33$.

This masks differences in who is willing-to-pay when monitoring is done at the group versus the peer level. Intuitively, free riders seem unlikely to pay for either group or peer monitoring. However, as discussed in Section 4, we predict that free riders will never pay for monitoring in the ENDG GroupMonitor (Unanimous) game, but they might in the ENDG PeerMonitor game.

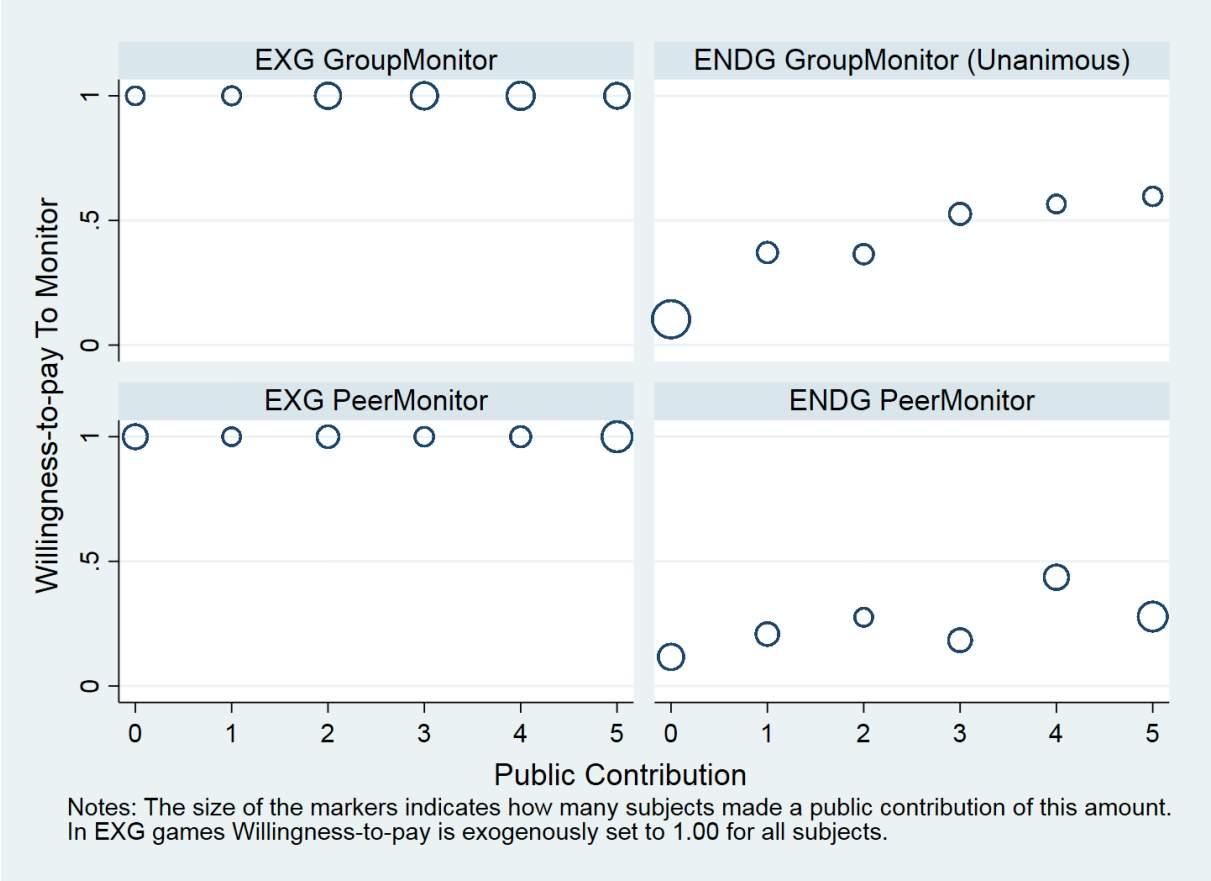



Figure 4: Correlation between willingness-to-pay to monitor and public contribution

Figure 4 shows (i) the proportion who contributed each amount and (ii) the correlation between willingness-to-pay for monitoring and amount contributed.[28] Larger circles indicate that a greater proportion of subjects have made a contribution of a specific amount. The vertical placement of a circle indicates the proportion of subjects who first made a specific

[28]Figure 4 shows the relationship between *absolute* free riding and willingness-to-pay for monitoring. In Appendix Figure C1, we show the patterns are very similar if one looks at *relative* rather than *absolute* free riding. And in Appendix Figure C2, we show the patterns if one looks at *group* rather than *individual* behavior.

contribution (e.g., 0, 1, 2 tokens) and then pledged money toward monitoring. Recall that in the exogenous games, subjects are forced to pay for monitoring, so the vertical placement of all circles is at 1 in the left-hand column of Figure 4.

In the top row of Figure 4, there is a large shift in the proportion of subjects giving nothing to the public good when the GroupMonitor game is made endogenous. In fact, the largest circle in Figure 4 is where 43% of the subjects give zero in the ENDG GroupMonitor (Unanimous) graph, yet looking at the analogous place on the EXG GroupMonitor graph, we see one of the smallest circles with only 9% giving zero. Endogenizing group monitoring causes a 34 percentage point rise in the proportion of subjects giving nothing. This is likely because free riders can guarantee there will be no monitoring in this game, so they choose to contribute zero, knowing they will not be monitored or punished in line with Prediction 2.[29] Moreover, the low horizontal placement of the large circle on the ENDG GroupMonitor (Unanimous) graph shows that only 10% of zero-contributing subjects pay to monitor.

In line with Prediction 2, zero-contributing subjects almost never pay for monitoring. However, we predicted that free riders of any level would never pay for monitoring, such that all the circles in the ENDG GroupMonitor setting would be at the bottom of the graph until it reaches full contribution. Instead we see an upward-sloping pattern. Willingness-to-pay for monitoring is positively correlated with greater public contributions in the same period, as indicated by the results in column 1 of Table 3.

We see the second- and third-largest circles across the four panels of Figure 4 are both in the PeerMonitor graphs. In the EXG PeerMonitor game, 30% of subjects gave all five tokens, and this only falls a small amount, to 25% of subjects in the ENDG PeerMonitor game. Similarly, the proportion giving zero is 19% in both the EXG PeerMonitor and ENDG PeerMonitor games (as indicated by the same size circles).[30] Endogenization has virtually no effect on the number of full contributors and full free riders under the PeerMonitor game. In line with Prediction 1, there is no strong correlation between public contribution and willingness-to-pay for monitoring in the ENDG PeerMonitor game (column 2 of Table 3). Because there is little change in the distribution of public contributions and no strong correlation between the level of contribution and willingness-to-pay for monitoring, we see similar levels of public good provision whether peer monitoring is exogenous or endogenous.

[29] This difference is statistically significant, as shown in column 3 of Table 2. However this rise in zero contributions is not specific to the unanimity rule. We find that when we only require three people to pledge for the GroupMonitor, there is still a statistically significant 18 percentage point rise in the proportion of subjects giving zero.

[30] These differences are not statistically significant, as shown in the coefficient on *EndogM* in columns 2 and 3 of Table 2.

Table 3: **Correlation between willingness-to-pay and public contribution choice**

| | WTP to Monitor | |
|---|---|---|
| | **1** | **2** |
| | **ENDG** | **ENDG** |
| | **GroupM (Unan.)** | **PeerM** |
| Public Cont. | 0.107*** | 0.038 |
| | (0.017) | (0.020) |
| Constant | 0.141** | 0.147** |
| | (0.043) | (0.041) |
| Adj. R2 | 0.166 | 0.024 |
| Obs. | 560 | 320 |

*Notes:* + significant at 10%; * significant at 5%; ** significant at 1%, *** significant at .1%. In all models, wild bootstrapped standard errors are clustered at the group level. There were a total of 88 subjects in 22 fixed groups of four who each played for 10 periods with 14 groups in column one and eight groups in column two. Contributions could be between zero and five tokens. In PeerMonitor games, WTP to Monitor is the proportion of subjects who paid $1 to monitor their right-side group member. In GroupMonitor (Unanimous) games, WTP to Monitor is the proportion of subjects who pledged $1 toward a group monitoring regime, but if the group did not collect $4, this amount was refunded back. The coefficients on *PublicContribution* in these two models are statistically significantly different $Prob > chi2 = 0.0060$.

**Result 3.** *When group monitoring is made endogenous, more people contribute nothing, and there is a negative correlation between free riding and willingness-to-pay for monitoring that might detect low contributions. In contrast, when peer monitoring is made endogenous, contributions are unchanged and there is no strong correlation between contributions and willingness-to-pay, which is partially why peer monitoring is robust to endogenization.*

### 5.3.2 When does the number monitored matter?

Endogenizing group monitoring leads to more full free riding and an unwillingness-to-pay for monitoring by those free riders. This leads to almost no one being monitored or punished in the ENDG GroupMonitor (Unanimous) game. However, looking at Tables 1 and 2, we also see a statistically significant drop in the number monitored, from 2.5 subjects in the EXG PeerMonitor game to only 0.86 subjects in the ENDG PeerMonitor one. Yet public contributions are almost the same across these two games (2.80 vs. 2.71 tokens). Even with this 65% drop in the number monitored, endogenizing monitoring has no measurable effect on contributions in the PeerMonitor games.[31]

[31] It is also interesting to note that when we endogenize either group or peer monitoring, we observe a drop in monitoring that is much larger than one might expect using a very simple model of decision-making. Let us assume a simple decision-making model where each subject has an independent 50% chance of being willing-to-pay for monitoring (regardless of public contribution level). Then in ENDG GroupMonitor (Unanimous), we would expect 0.16 persons to be monitored each period ($\frac{2}{3} * 4 * (0.50)^4 = 0.16$), but under ENDG

These findings imply that we simply may not need as much monitoring to maintain high contributions. In that case it would not matter if people were monitored by chance or rather by choice. To test this, we use the fact that in our exogenous games, there is a two-thirds chance that each subject is monitored, which creates exogenous variation in the number monitored, even when monitoring is mandated. By contrast, in our endogenous games, the number monitored is a mixture of endogenous choices to pay for monitoring, and there is a two-thirds chance that choice is realized.

In Table 4, we regress public contributions in period $t$ on the realized number monitored in period $t-1$, an indicator for whether the game is an endogenous game, and the interaction of these variables. This allows us to explore the correlation between realized number monitored in the endogenous and exogenous games.[32] The coefficient on $\text{NumMonitored}_{t-1}$ is positive but is statistically and economically insignificant.[33] In contrast, an additional person monitored in the endogenous games is associated with a 0.609 rise in tokens contributed the next period (see bottom row of Table 4), which is both statistically and economically significant.[34]

Increasing the number monitored does not raise contributions to the public good when the number monitored is purely exogenously determined. In our ENDG GroupMonitor (Unanimous) game, almost no one is endogenously monitored, which results in a large decline in public contributions. However, in our ENDG PeerMonitor games, the very small number of subjects monitored, 0.86, results in similar levels of public contribution as the EXG PeerMonitor game.

**Result 4.** *The number monitored only matters when monitoring is not purely exogenous. A small amount of monitoring can be associated with relatively high public contributions when*

PeerMonitor we expect 1.33 persons to be monitored ($\frac{2}{3} * 4 * (0.50) = 1.33$). However, in our data we observe only 0.03 persons monitored under the ENDG GroupMonitor (Unanimous) game and 0.86 persons monitored under the ENDG PeerMonitor one. Given these calculations, the difference between the expected versus the observed drop in monitoring is larger for GroupMonitor ($1 - 0.03/0.16 = 81\%$) than for PeerMonitor ($1 - 0.86/1.33 = 35\%$).

[32] A helpful reviewer pointed out that we should be able to break this into an endogenous and exogenous component when there is a "surprise" of nonrealized monitoring. However, such analysis is not informative because there are very few "surprises." In our EXG GroupMonitor (Unanimous) game, only one group pays for monitoring, which is implemented without a "surprise." In our ENDG PeerMonitor game, there are only 24 "surprises."

[33] In the exogenous games, the addition of one more monitored person results in a 0.011 token increase, representing less than a 1% increase from the baseline of 2.798 tokens when no one is monitored in an exogenous game.

[34] The 0.609 increase is a 35% rise from the baseline of 1.72 tokens when no one is monitored in an endogenous game.

Table 4: **Exogenous vs. endogenous number monitored in $t-1$ and public contributions in $t$**

| | **Public Contributions$_t$** |
|---|---|
| NumMonitored$_{t-1}$ | 0.011 |
| | (0.107) |
| EndogM*NumMonitored$_{t-1}$ | 0.598** |
| | (0.212) |
| EndogM | -1.078* |
| | (0.445) |
| Constant | 2.798*** |
| | (0.388) |
| Adj. R2 | 0.083 |
| Obs. | 1476 |
| NumMonitored$_{t-1}$ + | 0.609** |
| ENDG*NumMonitored$_{t-1}$ | (0.183) |

*Notes:* + significant at 10%; * significant at 5%; ** significant at 1%, *** significant at .1%. This table presents results from the EXG GroupMonitor, ENDG GroupMonitor (Unanimous), EXG PeerMonitor, and the ENDG PeerMonitor games. Standard errors are clustered at the group level. There were a total of 164 subjects in 41 fixed groups of four, who each played for 10 periods. Contributions could be between zero and five tokens.

*that monitoring is partially endogenously selected, as is the case in our ENDG PeerMonitor game.*

A final thing to note is that when we endogenize peer monitoring, those who actually observe an action have a 65% higher likelihood of punishing versus subjects in EXG PeerMonitor game.[35] So even though we catch fewer bad acts, the bad acts are punished more often when peer monitoring is an endogenous choice.

# 6 Extensions

In this section, we show our results are not specific to requiring unanimity for group monitoring nor to concentrating punishments on the lowest contributor. The Extensions panel of Table 1 shows summary statistics for the extensions discussed in this section.

[35]In the EXG PeerMonitor game, 2.5 of the four people's actions are observed on average, while in the ENDG PeerMonitor game, only 0.86 of the four peoples' actions are observed on average. But in the EXG PeerMonitor game, only 29% assign a punishment, whereas in the ENDG PeerMonitor game, that rises to 48% ($\frac{0.48}{0.29} = 1.65$).

## 6.1 Other voting rules

Endogenizing monitoring leads to a decline in public contributions under group monitoring when we require unanimity. We now present evidence from the ENDG GroupMonitor (Non-unanimous) game, which is identical to the ENDG GroupMonitor (Unanimous) game except that we only require three of the four subjects to be willing-to-pay for monitoring to be implemented. Importantly, all subjects still have to pay for monitoring to keep the costs of monitoring the same across games.

In the second row of Table 5, we show the effects of endogenizing group monitoring when requiring only three out of four subjects rather than unanimity, while the first row shows the results for unanimity. Comparing row 1 and row 2 of Table 5, we see most coefficients maintain their same sign and are still significantly different from the baseline EXG GroupMonitor game, although the magnitudes and level of significance are sometimes lower for the nonunanimous game relative to the unanimous game.

Table 5: **Unanimous vs. Non-unanimous in GroupMonitor games**

| Variable | 1<br>Public<br>Cont. | 2<br>% Gave<br>All | 3<br>% Gave<br>0 | 4<br>Number<br>Monitored^ | 5<br>WTP<br>to Monitor | 6<br>Total<br>Punishments | 7<br>Net<br>Earnings |
|---|---|---|---|---|---|---|---|
| ENDG GroupM (Unan) | -1.290*** | -0.083 | 0.338*** | -2.351*** | -0.682*** | -2.061** | 0.474 |
| | (0.356) | (0.063) | (0.068) | (0.093) | (0.039) | (0.678) | (0.759) |
| ENDG GroupM (Non-Unan) | -0.781* | -0.047 | 0.178*** | -2.013*** | -0.633*** | -1.740* | 1.028 |
| | (0.338) | (0.068) | (0.044) | (0.135) | (0.038) | (0.685) | (0.712) |
| Constant (Omitted EXG GroupM) | 2.935*** | 0.185** | 0.093** | 2.380*** | 1.000*** | 2.090** | 13.780*** |
| | (0.274) | (0.057) | (0.031) | (0.088) | (0.000) | (0.677) | (0.604) |
| Adj. R2 | 0.083 | 0.008 | 0.090 | 0.597 | 0.351 | 0.057 | 0.006 |
| Obs. | 1440 | 1440 | 1440 | 1440 | 1440 | 1440 | 1440 |

*Notes:* ^ excludes those detected because they gave their full endowment. + significant at 10%; * significant at 5%; ** significant at 1%, *** significant at .1%. In all models, standard errors are clustered at the group level. There were a total of 144 subjects in 36 fixed groups of four, who each played for 10 periods. Contributions could be between zero and five tokens. Number Monitored could be between zero and four subjects per group. WTP to Monitor is the proportion of subjects (0%–100%) who were willing-to-pay to monitor. 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 GroupMonitor games, WTP to Monitor is the proportion of subjects who pledged $1 toward a group monitoring regime, but if the group did not collect four in the unanimous games or three in the nonunanimous games, this amount was refunded back. Total Punishments is the cost of assigned punishments added to received punishments. Net Earnings represents the total earnings minus the costs of punishments both assigned and received.

Without unanimity, we find a negative and statistically significant drop in public contributions, number monitored, willingness-to-pay to monitor, and total punishments, as we did in Result 1. Similar to Result 3, we find that when group monitoring is made endogenous without unanimity, there is still an increase in the proportion of subjects giving nothing. We also see that the positive and statistically significant relationship between the amount contributed and willingness-to-pay to monitor is almost exactly the same in the nonunanimous and unanimous game (0.109 versus 0.107, as shown in Appendix Table A1 and Figure A2). Additionally, just as in our Result 4, decreases in the number monitored in the nonunanimous game are correlated with a fall in public contributions, just as they were in the unanimous

game (0.332 versus 0.598, as shown in Appendix Table A2).

**Result 5.** *The efficacy of group monitoring is harmed by endogenization whether the endogenous choice requires unanimous support or something lower than unanimity. However, the magnitude of the harms decrease when we no longer require unanimity.*

## 6.2 Other punishment mechanisms

Thus far, everything has a punishment mechanism that concentrates all punishments on the lowest contributor to the public good. This type of concentrated punishment captures the idea that often those most out of compliance are the only ones punished. However, another possible punishment mechanism is one where those closest to the violator are responsible for carrying out punishments.[36] This type of punishment is related to commonly used peer-to-peer punishments. In peer-to-peer punishments, any subject can punish any other subject, so punishments still take place at the peer level but are not limited to a single peer.

Although we are interested in the efficacy of various monitoring regimes, the monitoring regime may interact with the punishment mechanism. We ran a set of games identical to those already discussed but with the difference that a subject could only be punished by a single neighbor. The own-payoff-maximizing equilbria of these games were the same as those with concentrated punishments that we have already discussed.

The bottom four rows of Table 1 show summary statistics for games with this alternative neighbor punishment mechanism and show that outcomes are quite similar to the results of main games with a concentrated punishment mechanism. To formalize this, we present the results from a triple difference regression in Table 6. This model includes the following three indicator variables and their interactions: (1) whether monitoring is done at the group level, (2) whether monitoring is endogenous, and (3) whether a subject can only be punished by a single neighbor.[37] Table 6 shows that the use of a NeighborPunishment mechanism instead

[36]For example, if a homeowner illegally dumps their waste, it is likely their neighbor (whose health and well-being are immediately affected) will mete out punishments.

[37]The omitted category in Table 6 is the EXG PeerMonitor game with concentrated punishments. This is the same omitted category in the model in Table 2. The specific estimating equation used in Table 6 is

$y_{it} = \beta_1 * GroupM(Unan.)_i + \beta_2 * EndogM_i + \beta_3 * EndogM_i * GroupM(Unan.)_i + \beta_4 * NeighborP_i + \beta_5 * NeighborP_i * EndogM_i + \beta_6 * NeighborP_i * GroupM(Unan.)_i + \beta_7 * NeighborP_i * EndogM_i * GroupM(Unan.)_i + c + \epsilon_{it}$,

where $y_{it}$ is an outcome for subject $i$ in period $t$. The variable $GroupM(Unan.)_i$ takes the value 1 if person $i$ is in a game with Group Monitoring and 0 otherwise. The variable $EndogM_i$ takes the value 1 if person $i$ is in a game with Endogenous Monitoring and 0 otherwise. The variable $NeighborP_i$ takes the value 1 if person $i$ is in a game where a person is only punished by a single neighbor and 0 otherwise.

of the ConcentratedPunishment mechanism did not have any statistically significant effect on the outcome variables.[38]

Table 6: **Other Punishment Mechanism Using Triple Difference Regressions**

| Variable | 1<br>Public<br>Cont. | 2<br>% Gave<br>All | 3<br>% Gave<br>0 | 4<br>Number<br>Monitoredˆ | 5<br>WTP<br>to Monitor | 6<br>Total<br>Punishments | 7<br>Net<br>Earnings |
|---|---|---|---|---|---|---|---|
| GroupM(Unan.) | 0.132 | -0.118 | -0.099+ | -0.120 | 0.000 | 0.457 | -0.192 |
| | (0.437) | (0.109) | (0.052) | (0.088) | (0.000) | (0.723) | (0.984) |
| EndogM | -0.090 | -0.056 | -0.004 | -1.638*** | -0.750*** | -0.308 | 0.878 |
| | (0.451) | (0.122) | (0.069) | (0.115) | (0.036) | (0.346) | (1.002) |
| GroupM(Unan.)*EndogM | -1.200* | -0.027 | 0.342*** | -0.714*** | 0.068 | -1.753* | -0.404 |
| | (0.573) | (0.137) | (0.096) | (0.148) | (0.052) | (0.757) | (1.254) |
| NeighborP | -0.328 | -0.068 | 0.071 | 0.000 | 0.000 | -0.021 | -0.635 |
| | (0.410) | (0.106) | (0.075) | (0.064) | (0.000) | (0.351) | (0.914) |
| NeighborP*EndogM | 0.274 | 0.110 | -0.008 | 0.115 | 0.031 | -0.182 | 0.699 |
| | (0.610) | (0.160) | (0.105) | (0.278) | (0.079) | (0.530) | (1.389) |
| NeighborP*GroupM(Unan.) | -0.120 | 0.037 | 0.030 | 0.120 | -0.000 | -1.019 | 0.780 |
| | (0.520) | (0.127) | (0.092) | (0.088) | (0.000) | (0.847) | (1.273) |
| NeighborP*GroupM(Unan.)*EndogM | 0.236 | -0.030 | -0.063 | -0.264 | -0.013 | 1.193 | -0.683 |
| | (0.779) | (0.182) | (0.143) | (0.293) | (0.104) | (0.936) | (1.803) |
| Constant | 2.803*** | 0.303** | 0.192*** | 2.500*** | 1.000*** | 1.633*** | 13.972*** |
| | (0.342) | (0.093) | (0.042) | (0.044) | (0.000) | (0.263) | (0.780) |
| Adj. R2 | 0.064 | 0.029 | 0.074 | 0.620 | 0.531 | 0.041 | 0.008 |
| Obs. | 2920 | 2920 | 2920 | 2920 | 2920 | 2920 | 2920 |

*Notes:* ˆ excludes those detected because they gave their full endowment. + significant at 10%; * significant at 5%; ** significant at 1%, *** significant at .1%. In all models, standard errors are clustered at the group level. There were a total of 292 subjects in 73 fixed groups of four, who each played for 10 periods. Contributions could be between zero and five tokens. Number Monitored could be between zero and four subjects per group. WTP to Monitor is the proportion of subjects (0%–100%) who were willing-to-pay to monitor. 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 PeerMonitor games, WTP to Monitor is the proportion of subjects who paid $1 to monitor their right-side group member. In GroupMonitor (Unanimous) games, WTP to Monitor is the proportion of subjects who pledged $1 toward a group monitoring regime, but if the group did not collect $4, this amount was refunded back. Total Punishments is the cost of assigned punishments added to received punishments. Net Earnings represents the total earnings minus the costs of punishments both assigned and received.

Whether we use neighbor or concentrated punishments, there is only a positive and significant relationship between contributions and willingness-to-pay for monitoring under group rather than peer monitoring, similar to our Result 3 (see Appendix Table B1 and Figure B2). Like our Result 4, under the neighbor punishment mechanism, there is only a positive correlation between the number monitored and subsequent public contributions when the number monitored is partially endogenous (Appendix Table B2).

**Result 6.** *The efficacy of a monitoring regime is unaffected by the type of punishments it is paired with. Outcomes by monitoring regimes (EXG GroupMonitor, ENDG GroupMonitor (Unanimous), EXG PeerMonitor, ENDG PeerMonitor) do not differ by whether monitoring is followed by punishments that are concentrated on the largest free rider or punishments that are meted out by a single neighbor.*

[38] Wilcoxon rank-sum tests with one observation per group for pairwise comparisons by punishment mechanism are never statistically significant at traditional levels. Additionally, the patterns in Figure 4 are quite similar regardless of punishment mechanism, as can be seen in Appendix Figure B1.

# 7 Summary & concluding remarks

Monitoring is a crucial step in punishing and hopefully deterring bad acts. At some point a society must endogenously choose how to monitor bad acts. Yet, research often assumes the actions of the whole collective are monitored and broadcast to all members. In this paper, we show that when monitoring is endogenous, a lack of agreement on implementing group monitoring leads to a 44% decline in public good provision. However, this fall is avoided when a society uses peer monitoring.

When monitoring is not automatic, group-level monitoring fails because there is a correlation between free riding and being unwilling-to-pay for monitoring. As such, the groups that most need monitoring fail to provide it. Peer-level monitoring survives endogenization perhaps because under peer monitoring, free riders willingly pay to detect others' wrongdoing to take attention away from their own bad acts.

To our knowledge, this is the first paper to contrast endogenous and exogenous monitoring in a public goods game while exploring the interaction of endogenization with peer versus group monitoring. A limitation of the current study is that we do not allow a mixture of peer and group monitoring. Yet, crimes are detected by both the police and peers calling the tip line; piracy is deterred by both navies and private security aboard vessels; and sports are regulated by referees as well as informal "enforcers" [DeAngelo and Smith, 2020, DeAngelo et al., 2017]. Additionally, our study hints at the importance of the interaction of the amount of consensus required to monitor with the number of people monitored. What would happen if we allowed a single person to implement group-level monitoring, or if we required group consensus to monitor even a single person? Although these are clearly important and desirable next steps in this research agenda, they are beyond the scope of this study.

From our current study, we offer two implications. The first is specific to how we interpret the previous work that assumes exogenous group monitoring, and the second speaks more broadly to how a society might choose between different monitoring regimes.

First, our results imply that findings from previous work using exogenous group monitoring might be replicated using an endogenous monitoring regime when monitoring does not require group consensus. This is because we find that endogenous peer monitoring leads to similar levels of public good provision as our exogenous group monitoring regime, which closely resembles the setup commonly used in previous experiments. We need not question the previous results but rather should be cautious in interpreting them when monitoring is costly and requires a large portion of the group to support implementation.

Second, this paper illustrates that when a population's appetite for monitoring bad acts

is unknown, then monitoring done at the peer level will lead to greater public good provision than group monitoring that requires consensus. Our endogenous peer monitoring regime is just one example of a regime that leads to similar levels of public good provision and net earnings without assuming monitoring is automatic or done at the group level. In short, when a community might let free riding go undiscovered, relying on the efforts of a few peer monitors is likely best.

# References


M. Agranov, J. K. Goeree, J. Romero, and L. Yariv. What makes voters turn out: The effects of polls and beliefs. *Journal of the European Economic Association*, 16(3):825–856, 2017.

J. Alm, E. Kirchler, and S. Muehlbacher. Combining psychology and economics in the analysis of compliance: From enforcement to cooperation. *Economic Analysis and Policy*, 42(2):133–151, 2012.

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

L. R. Anderson, G. DeAngelo, W. Emons, B. Freeborn, and H. Lang. Penalty structures and deterrence in a two-stage model: Experimental evidence. *Economic inquiry*, 55(4): 1833–1867, 2017.

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

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

J. Andreoni and R. Petrie. Public goods experiments without confidentiality: a glimpse into fund-raising. *Journal of Public Economics*, 88(7-8):1605–1623, 2004.

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

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

A. Bar-Ilan and B. Sacerdote. Response to fines and probabilities in a natural experiment. *Journal of Law and Economics*, 47(1):1–17, 2004.

L. Boosey and R. M. Isaac. Asymmetric network monitoring and punishment in public goods experiments. *Journal of Economic Behavior & Organization*, 132:26–41, 2016.

J. P. Carpenter, S. Kariv, and A. Schotter. Network architecture and mutual monitoring in public goods experiments. *The Review of Economic Design*, 2012.

M. Casari and C. R. Plott. Decentralized management of common property resources: experiments with a centuries-old institution. *Journal of Economic Behavior & Organization*, 51(2):217–247, 2003.

T. N. Cason and F. U. Khan. A laboratory study of voluntary public goods provision with imperfect monitoring and communication. *Journal of Development Economics*, 58(2): 533–552, 1999.

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

A. Cohen. *The Perfect Store: Inside EBay*. Little, Brown, 2003. ISBN 9780316164931. URL https://books.google.com/books?id=cANUr5KyTOsC.

P. Dal Bó, A. Foster, and L. Putterman. Institutions and behavior: Experimental evidence on the effects of democracy. *American Economic Review*, 100(5):2205–29, 2010.

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

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

G. DeAngelo and T. Smith. Private security, piracy and the international provision of public goods. *Journal of Risk & Uncertainty*, 2020.

G. DeAngelo, B. Humphreys, and I. Reimers. Are public and private enforcement complements or substitutes evidence from high frequency nhl. *Journal of Economic Behavior & Organization*, 2017.

D. Dickinson and M.-C. Villeval. Does monitoring decrease work effort?: The complementarity between agency and crowding-out theories. *Games and Economic Behavior*, 63(1): 56–76, 2008.

M. Egas and A. 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.

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

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

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

E. Fatas, M. A. Melendez-Jimenez, and H. Solaz. An experimental analysis of team production in networks. *Experimental Economics*, 13(4):399–411, 2010.

E. Fehr and S. 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.

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

E. Fehr and T. Williams. Social norms, endogenous sorting and the culture of cooperation. *Working Paper*, 2018.

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

S. Fischer, K. Grechenig, and N. Meier. Monopolizing sanctioning power under noise eliminates perverse punishment but does not increase cooperation. *Frontiers in Behavioral Neuroscience*, 10:180, 2016.

L. Friesen. Certainty of punishment versus severity of punishment: An experimental investigation. *Southern Economic Journal*, 79(2):399–421, 2012.

A. Fuster and S. Meier. Another hidden cost of incentives: The detrimental effect on norm enforcement. *Management Science*, 56(1):57–70, 2010.

S. Gächter, E. Renner, and M. Sefton. The long-run benefits of punishment. *Science*, 322 (5907):1510–1510, 2008.

L. K. Gee and M. J. Schreck. Do beliefs about peers matter for donation matching? experiments in the field and laboratory. *Games and Economic Behavior*, 107:282–297, 2018.

K. Grechenig, A. Nicklisch, and C. Thöni. Punishment despite reasonable doubt a public goods experiment with sanctions under uncertainty. *Journal of Empirical Legal Studies*, 7 (4):847–867, 2010.

B. Greiner. Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125, 2015. ISSN 2199-6784. doi: 10.1007/s40881-015-0004-4. URL http://dx.doi.org/10.1007/s40881-015-0004-4.

Y. Halevy. Ellsberg revisited: An experimental study. *Econometrica*, 75(2):503–536, 2007. ISSN 1468-0262. doi: 10.1111/j.1468-0262.2006.00755.x. URL http://dx.doi.org/10.1111/j.1468-0262.2006.00755.x.

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

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

R. Moir. Spies and swords. *Environmental Economics, Experimental Methods*, 10:212, 2007.

A. Nicklisch, K. Grechenig, and C. Thöni. Information-sensitive leviathans. *Journal of Public Economics*, 144:1–13, 2016.

A. Nicklisch, L. Putterman, and C. Thöni. Trigger-happy or precisionist? on demand for monitoring in a noisy social dilemma game. *Working Paper*, 2019.

C. Noussair and S. Tucker. Public observability of decisions and voluntary contributions in a multiperiod context. *Public Finance Review*, 35(2):176–198, 2007.

E. Ostrom, J. Walker, and R. Gardner. Covenants with and without a sword: Self-governance is possible. *American Political Science Review*, 86(02):404–417, 1992.

A. Ramalingam, S. Godoy, A. J. Morales, and J. M. Walker. An individualistic approach to institution formation in public good games. *Journal of Economic Behavior & Organization*, 129:18–36, 2016.

M. Rege and K. Telle. The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics*, 88(7):1625–1644, 2004.

J. M. Spraggon, L. A. V. Sobarzo, and J. K. Stranlund. A note on stochastic public revelation and voluntary contributions to public goods. *Economics Letters*, 126:144–146, 2015.

A. Traulsen, T. Röhl, and M. 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.

J.-R. Tyran and L. P. Feld. Achieving compliance when legal sanctions are non-deterrent. *Scandinavian Journal of Economics*, 108(1):135–156, 2006.

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

# Online Appendix

## A Other voting rules

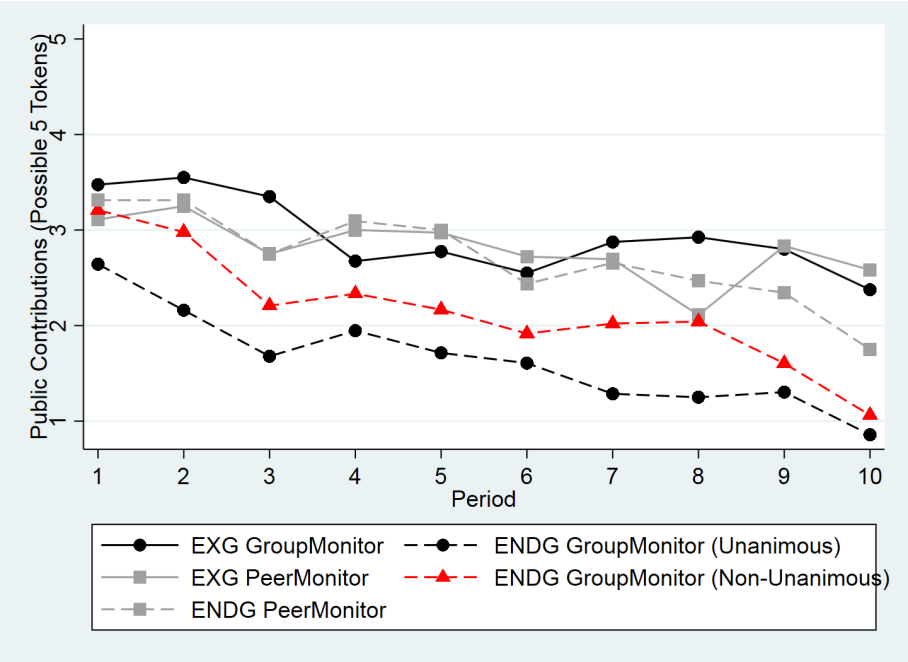



Figure A1: Public Contributions (Replicates data reported in Figure 3)

Table A1: **Correlation between willingness-to-pay and public contribution choice (replicates data reported in Table 3)**

| | WTP to Monitor | | |
|---|---|---|---|
| | **1** | **2** | **3** |
| | **ENDG GroupM (Non-unan.)** | **ENDG GroupM (Unan.)** | **ENDG PeerM** |
| Public Cont. | 0.109*** | 0.107*** | 0.038 |
| | (0.020) | (0.017) | (0.020) |
| Constant | 0.133* | 0.141** | 0.147** |
| | (0.043) | (0.043) | (0.041) |
| Adj. R2 | 0.158 | 0.166 | 0.024 |
| Obs. | 480 | 560 | 320 |

*Notes:* + significant at 10%; * significant at 5%; ** significant at 1%, *** significant at .1%. In all models, wild bootstrapped standard errors are clustered at the group level. There were a total of 136 subjects in 34 fixed groups of four, who each played for 10 periods. Contributions could be between zero and five tokens. In PeerMonitor games, WTP to Monitor is the proportion of subjects who paid $1 to monitor their right-side group member. In GroupMonitor (Unanimous) games, WTP to Monitor is the proportion of subjects who pledged $1 toward a group monitoring regime, but if the group did not collect $4, this amount was refunded back. The coefficient on $PublicContribution$ in the two GroupM models are statistically significantly different from the PeerM model $Prob > chi2 = 0.0060$ and $Prob > chi2 = 0.0096$

Table A2: **Exogenous vs. endogenous number monitored in $t-1$ and public contributions in $t$ with both unanimous and nonunanimous GroupMonitor games (replicates data reported in Table 4)**

| | Public Contribution$_t$ | |
|---|---|---|
| | **1** | **2** |
| | **ENDG GroupM (Non-unan.)** | **ENDG GroupM (Unan.)** |
| NumMonitored$_{t-1}$ | 0.011 | 0.011 |
| | (0.107) | (0.107) |
| EndogM*NumMonitored$_{t-1}$ | 0.332* | 0.598** |
| | (0.153) | (0.212) |
| EndogM | -0.728 | -1.078* |
| | (0.436) | (0.445) |
| Constant | 2.798*** | 2.798*** |
| | (0.388) | (0.388) |
| Adj. R2 | 0.037 | 0.083 |
| Obs. | 1404 | 1476 |
| NumMonitored$_{t-1}$+ | .342*** | 0.609** |
| ENDG*NumMonitored$_{t-1}$ | (0.109) | (0.183) |

*Notes:*+ significant at 10%; * significant at 5%; ** significant at 1%, *** significant at .1%. In column 1, the model includes data from the EXG GroupMonitor, ENDG GroupMonitor (Non-unanimous), EXG PeerMonitor, and the ENDG PeerMonitor games. In column 2, the model includes data from the EXG GroupMonitor, ENDG GroupMonitor (Unanimous), EXG PeerMonitor, and the ENDG PeerMonitor games. Standard errors are clustered at the group level. In column 1, there were a total of 156 subjects in 39 fixed groups of four, who each played for 10 periods. In column 2, there were a total of 164 subjects in 41 fixed groups of four, who each played for 10 periods. Contributions could be between zero and five tokens.

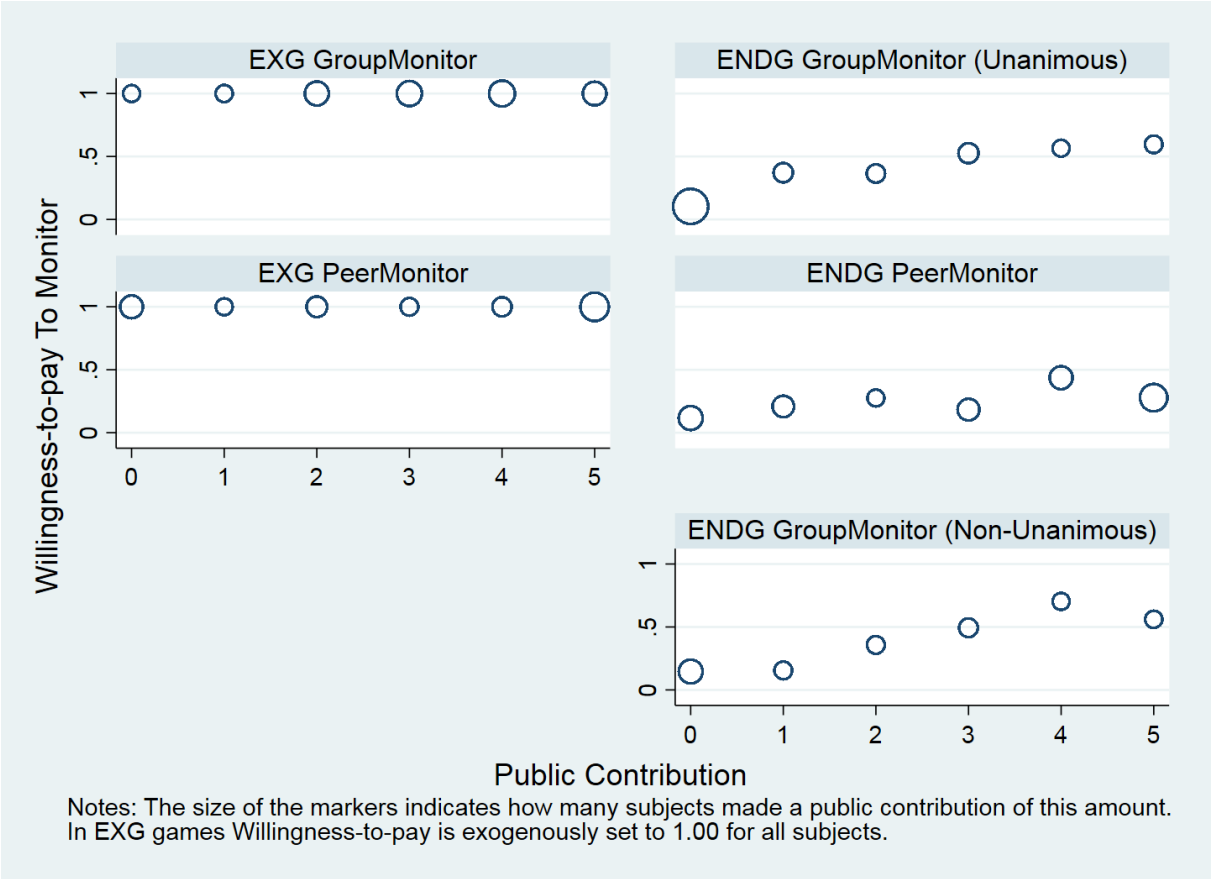



Figure A2: Correlation between WTP to monitor and contribution (replicates data reported in Figure 4)

# B Other punishment mechanisms

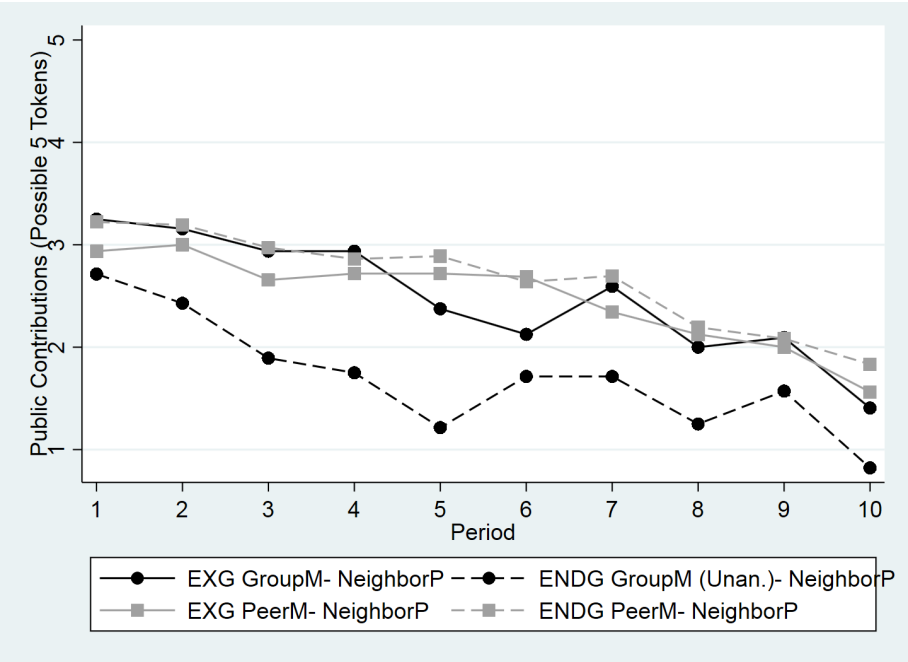



Figure B1: Public contributions with neighbor punishment

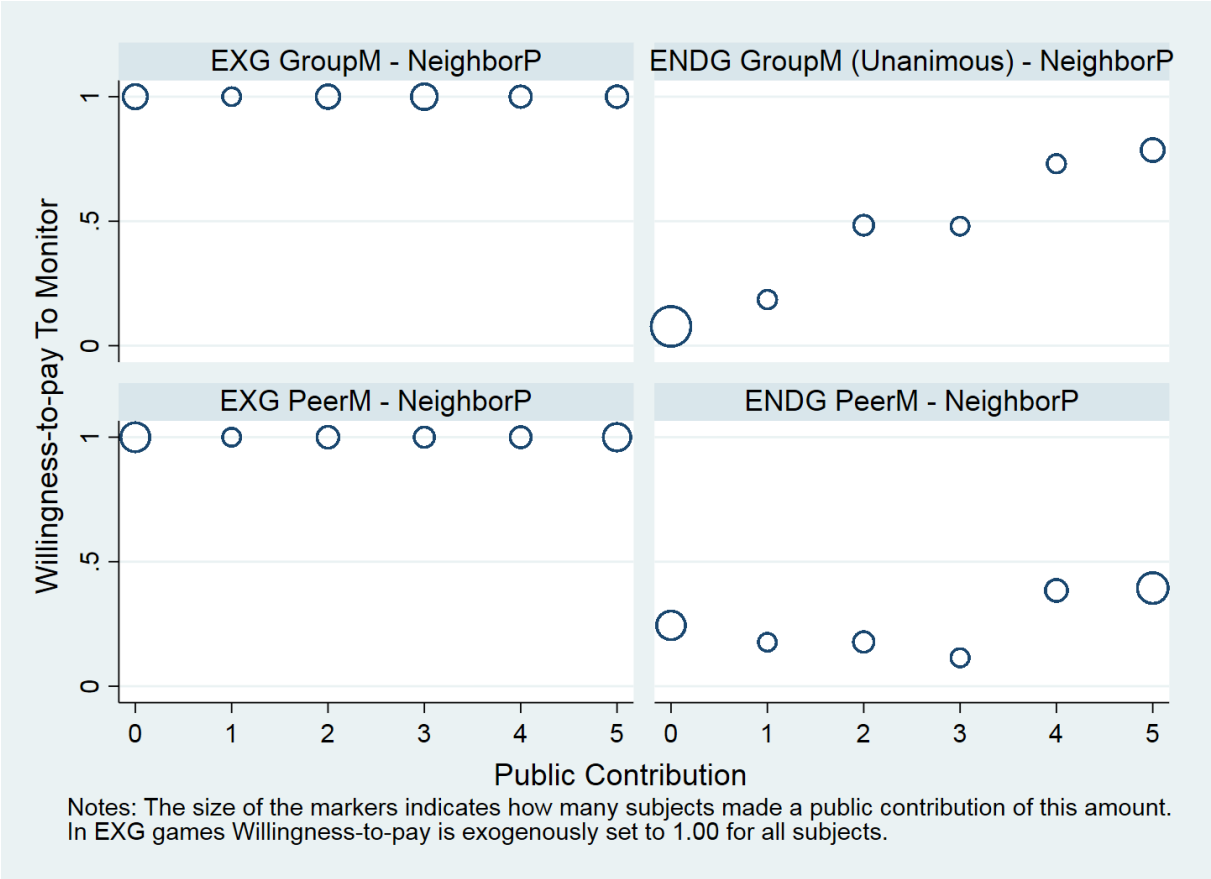



Figure B2: Correlation between willingness-to-pay to monitor and public contribution with neighbor punishment

Table B1: **Correlation between willingness-to-pay and public contribution choice with neighbor punishment (similar to Table 3)**

| | WTP to Monitor | |
|---|---|---|
| | **1** | **2** |
| | **ENDG** | **ENDG** |
| | **GroupM(Unan.)** | **PeerM** |
| | **NeighborP** | **NeighborP** |
| Public Cont. | 0.148*** | 0.036 |
| | (0.005) | (0.020) |
| Constant | 0.084* | 0.186+ |
| | (0.030) | (0.082) |
| Adj. R2 | 0.356 | 0.022 |
| Obs. | 280 | 360 |

*Notes:* + significant at 10%; * significant at 5%; ** significant at 1%, *** significant at .1%. In all models, wild bootstrapped standard errors are clustered at the group level. Contributions could be between zero and five tokens. In PeerMonitor games, WTP to Monitor is the proportion of subjects who paid $1 to monitor their right-side group member. In GroupMonitor (Unanimous) games, WTP to Monitor is the proportion of subjects who pledged $1 toward a group monitoring regime, but if the group did not collect $4, this amount was refunded back. The coefficient on Public Contribution in the GroupM model is statistically significantly different from the PeerM model.

Table B2: **Exogenous vs. endogenous number monitored in $t-1$ and public contributions in $t$ (similar to Table 4)**

| | **Public Contributions$_t$** |
|---|---|
| NumMonitored$_{t-1}$ | -0.021 |
| | (0.078) |
| EndogM*NumMonitored$_{t-1}$ | 0.622** |
| | (0.207) |
| EndogM | -0.649+ |
| | (0.348) |
| Constant | 2.469*** |
| | (0.256) |
| Adj. R2 | 0.040 |
| Obs. | 1152 |
| NumMonitored$_{t-1}$ + | .600*** |
| ENDG*NumMonitored$_{t-1}$ | (0.192) |

*Notes:* + significant at 10%; * significant at 5%; ** significant at 1%, *** significant at .1%. This table presents results from the EXG GroupMonitor-NeighborPunishment, ENDG GroupMonitor (Unanimous)-NeighborPunishment, EXG PeerMonitor-NeighborPunishment, and the ENDG PeerMonitor-NeighborPunishment games. Standard errors are clustered at the group level. There were a total of 128 subjects in 32 fixed groups of four, who each played for 10 periods. Contributions could be between zero and five tokens.

# C Additional Analysis

Figure C1 shows the relationship between *relative* free riding and willingness-to-pay for monitoring. This complements Figure 4 in the main text, which shows the relationship between *absolute* free riding and willingness-to-pay for monitoring.

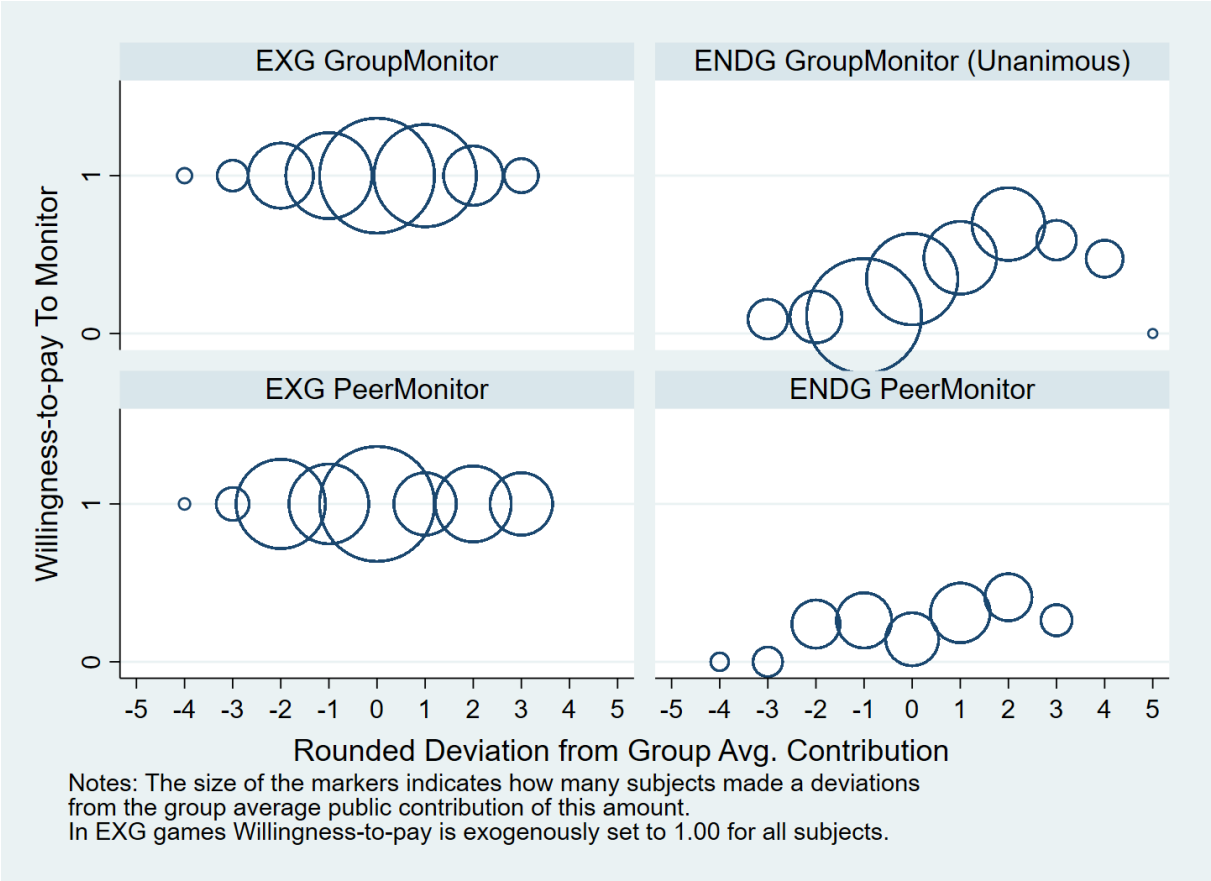



Figure C1: Correlation between willingness-to-pay to monitor and $\text{PublicContribution}_{i,t} - \text{MeanGroupPublicContribution}$

First, we compute the deviation between person $i$'s contribution in period $t$ from their group's mean contribution over all 10 periods. We then round that deviation to the nearest integer. Figure C1 shows the correlation between individual $i$'s willingness-to-pay for monitoring in period $t$, and $i$'s rounded deviation from the group's mean contribution. Larger circles indicate that a greater proportion of subjects have this size deviation. The vertical placement of a circle indicates the proportion of subjects who pledge money toward monitoring during the same period they had a certain deviation from the group mean (e.g., -5, -4, ..., 4, 5 tokens). Recall that in both exogenous games, subjects are forced to pay for monitoring, so that vertical placement of all circles is at 1.00 in the graphs in the left-hand column of Figure C1.

The story told by Figure C1 is similar to that told by Figure 4 in the main text. Concentrating on the top-right panel, we see in the ENDG GroupMonitor (Unanimous) game that there is a negative correlation between relative free riding and willingness-to-pay for monitoring. Moving to the bottom-right panel in the ENDG PeerMonitor game, we see no strong correlation between relative free riding and willingness-to-pay for monitoring, so the right-hand side panels of Figure C1 look very similar to the right-hand side panels of Figure 4 in the main text.

Figure C2 shows the relationship between *group* free riding and *group* willingness-to-pay for monitoring. This complements Figure 4 in the main text, which shows the relationship between *individual* free riding and *individual* willingness-to-pay for monitoring. We plot the mean group willingness-to-pay for monitoring against the mean group public contribution over all 10 periods of play. Each point in Figure C2 represents a single group.

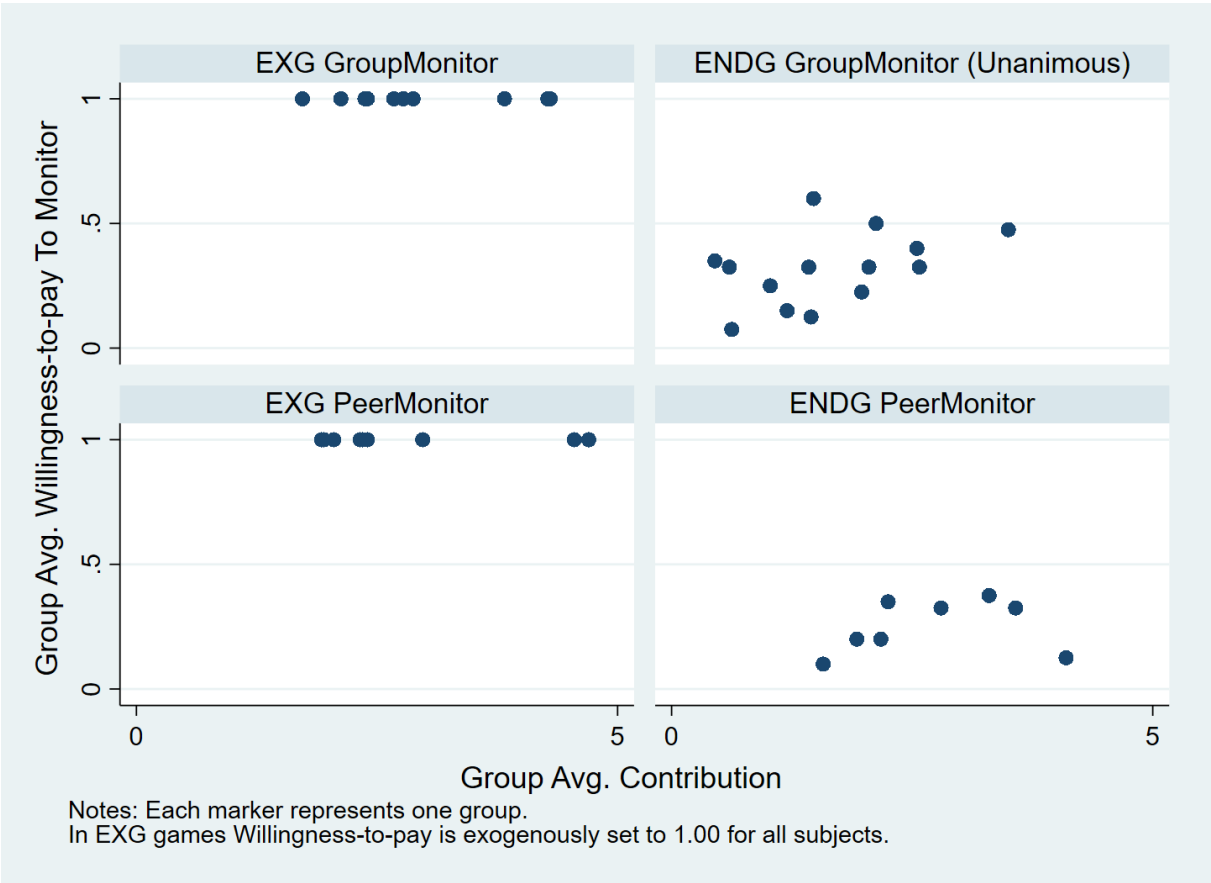



Figure C2: Correlation between group willingness-to-pay to monitor and group public contribution

Starting with the top-right panel, we see in the ENDG GroupMonitor (Unanimous) game that there is a negative correlation between group free riding and group willingness-to-pay for monitoring, which is very similar to the story told by Figure 4 in the main text.

Turning to the bottom-right panel, we see an inverted U-shape rather than the more erratic pattern in Figure 4 in the main text. Figure 4 in the main text shows a lack of a correlation between *individual* free riding behavior and willingness-to-pay for monitoring. In contrast, Figure C2 shows that *both* groups with many free riders and very few free riders are less likely to pay for monitoring. This may be because groups with very few free riders have no bad actions to monitor and groups with many free riders do not want to catch their own bad actions.