Exp Econ (2015) 18:154–171
DOI 10.1007/s10683-014-9392-1

ORIGINAL PAPER

Gunning for efficiency with third party enforcement in threshold public goods

James Andreoni · Laura K. Gee

Received: 3 September 2013 / Accepted: 9 January 2014 / Published online: 29 January 2014

Abstract When public goods can only be provided when donations cross a minimum threshold, this creates an advantage in that Pareto Efficient outcomes can be Nash Equilibria. Despite this, experiments have shown that groups struggle to coordinate on one of the many efficient equilibria. We apply a mechanism used successfully in continuous public goods games, the Hired Gun Mechanism (Andreoni and Gee in J. Public Econ. 96(11–12):1036–1046, 2012), to see if it can successfully get subjects across the threshold. When we use the mechanism to eliminate only inefficient equilibria, without addressing coordination, there is a modest but statistically insignificant improvement with the mechanism. However, when we hone the mechanism to eliminate all but one of the provision-point equilibria, thereby addressing the coordination issue, the mechanism moves all subjects to the desired efficient outcome almost immediately. In fact, after only one round using the hired gun mechanism, all subject are coordinating on the chosen equilibrium. The mechanism can be applied in settings

Andreoni would like to thank the National Science Foundation (SES-1024683), and the Science of Generousity Initiative for financial support. This research was approved by the UCSD IRB. We would also like to thank Mark Isaac, James Walker, two anonymous referees, Christopher Cotton, Jennifer Coats, Joseph Falkinger, Rosemarie Nagel, David Scmidtz, Jeff Zabel, and participants at the ESA and BABEEW conferences for their helpful comments.

Electronic supplementary material The online version of this article (doi:10.1007/s10683-014-9392-1) contains supplementary material, which is available to authorized users.

J. Andreoni
Department of Economics, University of California, San Diego, La Jolla, CA 92093, USA
e-mail: andreoni@ucsd.edu

J. Andreoni
NBER, 1050 Massachusetts Ave., Cambridge, MA 02138, USA

L.K. Gee (✉)
Department of Economics, Tufts University, Medford, MA 02155, USA
e-mail: laura.gee@tufts.edu

where a group (1) has a plan for public good provision, (2) can measure contributions, (3) can fine members and (4) has an agreed upon standard for expected contributions. In these settings simple punishments, when focused on solving coordination as well as free riding, can greatly improve efficiency.

Keywords Public goods · Experiment · Laboratory · Equilibrium selection · Punishment · Free riding

JEL Classification C72 · C91 · C92 · D7 · H41 · H42

1 Introduction

Previous experimental work has found that there is serious under provision of public goods in the lab setting. If the public good is linear, the Nash equilibrium predicts zero provision and in experiments subjects move toward this full free riding equilibrium with repeated play (Ledyard 1995). When public goods can only be provided when contributions reach a minimum threshold, this creates an advantage in that Pareto efficient outcomes can be Nash equilibria. Yet, in threshold public goods games experiments, we still see significant under-provision of the public good (Croson and Marks 2000).

So, it is natural to ask, if we were to add some minimal extra structure, could we overcome free riding? A sanctioning mechanism is a common starting point (Falkinger et al. 2000; Ostrom et al. 1992; Yamagishi 1986). We would want our mechanism to be something that is simple, and low cost to enforce.

In an earlier paper (Andreoni and Gee 2012), we aimed this question at continuous linear public goods and introduced what we called the Hired Gun mechanism. The inspiration is that, in the real world, groups are often able to overcome free riding by adopting a sanctioning mechanism, perhaps at a small cost, that has the following flavor: First, hire or appoint someone with a small bit of authority (a team leader, a professor for a class, or a municipal authority). Second, set a fair standard of compliance that is Pareto improving if followed by everyone. Third, set up small punishments to the least compliant individual. In our specific mechanism, the punishments are chosen to be just big enough that the least compliant person would have been better off as the second least compliant person. This transforms the game to a "race for second place," which sets off a dynamic that pushes groups toward complete compliance, all while exacting typically small punishments.

Using lab experiments, we find this hired gun mechanism improves provision in the linear public goods game, and does so quite dramatically. Here we ask whether this mechanism can also help with threshold public goods provision. Threshold public goods are known to suffer from two challenges. First is standard free riding; pledging zero is one of the Nash equilibria. The second is coordination. Typically there is a large set of Pareto efficient equilibria in these games and, with simultaneous play, it is hard for any player to know which equilibria is being selected by others, and if their contribution is pivotal (Corazzini et al. 2013; Ostrom 2001;

Bagnoli and Lipman 1989). The problem goes beyond an informational one, however; when an experimenter tries to coordinate players by suggesting an equilibrium, if players are homogenous the suggestion is ineffective (Croson and Marks 2001).

To increase public good provision in the threshold public goods case, theoretically we only need to push subjects toward the multiple interior equilibria. However, even if we eliminate the full free riding equilibria subjects may still not be able to coordinate on a single efficient equilibrium. In this paper we test a hired gun mechanism which pushes subjects toward many efficient equilibria (GunEff) against one that pushes subjects toward a single selected equilibrium (GunSelect).

Because individuals may care about equality (Engelmann and Strobel 2004; Andreoni and Miller 2002; Bolton and Ockenfels 2000; Fehr and Schmidt 1999), a natural starting point is to select a single equilibrium as the most socially "fair" outcome. In our case, this is where all subjects make equal contributions. We find that the mechanism which eliminates inefficiency mildly improves public good provision, but that the mechanism which selects a single equilibrium greatly enhances provision of the public good in the lab.

Does this finding have any implications for the real-world? Many groups struggle with the efficient provision of public goods, from small teams working toward a collective project to large municipalities trying to protect a common pool resource from exploitation. The hired gun mechanism can be implemented in a setting with the following attributes: (1) there is a group with a plan for public good provision, (2) we can approximate or measure exact contributions, (3) there is a way to sanction group members either through withholding benefits or assigning fines and (4) an agreed upon standard for expected contributions. There are many examples of real world groups that meet these criteria including: a group of students working on a project, a team of lawyers working on a single case, a community group fundraising for a shared resource, or a municipality trying to protects its watershed from depletion by its citizens.[1]

Our results suggest that students working on a group project could be asked to rank the level of effort by each group member and the largest free riders could receive a lower grade; a common practice in MBA courses. Teams at a law firm working on a single case could review their colleagues, and those reviews could help determine annual bonuses. A home owners' association fundraising for new community benches could withhold prizes, like named recognition, from those who do not make large enough pledges.[2] Municipalities could measure individual water consumption and single out the largest users to be punished. In fact, the city of San Antonio already

[1] In some of these examples contributions cannot be fully refunded after they have been contributed (e.g. time given to a group project), while others can be refunded (e.g. donations pledged toward fundraising). We test the hired gun mechanism in a setting with a money back guarantee, which is more closely related to the pledges which can be refunded. We chose the money back guarantee because it has been shown to increase provision of threshold public goods, and we wanted to show if we could improve on this best-case baseline (Isaac et al. 1989).

[2] In fact, online fundraising websites like neighbor.ly already give prizes like stickers and t-shirts to more generous contributors, so in essence they withhold prizes from lower contributors.

publicly lists the largest water users each year as a form of sanctioning the biggest free riders.[3]

In the next section we describe how the hired gun mechanism applies to this domain, and derive the equilibrium predictions. Section 3 presents our experimental procedures. Section 4 gives our results, and Section 5 concludes.

2 Games

The experiment contains three different games all using a threshold public goods game similar to one used in previous studies (Cadsby et al. 2008; Croson and Marks 1999, 2000, 2001; Marks and Croson 1998, 1999; Cadsby and Maynes 1999; Suleiman and Rapoport 1992). Below we present the details for these three games: first our specific baseline threshold public goods game (TPG), then a game with a hired gun to increase efficiency (GunEff), and last a game with a hired gun that coordinates players on the selected equitable equilibrium (GunSelect).

2.1 Threshold public goods (TPG) game

Subjects in groups of four allocate 8 tokens between a public and a private good. Tokens invested in the private good pay \$2 to only the individual who made the investment. If the sum of contributions to the public good $\sum_{j=1}^{4} g_j$ is greater than or equal to the threshold of 20 tokens, then the public good is provided. If not, there is a money back guarantee, and each subject earns \$16, the value of the 8 token endowment invested in the private good (Coats et al. 2009; Dawes et al. 1986). If the sum of contributions is strictly greater than 20 tokens, there is no rebate and the over-payment is a deadweight loss. By construction, subjects have homogeneous endowments and returns from the public good. These were chosen such that in the GunSelect game the symmetric equilibrium coincides with equitable contributions.[4] Let g_i be player i's contribution to the public good.

The earnings for a subject for a period are:

$$\pi_i^{TPG} = \begin{cases} 2(8 - g_i) + 20 & \text{if } \sum_{j=1}^{4} g_j \geq 20 \\ 16 & \text{if } \sum_{j=1}^{4} g_j < 20 \end{cases}$$

There are many Nash equilibria for this game. The inefficient equilibria occur when groups contribute less than a total of 20 tokens and no player can deviate so that 20 tokens is reached: for example $(0, 0, 0, 0)$, $(0, 3, 4, 4)$, $(1, 2, 3, 4)$, or $(0, 0, 5, 6)$. Additionally, efficient equilibria occur when groups meet the threshold

[3] Thank you to Aaron Schroeder for suggesting this example (http://www.mysanantonio.com/news/environment/article/Biggest-water-users-revealed-4147851.php).

[4] Additionally we chose homogenous endowments and valuations because Croson and Marks (2001) found that in the heterogenous case the mere suggestion of contributions had an effect on actions. Thus we are biased away from suggestion driving our results by choosing the homogenous setting.

exactly, $\sum_{j=1}^{4} g_j = 20$. For example: (8, 8, 4, 0) or (8, 6, 4, 2) or (5, 5, 5, 5).[5] Only one of these efficient equilibria will also be what we will refer to as "fair": (5, 5, 5, 5). This threshold public goods (TPG) game is the basic framework for all games in the experiment, and acts as our control.

2.2 Delegated sanctioning: the hired gun mechanism

Our goal is a low cost mechanism that eliminates inefficient equilibria. Our 2012 hired gun mechanism is one example of such a device. The intuition for the hired gun comes from two sources. First is simple observation of real life mechanisms. Speeding tickets are not generally issued to everyone on the freeway, but rather are assigned to the fastest car on the road. To avoid a ticket, one only needs to be the second fastest car. Likewise, the largest shirker on the team project will be pulled aside for a performance review, and the smallest bonus will go to the lawyer with the least billable hours. That is, enforcement of compliance in organizations in the real world often focuses first, and often exclusively, on the most egregious violators.[6]

The second source of intuition is from the Guessing Game of Nagel (1995) (see also the Keynesian p-beauty contest games of Ho et al. (1998)). Players choose a number between zero and 100, and the one who chooses the number closest to two-thirds of the average of all the guesses is the winner. As long as there is common knowledge of rationality, people realize that, through iterated deletion of dominated strategies, the only way for everyone to be two-thirds of the average is if they all guess zero, the Nash equilibrium. Our mechanism turns this intuition upside down. Here the loser is the one who gained the most by deviating from full compliance, and the penalty is enough to make them wish they were the second biggest cheater. The only way in which everyone can be the second biggest cheater is if no one cheats at all. That is, full compliance with the chosen standard, be it efficiency or fairness, becomes the new equilibria.

2.3 Gunning for efficiency (GunEff) game

Our primary goal is removing inefficient equilibria from the TPG game. Inefficiency is eliminated, if the hired gun mechanism punishes the lowest contributor when the threshold is not met ($\sum_{j=1}^{4} g_j < 20$). The punishment must be non-zero, so we chose \$2.25 to be comparable to our GunSelect treatment. If many players tie for lowest contributor, all those tied are punished. The punishment only occurs when the threshold isn't met, so initial earnings are \$16 and the lowest contributor(s) earn \$13.75 (\$16 minus \$2.25).

[5] Given our parameter choices any set of contributions that exactly meet the threshold are a Nash equilibria of the game, so we avoid the "cheap riding" problem described by Isaac et al. (1989) where some sets of contributions meeting the threshold are not equilibria.

[6] See Savikhin Samek and Sheremeta (2014) for a mechanism where the sanctioning takes the form of listing the largest free riders rather than a monetary fine. See Bornstein et al. (2002) for a mechanism where the largest free riding team is punished rather than the largest free riding individual.

Formally, let g_z denote the contribution of the lowest contributor to the public good, $g_z = \min\{g_1, g_2, g_3, g_4\}$. Let g_y denote the second lowest contribution, $g_y = \min\{g_1, g_2, g_3, g_4 \backslash g_z\}$. The payoffs can be summarized by:

$$\pi_i^{GunEff} = \begin{cases} 2(8 - g_i) + 20 & \text{if } \sum_{j=1}^4 g_j \geq 20 \\ 16 & \text{if } \sum_{j=1}^4 g_j < 20 \text{ and } g_i \neq \min\{g_1, g_2, g_3, g_4\} \\ 13.75 & \text{if } \sum_{j=1}^4 g_j < 20 \text{ and } g_i = \min\{g_1, g_2, g_3, g_4\} \end{cases}$$

The Nash equilibria include any combination of contributions summing to exactly the threshold of 20. Notice if the good isn't provided, being the lowest contributor by choosing $g_z < g_y$ results in earning \$13.75 instead of \$16. This choice is strictly dominated by a choice of $g_i = g_y + \epsilon > g_y$, where $\epsilon > 0$ is the smallest positive increment of g_i. That is, the best response of the lowest contributor is to be just slightly higher than g_y, when the good isn't provided.

Setting $g_i = 0$ is never a best response. With common knowledge of rationality, a subject knows that no one will choose g_i to be zero, because this guarantees punishment. So, a subject chooses g_i equal to the next discrete amount above zero, $g_i = 1$. But knowing everyone else is using similar reasoning, subjects will choose the next discrete amount above $g_i = 1$, and move to $g_i = 2$. In short, the best response for any player when the threshold isn't met is to find the lowest contribution, and set their contribution slightly above it until the threshold is met. After the threshold has been met, the best response is to be the lowest contributor to the public good. The Nash equilibria include any combination of contributions summing exactly to 20 tokens. This leads us to our first prediction:

Prediction 1 *We expect to see more efficient Nash equilibria play in GunEff than TPG, because the inefficient equilibria are no longer present in the GunEff game. That is we expect more choices of contributions such that* $\sum_{j=1}^4 g_j = 20$ *in GunEff than in TPG.*

2.4 Gunning for selecting equity (GunSelect) game

Although eliminating the inefficient equilibria is important, individuals may want to select an equilibrium with equitable distribution. Because our players are homogeneous by construction, the "fair" action is naturally for each subject to pay 5 tokens towards the public good. We make this the unique equilibrium of the game by extending our hired gun so that it punishes even when the threshold *is* met.

When the threshold is met ($\sum_{j=1}^4 g_j \geq 20$) but the lowest contributor gives less than their fair share of 5 tokens ($g_z < 5$), then that lowest contributor is punished. The deduction makes the lowest contributor just slightly worse off (in terms of net subgame payoff) than the second lowest contributor. The lowest contributor earns the amount that the second lowest contributor earns minus a constant, say M. M must be strictly greater than the value of one unit of the private good (\$2), so we set $M = \$2.25$, which now should make clear the reason for choosing \$2.25 as the punishment in the GunEff treatment. If all players give at least 5 tokens, then there is no punishment. Using the definitions of g_y and g_z above, the punishment P for player z is equal to:

$$2(g_y - g_z) + 2.25 \quad \text{if} \sum_{j=1}^{4} g_j \geq 20 \text{ and } g_z = \min\{g_1, g_2, g_3, g_4\} < 5$$

We can in turn write payoffs as:

$$\pi_i^{GunSelect} = \begin{cases} 2(8 - g_i) + 20 & \text{if } \sum_{j=1}^{4} g_j \geq 20 \text{ and } g_i \geq 5 \\ 2(8 - g_i) + 20 - P & \text{if } \sum_{j=1}^{4} g_j \geq 20 \text{ and } g_i = \min\{g_1, g_2, g_3, g_4\} < 5 \\ 16 & \text{if } \sum_{j=1}^{4} g_j < 20 \text{ and } g_i \neq \min\{g_1, g_2, g_3, g_4\} \\ 13.75 & \text{if } \sum_{j=1}^{4} g_j < 20 \text{ and } g_i = \min\{g_1, g_2, g_3, g_4\} \end{cases}$$

Recall in the GunEff game that the remaining Nash equilibria were those where the threshold of 20 was exactly met. Now even when the threshold is met a subject will be punished if contributing less than 5 tokens. The punishment is such that the lowest contributor makes slightly less than the second lowest contributor, so once again the best response is to give just slightly more than the second lowest contributor. Now, we have selected a unique Nash equilibrium with symmetric equally shared cost of public good provision $g_i = 5$ for all four subjects. With this unique equilibrium the subjects coordinate on the equitable outcome. This gives us our second prediction:

Prediction 2 *We expect to see more symmetric efficient Nash equilibrium play in GunSelect. That is we expect more choices of* $(5, 5, 5, 5)$ *in GunSelect than in either GunEff or TPG.*

3 Procedures

Each session involved 12 subjects and 20 periods: 10 periods of the Threshold Public Goods game followed by 10 periods of either (1) 10 more periods of the threshold public goods game without a mechanism (TPG), (2) the efficiency hired gun mechanism (GunEff), or (3) the hired gun mechanism which selects the unique fair equilibrium (GunSelect). Each session was conducted using z-tree software (Fischbacher 2007), lasted under 90 minutes and subjects earned \$18 on average.

To minimize repeated game effects, participants were randomly and anonymously re-matched into a new group of 4 participants at the beginning of each period.[7] Subjects were given instructions for the first 10 periods of play, a quiz, and then played that game for 10 periods. This was done again before periods 11–20. To remove experimenter effects, all sessions were run by the same person. Subjects were informed

[7]Note, the use of strangers matching also is known to add variance to the data (Andreoni and Croson 2008), which handicaps the analysis against finding significant effects.

that they would be paid for a single randomly selected period from the 20 periods in the session.[8]

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 referring to all punishments as "deductions." The example of fair contributions (5, 5, 5, 5) was used in all three treatments to try to keep salience the same across all treatments. Merely suggesting the symmetric equally shared outcome has been shown to have no effect on efficiency (Croson and Marks 2001), so we are confident that our results are driven by the possibility of punishment rather than an implicit suggestion of equal contributions. Full instructions and screen shots are available in the online Appendix.

Subjects were always informed on the decision screen if a punishment mechanism was in place during that period. In all treatments, after all subjects made contribution choices, they were given anonymous information about the contributions to the public good, private good, and initial earnings for each of their group members. In the GunEff and GunSelect treatments they were also anonymously informed if another subject had been punished and by how much at the end of each Period.

4 Results

Table 1 contains summary statistics for the treatments. For brevity, we have pooled the first 10 periods of play across all three treatments (TPG, GunEff, and GunSelect). In the first columns of Table 1 we report the individual level statistics: average contribution, percent of subjects giving 0, percent who gave the fair contribution of 5, and the absolute deviation from 5 tokens, which is a measure of inefficiency.

Table 1 also lists the collective contributions to the public good, $\sum_1^4 g_i$, and average earnings. Contributions, public good provision, and net earnings are higher in the GunEff and GunSelect treatments than in the TPG treatment where there is no mechanism. However, although contributions are highest in GunEff the net earnings are lower than in GunSelect due to overpayment for the public good.

In the next sections we test our predictions. Because the groups are randomly rematched each period, the actual level of public good provision is partially driven by chance, so we report average realized outcomes as well as the average for all 495 unique groups which could have been formed each period.[9]

[8]To choose the random period after the end of the 20th period, a subject was given a 20 sided die. The subject was asked to verify if the die had 20 sides, roll the die, and announce the outcome on the die out loud.

[9]There are 12 subjects in each session in each period, so there are 12 choose 4, or 495 unique combinations. To see how actual versus possible groups may change the outcomes consider an example where three of the 12 subjects chose $g = 0$ while the remaining 9 subjects each chose $g = 5$. If the three selfish subjects were each assigned to different groups, then no group would reach the threshold [(0, 5, 5, 5); (0, 5, 5, 5); (0, 5, 5, 5)], while if they ended up together in one group then two of three groups would succeed [(0, 0, 0, 5); (5, 5, 5, 5); (5, 5, 5, 5)]. In both cases, however, the choices of subjects were the same but the impression of the outcome is different.

Table 1 Summary statistics of experimental results

Treatment	Average cont. g_i	Percent $g_i = 0$	Percent $g_i = 5$	Average $\lvert 5 - g_i \rvert$	Average $\sum_1^4 g_i$	Average earnings	Average punishment
Rounds 1–10:							
TPG	4.88	2 %	61 %	0.60	19.54	$21.75	–
Rounds 11–20:							
TPG	4.98	0 %	67 %	0.42	19.93	$22.27	–
GunEff	5.07	0 %	74 %	0.28	20.28	$23.34	$0.15
GunSelect	5.01	0 %	97 %	0.03	20.03	$25.73	$0.01

108 subjects total, 36 subjects per treatment, 3 Sessions per treatment

10 Periods of each game per Session, 3 Groups per Session, 4 Subjects per Group

TPG: threshold public good, GunEff: promotes efficiency, GunSelect: select equitable equilibrium

4.1 All public good provision

Although our predictions are specifically about efficient or equitable public good provision, we still care if the threshold is met in any manner. Figure 1 shows the percentage of groups each period who met or exceeded the threshold.

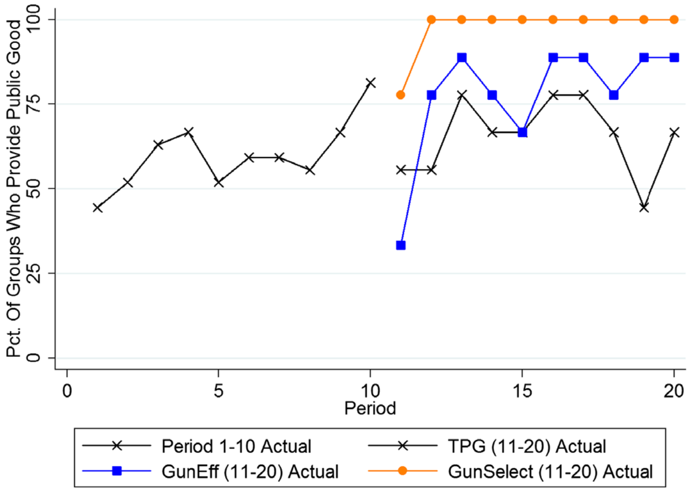
Panel (a) of Fig. 1 shows the average of the realized groups, while panel (b) shows the average of the 495 possible groups. In periods 1–10 subjects provide the public good about 40 to 80 percent of the time, so there is not a huge under-provision problem. In Periods 11–20 if there is no mechanism (TPG) the public good continues to be provided about 60 % of time. For most periods, the GunEff mechanism has similar levels of provision to having no mechanism (TPG). However, there are some larger improvements in periods 19 and 20, so it may be that GunEff mechanism takes some time to increase provision.

In fact, the realized level of provision in the GunEff versus the TPG treatment are not statistically significantly different from each other.[10] However, using the possible permuted groups there is an improvement in provision between GunEff versus TPG.[11] The level of provision may be improved by punishing for efficiency, but this depends on the "luck" of being in a good group, and the length of time the mechanism has been in place.

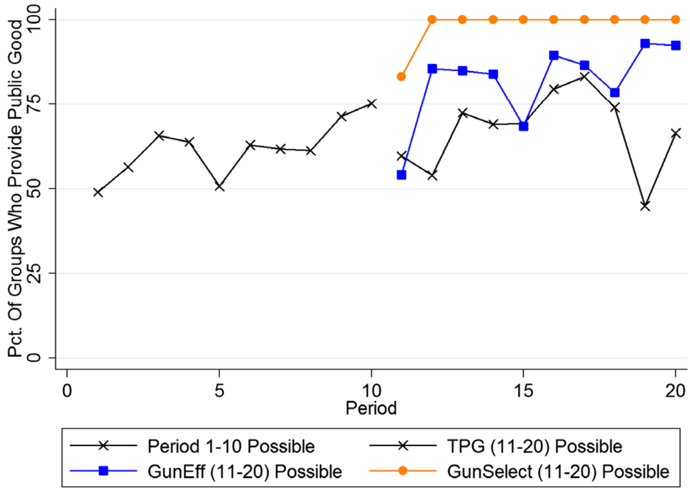
In contrast, public good provision in the GunSelect treatment, which pushes subjects to the selected equitable equilibrium, jumps almost immediately up to 100 %.

[10]For actual level of public good provision $p = 0.60$ using a Kolomogrov-Smirnov test at the session level. We use a Kolomogrov-Smirnov test because we only have 3 observations at the session level for the TPG (11–20) and GunEff games. The same result can be shown with a random effects regression $P > |z| = 0.112$.

[11]For the possible level of public good provision using all the possible group permutations $p = 0.10$ using a Kolomogrov Smirnov test at the session level. The same result can be shown with a random effects regression $P > |z| = 0.001$.



(a) Actual Outcomes



(b) Possible Outcomes

Fig. 1 Actual and possible percentage of groups providing public good

The level of provision is robustly significantly higher in GunSelect than in either of the other treatments.[12]

Although general provision is interesting, our predictions are specifically about efficient or equitable provision. Table 2 summarizes the average level of overall, efficient and equitable provision by treatment.

[12]For actual or possible level of public good provision $p = 0.10$ using a Kolomogrov Smirnov test at the session level. The same result can be shown with a random effects regression $P > |z| = 0.000$.

Table 2 Percentage of groups providing public good

Treatment	Provision of PG: $\sum_1^4 g \geq 20$		Efficient provision: $\sum_1^4 g = 20$		Equitable & efficient: $g_i = 5$, all i	
	Actual	Possible	Actual	Possible	Actual	Possible
Rounds 1–10:						
TPG	60 %	62 %	32 %	30 %	19 %	21 %
Rounds 11–20:						
TPG	66 %	67 %	24 %	29 %	17 %	19 %
GunEff	78 %	82 %	39 %	45 %	30 %	34 %
GunSelect	98 %	98 %	91 %	91 %	91 %	91 %

"Actual" is mean of realized groups. "Possible" is mean of all 495 potential groups. 108 subjects total, 36 subjects per treatment, 3 Sessions per treatment

10 Periods of each game per Session, 12 Subjects per Session, 4 Subjects per Group TPG: threshold public good, GunEff: promotes efficiency, GunSelect: select equitable equilibrium

4.2 Efficient public good provision

The GunEff mechanism should eliminate below-threshold equilibria, so efficient provision should be higher in GunEff than in TPG. In Fig. 2 we show the proportion of groups providing the public good efficiently each period by treatment. Panel (a) graphs the realized groups, and panel (b) is the average of 495 possible groups.

The GunEff mechanism results in slightly higher levels of efficient provision than the baseline TPG. However, this difference is insignificant.[13] When looking at the trend in Fig. 2 the GunEff mechanism is more effective over time, so it may be that subjects simply need to get used to the mechanism. On the other hand, the GunSelect mechanism quickly increases the level of efficient public good provision, with all the groups jumping to efficiency after only 3 periods.[14]

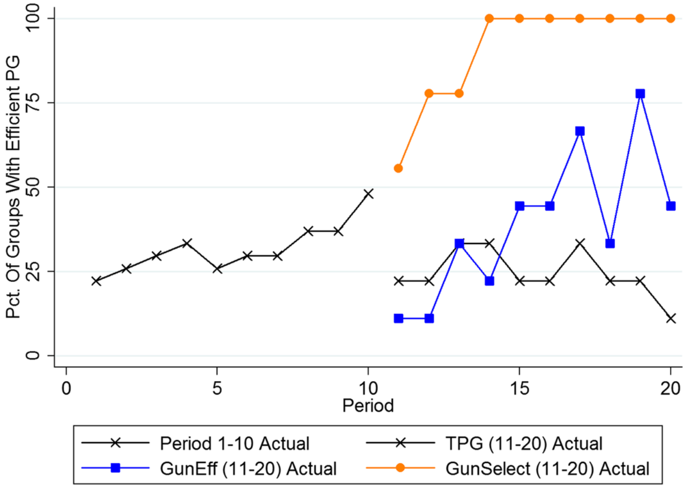
Result 1 We do not find support for Prediction 1 because there is no significant difference in the level of efficient public good provision between the baseline game (TPG) when compared to a game with a mechanism which should remove inefficient equilibria (GunEff). We do however see an improvement in efficiency with the addition of a mechanism which selects the equal contributions equilibrium (GunSelect).

4.3 Equitable public good provision

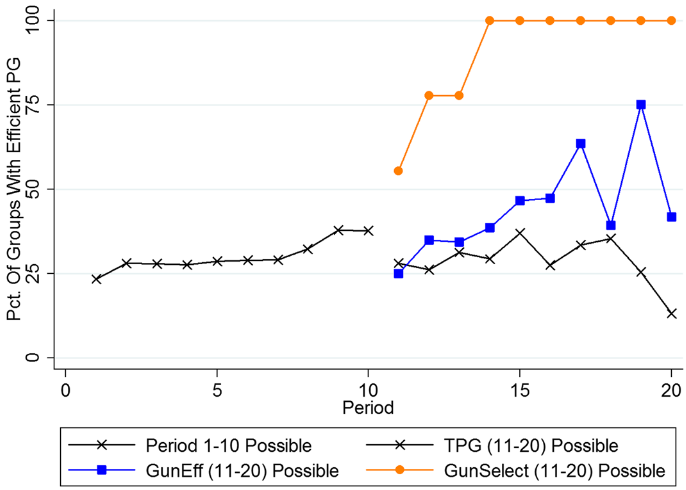
The GunSelect mechanism pushes subjects to coordinate on the selected unique equal contributions equilibrium. Figure 3 shows the proportion of groups providing the public good equitably each period by treatment. Panel (a) graphs the realized groups,

[13]For actual (possible) level of efficient public good provision $p = 0.60$ ($p = 0.40$) using a Kolomogrov Smirnov test at the session level. The same results can be shown with a random effects regression.

[14]For actual and possible level of efficient public good provision $p = 0.10$ using a Kolomogrov Smirnov test at the session level. The same results can be shown with a random effects regression.



(a) Actual Outcomes

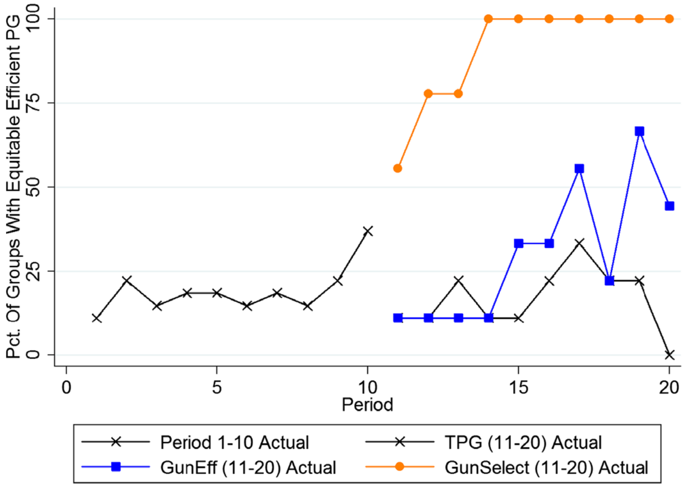


(b) Possible Outcomes

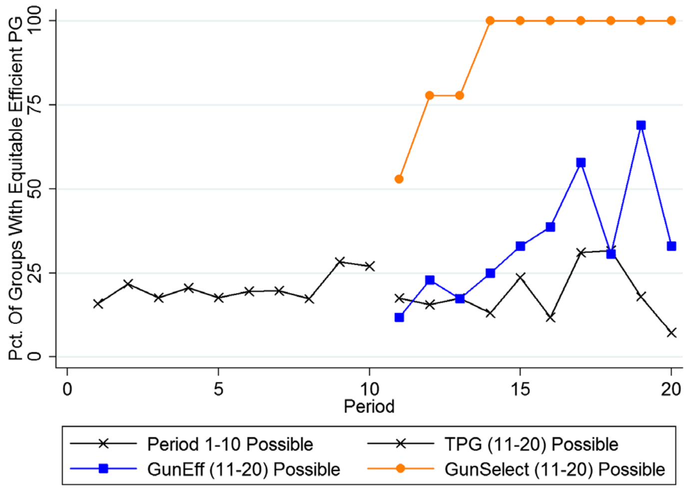
Fig. 2 Actual and possible percentage of groups efficiently providing public good

and panel (b) is the 495 possible groups. Equitable provision is extremely low in the other treatments, reaching less than 20 % under TPG, and inching up to 30 % in GunEff (see Table 2). In contrast, in GunSelect equitable provision occurs 90 % of the time.[15] So, we find strong support for our prediction that the public good will be provided more equitably under the GunSelect mechanism.

[15]The difference is statistically significant. For the actual and possible level of equitable and efficient public good provision $p = 0.10$ using a Kolomogrov Smirnov test at the session level. The same results can be shown with a random effects regression.



(a) Actual Outcomes



(b) Possible Outcomes

Fig. 3 Actual and possible percentage of groups fairly providing public good

Result 2 We find support for Prediction 2 because the equitable and efficient provision of the public good rises from 20–30 % up to 90 % when the GunSelect mechanism is in place.

A byproduct of speedy convergence to the equitable outcome is there are very low punishment costs in the GunSelect treatment as pictured in Fig. 4.[16] This is particu-

[16]The difference is statistically significant. For the actual proportion of groups punished $p = 0.10$ and level of punishment $p = 0.10$ using a Kolomogrov Smirnov test at the session level.

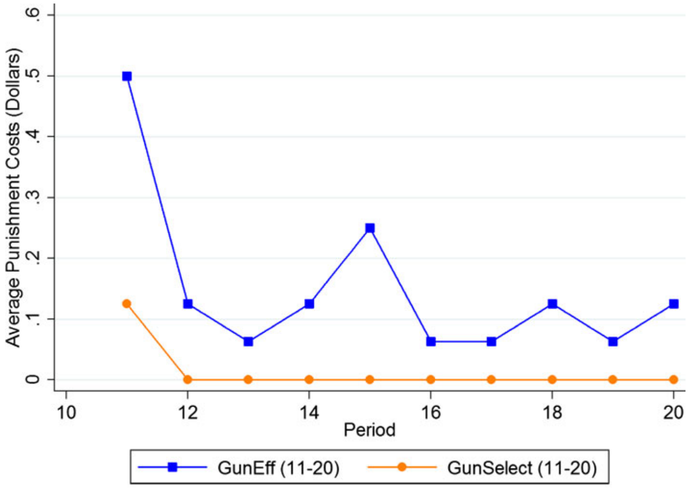


Fig. 4 Actual average costs of punishment in GunEff and GunSelect

larly surprising because punishment is always a flat \$2.25 in the GunEff treatment, whereas it can be as high as \$18.25 in the GunSelect treatment and can be levied in more circumstances.[17] In spite of the fact that punishments can be larger and happen under more circumstances, punishment costs are actually lower in GunSelect than in GunEff.

Result 3 Punishment is levied less often and is less costly under the mechanism selecting the equitable equilibrium (GunSelect) than under the mechanism only promoting efficiency (GunEff). This is especially surprising given that GunSelect has more opportunities for and higher possible levels of punishments than GunEff.

4.4 What is different about GunEff?

Our GunSelect mechanism is very effective at pushing groups toward public good provision. Yet, our GunEff mechanism doesn't push groups toward efficient provision in the same way. Here we offer speculation as to why.

It could be that GunSelect makes the equitable outcome more salient, but in an effort to keep the salience similar across treatments, we included an example of giving $(5, 5, 5, 5)$ in all our instructions.[18] It could also be that the GunSelect mechanism gives an implicit suggestion for contributions. However, previous studies have found that suggested contributions when there are symmetric endowments and payoffs do

[17]This would happen if the largest free rider gave 0 tokens and all the other players gave 8 tokens: $P = 2(8 - 0) + 2.25 = 18.25$. In fact, in the GunSelect treatment only 2 subjects are ever punished at the low level of \$2.25.

[18]Additionally in all treatments we stated that "if each person in your group invests 5 tokens in the BLUE investment, this will be the most equal way to reach 20 tokens." Recall the BLUE investment is the public good.

not affect contributions (Croson and Marks 2001). The GunSelect mechanism punishes both actual and intended free riding, while the GunEff mechanism only punishes actual free riding. So, it may be punishing intentions is the key to the success of the GunSelect mechanism. However, we find that patterns of actions are not statistically significantly different across the two treatments.[19] Beyond salience, suggestion, and punishing intentions, it may be that learning is more pronounced in the GunSelect mechanism.

A measure of a slow learner could be that a player receives a punishment but does not respond with a larger contribution in the next round. Call such a player an *unimproving player*. In the GunEff treatment there are four unimproving players, whereas there are none in the GunSelect treatment.[20] While this, along with the low punishment costs in GunSelect described in the prior section, could indicate a small role for faster learning in GunSelect. They could also provide evidence for a main hypothesis of this paper: Because the GunSelect mechanism focused on a single equilibrium, subjects could form better guesses about whether they were pivotal or would face a sanction in the next round.

This raises a second question that is potentially important for future research. We selected the symmetric equilibrium since it seemed most natural and realistic as a starting point for our hired gun mechanism. However, what if we had more asymmetry in our baseline, and chose to incentivize an equilibrium that was also asymmetric in some way? For instance, suppose subjects had unequal endowments, returns, or price of punishment (Anderson and Putterman 2006; Croson and Marks 2001; Rapoport and Suleiman 1993; van Dijk and Grodzka 1992). We could imagine trying to coordinate on either equal contributions (but different payoffs), or equal payoffs (but different contributions). Would our mechanism work equally well in these cases as it did in the symmetric case? We view this as an important question that could be the topic of study itself.

5 Conclusion

Threshold public goods games have the advantage that some of the equilibria are Pareto efficient. However, experimental games have revealed that subjects have a difficult time reaching an equilibrium that crosses the threshold, and inefficiency is a

[19] We thank an anonymous reviewer for suggesting this alternative explanation for why the GunSelect mechanism works so well. Punishment is only realized in the GunEff treatment if the threshold is not met, so it only punishes actualized free riding instead of the intention to free ride. In the GunEff treatment when the threshold is met, the best response is to be the lowest contributor. In contrast the GunSelect treatment punishes both actual and intended free riding, so the best response is always the fair contribution of 5. If a subject was in a group which met the threshold in period T, then that subject might want to lower her contribution in period $T+1$ in the GunEff treatment, but not in the GunSelect treatment. We found that subjects lower their contribution in the period after the public good is provided by -0.08 tokens in GunEff and by -0.02 in GunSelect, however this difference is not statistically significantly different (Kolomogrov Smirnoff $p=0.04$).

[20] We ran three sessions of the GunEff treatment. Two of those sessions had a single unimproving player, and the remaining session had two unimproving players.

common outcome. There are likely two reasons for this. First is standard free riding. The second is coordination; because there are typically many efficient equilibria, participants have difficulty selecting one.

Small self-governing groups innovate to overcome inefficiencies by adopting some mild sanctioning rules, perhaps enforced by a delegated leader like the class professor, the team leader, the home owners' association, or the municipality official. We stylize this idea in a simple and low-cost mechanism we have called the hired gun mechanism. The mechanism can be applied in settings where a group (1) has a plan for public good provision, (2) can measure contributions, (3) can sanction members and (4) has a standard for expected contributions. The sanctions, moreover, tend to be small. They are just enough to make the least compliant person wish they had been the second-least compliant person. This mechanism, by creating a "race for second place," drives everyone toward the standard and, as a result, enforces equilibria that are consistent with the standard. In the real world this punishment could take the form of lower grades on group projects, below average bonuses for shirking workers, fewer prizes for poor fundraisers, or fines for the biggest takers from a common pool resource.

We compare two standards of compliance. First is efficiency—meeting the threshold—and second is efficiency and fairness—is the threshold met at equal contributions by all. The first standard eliminates all inefficient equilibria, addressing free riding, while the second goes further by eliminating all equilibria but the symmetric efficient equilibrium, addressing coordination as well.

Our experimental test consists of three treatments. In the first 10 rounds of all three, subjects play a standard homogeneous threshold public goods game, with strangers matching each round. This gives all subjects comparable experience with the game. Our manipulations come in rounds 11 to 20. We first have a baseline condition that continues without a hired gun mechanism. Second, the GunEff treatment has a hired gun punish only when the threshold is not met. Third, the GunSelect treatment punishes whenever the equitable equilibrium is not met.

First, we find that GunEff increases provision of public good from the baseline by 18 %. This is a modest, but not statistically significant improvement. Second, and most importantly, GunSelect increases provision from the baseline by 48 % and, from round 12 onward 100 % of the groups are providing the public good. After one round of the GunSelect mechanism, all of our subjects were playing the symmetric Pareto efficient equilibrium and paying no punishment costs. In fact, only two of 36 subjects in GunSelect ever received a punishment. Importantly, the GunSelect mechanism not only dramatically helped reach the threshold, but it increased efficiency further by not exceeding the threshold.

In sum, using a simple and very low cost delegated punishing mechanism, the hired gun mechanism, we are able to encourage people to play the symmetric equilibrium almost immediately. Our study does, however, raise some interesting questions for future research. Most importantly, while it is natural that we selected the symmetric equilibrium for GunSelect to enforce, would the mechanism have worked just as well if we had solved the coordination problem at another asymmetric equilibrium? We view this as an obvious and important next step in research on delegated enforcement.

References

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

Andreoni, J., & Croson, R. (2008). Partners versus strangers: random rematching in public goods experiments. In C. R. Plott & V. L. Smith (Eds.), *Handbook of experimental economics results* (Vol. 1, pp. 776–783). Amsterdam: Elsevier. Chap. 82.

Andreoni, J., & Gee, L. K. (2012). Gun for hire: delegated enforcement and peer punishment in public goods provision. *Journal of Public Economics*, *96*(11–12), 1036–1046.

Andreoni, J., & Miller, J. (2002). Giving according to GARP: an experimental test of the consistency of preferences for altruism. *Econometrica*, *70*(2), 737–753.

Bagnoli, M., & Lipman, B. L. (1989). Provision of public goods: fully implementing the core through private contributions. *Review of Economic Studies*, *56*(4), 583–601.

Bolton, G. E., & Ockenfels, A. (2000). ERC: a theory of equity, reciprocity, and competition. *The American Economic Review*, *90*(1), 166–193.

Bornstein, G., Gneezy, U., & Nagel, R. (2002). The effect of intergroup competition on group coordination: an experimental study. *Games and Economic Behavior*, *41*(1), 1–25.

Cadsby, C. B., & Maynes, E. (1999). Voluntary provision of threshold public goods with continuous contributions: experimental evidence. *Journal of Public Economics*, *71*(1), 53–73.

Cadsby, C. B., Croson, R., Marks, M., & Maynes, E. (2008). Step return versus net reward in the voluntary provision of a threshold public good: an adversarial collaboration. *Public Choice*, *135*(3/4), 277–289.

Coats, J. C., Gronberg, T. J., & Grosskopf, B. (2009). Simultaneous versus sequential public good provision and the role of refunds—an experimental study. *Journal of Public Economics*, *93*(1–2), 326–335.

Corazzini, L., Cotton, C., & Valbonesi, P. (2013). Too many charities? Insight from an experiment with multiple public goods and contribution thresholds. Marco Fanno working paper.

Croson, R., & Marks, M. (1999). The effect of heterogenous valuations for threshold pubic goods: an experimental study. *Risk Decision and Policy*, 99–115.

Croson, R., & Marks, M. (2000). Step returns in threshold public goods: a meta- and experimental analysis. *Experimental Economics*, *2*(3), 239–259.

Croson, R., & Marks, M. (2001). The effect of recommended contributions in the voluntary provision of public goods. *Economic Inquiry*, *39*(2), 238–249.

Dawes, R. M., Orbell, J. M., Simmons, R. T., & Van De Kragt, A. J. C. (1986). Organizing groups for collective action. *American Political Science Review*, *80*(4), 1171–1185.

Engelmann, D., & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *The American Economic Review*, *94*(4), 857–869.

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

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868.

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

Ho, T.-H., Camerer, C., & Weigelt, K. (1998). Iterated dominance and iterated best response in experimental "p-beauty contests". *The American Economic Review*, *88*(4), 947–969.

Isaac, R., Schmidtz, D., & Walker, J. (1989). The assurance problem in a laboratory market. *Public Choice*, *62*, 217–236. doi:10.1007/BF02337743.

Ledyard, J. (1995). Public goods: a survey of experimental research. In J. Kagel & A. Roth (Eds.), *Handbook of experimental economics*. Princeton: Princeton University Press.

Marks, M., & Croson, R. (1998). Alternative rebate rules in the provision of a threshold public good: an experimental investigation. *Journal of Public Economics*, *67*(2), 195–220.

Marks, M., & Croson, R. (1999). The effect of incomplete information in a threshold public goods experiment. *Public Choice*, *99*, 103–118. doi:10.1023/A:1018316500800.

Nagel, R. (1995). Unraveling in guessing games: an experimental study. *The American Economic Review*, *85*(5), 1313–1326.

Ostrom, E. (2001). Social dilemmas and human behavior. In *Economics in Nature* (pp. 23–41).

Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a sword: self-governance is possible. *American Political Science Review*, *86*(2), 404–417.

Rapoport, A., & Suleiman, R. (1993). Incremental contribution in step-level public goods games with asymmetric players. *Organizational Behavior and Human Decision Processes*, *55*(2), 171–194.

Savikhin Samek, A., & Sheremeta, R. M. (2014). Recognizing contributors: an experiment on public goods. *Experimental Economics*. doi:10.1007/s10683-013-9389-1.

Suleiman, R., & Rapoport, A. (1992). Provision of step-level public goods with continuous contribution. *Journal of Behavioral Decision Making*, *5*(2), 133–153.

van Dijk, E., & Grodzka, M. (1992). The influence of endowments asymmetry and information level on the contribution to a public step good. *Journal of Economic Psychology*, *13*(2), 329–342.

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, *51*(1), 110–116.