Journal of Public Economics 96 (2012) 1036–1046

Contents lists available at SciVerse ScienceDirect

Journal of Public Economics

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

Gun for hire: Delegated enforcement and peer punishment in public goods provision☆

James Andreoni [a,b,*], Laura K. Gee [a]

[a] University of California, San Diego, United States
[b] NBER, United States

ARTICLE INFO

Article history:
Received 2 May 2011
Received in revised form 3 August 2012
Accepted 10 August 2012
Available online 29 August 2012

JEL classification:
C72
C91
C92
D7
H41

Keywords:
Public goods
Experiment
Group behavior
Punishment
Free riding
Peer punishment

ABSTRACT

This paper compares two methods to encourage socially optimal provision of a public good. We compare the efficacy of vigilante justice, as represented by peer-to-peer punishment, to delegated policing, as represented by the "hired gun" mechanism, to deter free riding and improve group welfare. Small self-governing organizations often place enforcement in the hands of an appointed leader—the department chair, the building superintendent, and the team captain. This hired gun, we show, need only punish the least compliant group member, and then only punish this person enough so that the person would have rather been the second least compliant. The hired gun mechanism is an example of a low cost device that promotes complete compliance as the unique Nash equilibrium. We find that subjects are willing to pay to hire a delegated policing mechanism over 70% of the time and that this mechanism increases welfare between 15% and 40%. Moreover, the lion's share of the welfare gain comes because the hired gun crowds out vigilante peer-to-peer punishments.

1. Introduction

The title character of the 1950s television western show, Paladin, is described as a "gentleman" and "accomplished warrior" who "insists the *rule of law* be enforced, rejecting man-to-man frontier justice" (Hirschman, 2000). His calling card read simply, "Have Gun. Will Travel." The lawlessness of the "wild west" can be described in modern terms as "peer-to-peer punishment" in which scores are settled between parties, often with inefficient punishment. Paladin encouraged cooperative behavior by providing order through reason and, failing that, force. He was a gun for hire and was portrayed as a costly but superior alternative to shootouts, feuds, and endless retribution.

This paper provides a theoretical model and experimental analysis of Paladin. A special problem faced by small self-governing groups is that punishments meted out by members can often be quite deleterious. Peers often punish to the extent that they erase any gains brought on by the punishment, resulting in short-run net losses. Although a strong demand for peer-to-peer punishments exists in the laboratory setting, we observe little of this type of justice in the real world. Instead we often observe the development of delegated or appointed parties that sanction bad behavior. Consider the following examples: the homeowners' association, the building superintendent, the soccer coach, the department head, the committee chair, the parent and teacher association, and the synagogue or church elders. These authorities are created and often funded by a subset of the people and institutions that they monitor. One suspects that these mechanisms arise because they are more efficient than the alternative of vigilante justice. We see people and institutions choosing a hired gun instead of punishing each other for infractions in the real world, but a thorough investigation of the two different mechanisms has not been conducted.

We propose the "gun for hire" mechanism as one example of a third party mechanism based on a simple rule of punishment of noncompliance. The rule is low cost to enforce, in equilibrium results

☆ We would like to thank our colleagues at UCSD, participants in the Social Dilemmas Conference, participants at the ESA 2010 Conference, participants at the SWET 2011, Kristy Buzard, Simon Gachter, Nikos Nikiforakis and Anya Savikhina for comments. We would also like to thank three anonymous referees and the editor for their thoughtful and helpful comments. Andreoni also gratefully acknowledges the National Science Foundation, grant 1024683, for financial support.

* Corresponding author at: University of California, San Diego, United States.
E-mail address: andreoni@ucsd.edu (J. Andreoni).

0047-2727/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jpubeco.2012.08.003

in no punishments and full compliance, and when off the equilibrium path typically results in punishments that will be small. A central feature of the gun for hire is that the enforcer does not need to perfectly document all the noncompliance. The gun for hire only needs to know the exact actions of the two largest deviators from compliance. In many instances, the biggest deviators (think of loudest neighbors, worst teachers, most truant volunteers) are easy to identify. Moreover, the enforcer does not need to punish all non-compliant people, just the single biggest cheater. Finally, the punishment need not be large. It only needs to be just big enough that the most non-compliant person would rather have been the second most non-compliant person. If the second most non-compliant person is best responding to his or her environment, the two most non-compliant choices should be nearly identical, meaning that in expectation this difference should be trivial. It follows that punishments off the equilibrium path will likely be small. Hence, even if our mechanism requires some experience to reach equilibrium, the costs along that road should be minimal.

We show that (as seen on TV) a simple mechanism (a gun for hire) is an efficient and desirable substitute for lawless peer-to-peer punishment. We first use an experiment to show that our gun for hire mechanism works when it has been imposed exogenously on our subjects. Next, we show that subjects will choose to implement the gun for hire mechanism, and that it will work even when only a subset of subjects choose to implement the mechanism. Last, we show that even when vigilante justice is always available when there is a gun for hire, subjects discontinue the use of peer-to-peer punishments. That is, the mere presence of a centralized enforcement mechanism makes people less willing to employ vigilante justice.

In an effort to demonstrate the potential for research in this area, we use a series of linear public goods games to examine whether the gun for hire mechanism works first when it is exogenously assigned and second when it is endogenously chosen. In all of our games, subjects are randomly assigned to a group of four people, in which they are asked to allocate an endowment between a public good and a private good. We experiment with four enforcement regimes that players can use to discourage free riding. Subjects have three types of punishment conditions: an exogenously imposed gun for hire, an endogenously chosen gun for hire, peer-to-peer punishments only, or both peer-to-peer punishments with an endogenously chosen gun for hire. The gun for hire mechanism is meant to be just one of any number of examples of small scale self-policing devices; it is a stylized version of the homeowner's association or building superintendent. By looking at the peer-to-peer and gun for hire separately and jointly, we can identify their relative welfare effects.

We find that when our gun for hire is exogenously imposed it immediately improves welfare. When the mechanism is endogenously chosen, there is a significant demand for the gun for hire both when it is the only punishment option and when it is offered alongside peer-to-peer punishment. Welfare, as measured by group net earnings (that is earnings minus the costs of punishment), is significantly improved when groups can choose to hire a gun compared to when they can only peer-to-peer punish. Welfare is also improved when subjects can choose to hire a gun in addition to peer-to-peer punish compared to when they can only peer-to-peer punish. Furthermore, when both types of punishment are available and the gun is hired, the costs of peer-to-peer punishment decline precipitously.

In sum, when peer punishment is the only option, individuals use it, often with negative welfare consequences. When given the option of a centralized punishing mechanism, players prefer this to taking justice into their own hands; they cease to engage in peer punishment, and welfare improves dramatically. While our model and results are highly stylized, we will argue that the experimental observation is suggestive of a common real-world phenomenon: inefficient social institutions (such as peer-to-peer punishment) can be easily supplanted by lower cost, more efficient mechanisms that delegate enforcement.

2. Background

In previous experiments on costly peer-to-peer punishment, subjects can pay a fee to reduce the payoff to another subject in their group *only* once. While this type of peer-to-peer punishment leads to higher contributions to the public good, the effects on group welfare (group earnings minus punishment costs) have been ambiguous. Egas and Riedl (2008), Gachter et al. (2008), Herrmann et al. (2008), Botelho et al. (2007), Fehr and Gachter (2002), Fehr and Gachter (2000), and Ostrom et al. (1992) all found decreases in net earnings in the short run, while Masclet et al. (2003) found that adding a single round of punishment increased net earnings.[1] If the peer-to-peer punishment is repeated over many periods (50 periods of play, rather than 10) with the *same* groups intact Gachter et al. (2008) found a welfare improvement. In this case, it is possible that repeated interaction created reputation or reciprocity concerns that may have partially driven this result.

Notice that a single round of costly punishment does not take into account the possibility for revenge. When an opportunity for counter-punishment is added net earnings are dramatically reduced, as found by Denant-Boemont et al. (2007) and Nikiforakis (2008).[2] Hence, multiple rounds of costly punishment can create disastrous revenge cycles (Nikiforakis and Engelmann, 2011). One way to lower the costs is to allow non-monetary punishments such as disapproval messages or exposing only low contributors (Noussair and Tucker, 2005; Masclet et al., 2003; Savikhina and Sheremeta, 2010). Another avenue for lower costs is to allow subjects to threaten sanctions before contribution decisions are made (Masclet et al., 2011; Bochet and Putterman, 2007). The fact that people enjoy expressing their disapproval is convincingly shown by Fudenberg and Pathak (2010), who demonstrate that subjects still engaged in costly punishment even though it was not observed until the end of 10 rounds of play. In such a case, punishment logically could have no effect during the game. If people enjoy punishing, and if costly punishment is the only tool available, then the negative welfare effects of costly punishment are likely to be exacerbated by revenge cycles. It may be that people enjoy punishing because they believe these punishments carry out justice. If this is the case, a person may not actually want to carry out the punishing herself and would gladly hire a third party to mete out justice on her behalf.

This literature suggests that to improve welfare, we need to curb the enjoyment of punishment and prevent peer-to-peer revenge cycles. When the streets are full of vendettas, and desperados are roaming the frontier looking for a fight for fun, what do the town folks do? They call Paladin. That is to say that a natural method for solving both these problems is “hiring” or “appointing” someone to discipline the group. Note that by delegated we don't necessarily mean someone outside the group, but simply mean a commonly recognized conduit for complaints, who monitors and metes out punishments. The punishments need not be more severe than those available by peer-to-peer punishment. The key is that discipline is centralized and a credible threat. Some previous work has already shown that central coordination of punishment can be welfare improving both theoretically (Kube and Traxler, 2011; Boyd et al., 2010; Sigmund et al., 2010; Steiner, 2007) and in experiments (Baldassarri and Grossman, 2011; Dickinson and Villeval, 2008; Falkinger et al., 2000; Yamagishi, 1986).

[1] See Vesterlund (2012) for a review. Also note that if punishments are only carried out when at least two members of the group request them, then over time there is a welfare gain (Casari and Luini, 2009). While, if players receive a noisy signal of other group members' behavior then the addition of punishment is not only detrimental to welfare, but also decreases contributions to the public good. See Grechenig et al. (2010).

[2] Cinyabuguma et al. (2006) found that if subjects are not given information about who specifically punished them, then net earnings increase. This restriction on the information basically makes revenge motivated second round punishments impossible.

Yamagishi's experiment is most closely related to our gun for hire. Yamagishi allowed subjects to first play the public goods game and then contribute to a punishment fund that punished the lowest contributor to the public good. Unlike our mechanism, Yamagishi's punishment size was not related to the size of deviation from compliance. Yamagishi finds that public contributions were higher under punishment, but welfare was only improved under certain cost schemes. Although these results lend credence to the idea that there is a welfare gain from a delegated sanctioning mechanism, we believe that choosing the amount of punishment after the public contribution decisions is fundamentally different than choosing to hire a delegated mechanism before the public goods game has taken place. We also see our study as building on Yamagishi's insights by making punishments sensitive to the severity of the infraction.

If delegated punishment is the solution, will people voluntarily submit to a gun for hire? Clearly many positive examples exist in the real world on both a large and small scale, such as the police regulating public safety, the EPA assessing fines for emissions, the PTA socially penalizing those who don't sell raffle tickets, the building superintendent speaking to the noisy neighbors, or the department chair cracking down on bad teaching. There have been some experiments in which subjects have been able to choose if they would like to be punished either by each other (Sutter et al., 2010; Ertan et al., 2009; Gurerk et al., 2006; Botelho et al., 2007; Decker et al., 2003) or by a third party (O'Gorman et al., 2009; Kosfeld et al., 2009; Guillen et al., 2007; Tyran and Feld, 2006). These authors have found that, in some cases, subjects choose to allow punishing. Many of these experiments have made the implementation of a punishing mechanism monetarily costless. Monitoring, however, typically requires some resources or opportunity cost. By contrast, we make our punishment mechanism costly when it is endogenously chosen, but the cost is less than the gain realized through cooperation.

Our intuition for the hired gun comes from two sources. First is simple observation of real life mechanisms. Speeding tickets from police officers are not generally issued to everyone on the freeway going over the speed limit but rather to the fastest car on the road. To avoid a speeding ticket, one only needs to be the second fastest car. That is, enforcement of compliance in the real world often focuses first, and often exclusively, on the most egregious violators. The second source of intuition is from the Keynesian *p*-beauty contest games (Ho et al., 1998; Nagel, 1995). Imagine a game in which the winner of a prize is the person who guesses a number between 0 and 100 that is closest to two-thirds of the average of the others' guesses. As long as there is common knowledge of rationality, people will 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 0, which is the Nash equilibrium. Our mechanism turns this intuition upside down. Here the "loser" will be the largest free rider (that is, the one who gained the most by deviating from full compliance), and the penalty will be enough for her to wish she had been the second biggest cheater rather than the biggest. This gives everyone the incentive to be the second biggest cheater. The only set of choices in which everyone can avoid being the biggest cheater (again with common knowledge of rationality) is full compliance.

3. The games

The experiment contains five different public goods games. We use the linear public goods game with four players as the basic framework for each game, so we will begin by explaining the rules for this game.

3.1. The linear public goods (LPG) game

Subjects are given an "automatic payment" of $1 (to reduce within experiment income effects, as will be seen later) and an endowment of 5 tokens that they allocate between a public good and a private good. Each token invested in the public good pays a return of $2 to all group members for an aggregate social return of $8. Each token invested in the private good pays a return of $3 to only the individual who made the investment. Let g_i be player i's contribution to the public good. The earnings for a subject for a period are:

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

A selfish profit-maximizing player would choose to set $g_i = 0$ and if all players are selfish they will each earn $16. The group welfare maximizing level of contribution is $g_i = 5$. If all players choose this amount, their earnings would be $41 each. After all subjects have chosen g_i they are given anonymous information about the contribution to the public good, private good, and initial LPG earnings for each of their group members. This game will act as the basic framework for our other games. Next, we will explain how our gun for hire mechanism works when it has been exogenously imposed on players. We will call this the "gun hired" game.

3.2. The gun hired (GH) game

Subjects are given an "automatic payment" of $0.50 (to reduce within experiment income effects, as will be seen later) and an endowment of 5 tokens that they allocate between a public good and a private good. Subjects are informed that they are playing with a third party punishment mechanism that will punish the lowest contributor to the public good.

3.2.1. What the hired gun shoots

The gun for hire mechanism simply takes a deduction from the lowest contributor to the public good. The size of the bullet fired by the hired gun varies with the size of the infraction from the group behavior. The size of the deduction is set so as to make the lowest contributor to the public good just slightly worse off (in terms of net subgame payoff) than the second lowest contributor to the public good. In our mechanism the two payoffs will differ by the value of one unit of the private good, $3.

Formally, let g_z denote the contribution of the lowest contributor to the public good, $g_z = min\{g_1,g_2,g_3,g_4\}$. If there is a tie for the lowest contributor, then all those who tied will be punished. Let g_y denote the second lowest contribution to the public good, $g_y = min\{g_1,g_2,g_3,g_4 \backslash g_z\}$. The size of the punishment will be the difference between the initial payoffs of player z and player y plus a constant, M. We set M equal to the cost from taking one token of the player's private good, so $M = \$3$.

The punishment for player z is equal to:

$$P = \pi_z - \pi_y + 3 = 3\left(g_y - g_z\right) + 3.$$

In the special case in which all the players choose the same level of contribution to the public good, but still give below full contribution ($g_i = g_j < 5 \; \forall i,j$), all the subjects are punished P_0. We set P_0 to $3, the payoff from contributing a token to the private good. Lastly if all 4 subjects contribute the full 5 tokens to the public good, then no one is punished. To summarize, when the gun is hired, the size of the shot fired is equal to:

$$P = \begin{cases} 3 & \text{if} \quad g_i = g_j < 5 \text{ for all } i,j \\ 0 & \text{if} \quad g_i = 5 \text{ for all } i \\ 3\left(g_y - g_z\right) + 3 & \text{for lowest contributor}(s) \text{ in other cases} \end{cases}.$$

Subjects are aware of this punishment mechanism when they make their choices of contribution to the public good, g_i. After all

players have chosen g_i, they are given anonymous information about the contribution to the public good, private good, initial LPG earnings, size of punishment (if any), and final net payoffs for each of their group members.

3.2.2. GH equilibrium

Notice that any choice of $g_z < g_y$ will result in the subject earning \$3 less than player y. 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. The choice of $g_i = g_y + \epsilon > g_y$ will result in no punishment. That is, the best response of the lowest contributor is to change g_z to be just slightly higher than g_y. If all subjects reason this way, it is never a best response to set $g_i = 0$. Knowing that all subjects will not choose to set g_i to zero, a subject will choose g_i equal to the next discrete amount above zero, $g_i = 1$. But then knowing that everyone else is using similar reasoning, subjects will want to choose the next discrete amount above $g_i = 1$, and so they need to move to $g_i = 2$. In short, the best response for any player is to find what the lowest level of contribution is, and to set their contribution slightly above it. The only fixed point is full contribution to the public good $g_i = 5$. See the online Appendix for the generalized model and proofs.

The game is like a p-beauty contest (Nagel, 1995) in reverse. Each player is trying to guess the lowest amount given by the others in her group and then wants to give the closest contribution above that amount possible. This thought process eventually pushes all the players to contribute all of their endowment to the public good. Each player should choose $g_i = 5$ and will earn \$40 in the game (this does not include the \$0.50 fixed payment). We will compare earnings under this exogenously imposed mechanism to earnings when the mechanism is endogenously chosen. In the next section we explain how the game is played when we allow players to pay a fee to implement the gun for hire.

3.3. The gun for hire (G4H) game

The gun for hire (G4H) game is very similar to the gun hired (GH) game. The only difference is that we add a pre-play stage 1. In stage 1, each subject is given an endowment of 4 tokens worth \$0.25 each. Subjects choose $e_i, 0 \le e_i \le 4$, to contribute to the "hiring fund." If the sum of the 4 person group's contributions reach a threshold of 8 tokens, a delegated punishment mechanism will be implemented in stage 2. Subjects' stage 1 earnings equal the number of tokens they kept multiplied by \$0.25. Over-payments for hiring the gun are not refunded to the subjects. If the threshold for hiring is not met, subjects are refunded their e_i and earn \$1 in stage 1. Thus, if the gun is hired, then subjects play the aforementioned gun hired game. Note that it costs 8 tokens per group or 2 tokens per person (on average) to hire the gun. Because each token is worth \$0.25, the average cost is \$0.50 per person, which is equivalent to the automatic payment in the gun hired (GH) game. Also if the gun is not hired, then all tokens offered in stage 1 are refunded to each player, which is equivalent to the value of the "automatic payment" in the basic LPG game.

In short, the gun for hire game goes as follows in stage 1: subjects choose whether to hire the third party punishing mechanism. If they do not hire, they play the regular linear public goods game (LPG) with equilibrium earnings in the sub-game of \$15. If they do hire, they play the gun hired (GH) game with sub-game equilibrium earnings of \$40. A subject should be willing to pay any amount less than or equal to the gain from hiring the gun (\$25) to hire the gun. We have set the total group cost of hiring the gun to only \$2 per group. Any combination of contributions summing to exactly \$2 will be an equilibrium of the stage 1 game (Bagnoli and Lipman, 1989; Bagnoli and McKee, 1991; Marks and Croson, 1998).

Any two players could pay for the punishment mechanism, so one could interpret the implementation of our mechanism as requiring 50% of the group to agree on implementation. The average cost of the gun per person should be \$0.50, and with the gun hired second stage earnings which should be \$40, the resulting average earnings should be \$40.50 per subject in the G4H game.

3.4. The peer-to-peer (P2P) game

Our peer-to-peer punishment game is similar to that of previous experiments (see Fehr and Gachter, 2002; Cinyabuguma et al., 2006; Herrmann et al., 2008; Gachter et al., 2008). Subjects first play the LPG game with an automatic payment of \$1 (again to reduce income effects), then are given anonymous information about the contribution to the public good, private good, and about initial LPG earnings for each of their group members. At this point, each player i can pay \$1 to assign a punishment point to another player j, which we write as p_{ij}. Each point assigned reduces player j's payoff by \$3.[3] Final payoff are given by the following expression:

$$\pi_i = 1 + 3(5 - g_i) + 2\sum_{j=1}^{4} g_j - \sum_{j \neq i} p_{ij} - 3\sum_{k \neq i} p_{ki}.$$

Given that groups are randomly and anonymously rematched each period, own-profit maximizing subjects should choose to assign zero punishment points to all players ($p_{ij} = 0$), and the game should be the same as the LPG game. The predicted outcome under own-profit maximizing behavior is $g_i = 0$ for all subjects and final earnings per subject of \$16.

It is important to note that the own-profit maximizing equilibria predictions of the P2P and LPG games are the same, but that many previous works have found that subjects behave very differently in these two games. The fact that players engage in punishment at all is surprising, not only because it is not the equilibrium action, but more so because we do not observe much peer-to-peer punishment in many real world situations as found by Balafoutas and Nikiforakis (2011). One reason we observe such high amounts of peer punishment in the lab may be that players were never offered another alternative, such as hiring a delegated punishing mechanism in addition to peer punishments. Our final game allows the use of both a delegated punishment mechanism and peer-to-peer punishments.

3.5. The gun for hire and peer-to-peer (G4H/P2P) game

The last game combines the G4H and P2P games. In stage 1, subjects are given 4 tokens, and they make contributions toward a hiring fund. If the sum of those contributions is greater than 8 tokens, then a gun is hired and subjects get \$0.25 for each token they kept. If the gun is not hired, stage 1 earnings are \$1. Subjects are informed of their stage 1 earnings, group contributions to the hiring fund, and whether the gun has been hired. In stage 2, subjects get 5 tokens to contribute to either a public or private good. If the gun was hired, then the lowest contributor(s) to the public good will be punished by the delegated punishment mechanism. In stage 3, subjects are given anonymous details of group members' contributions to the public good, the private good, and their initial earnings (earnings before punishments from the hired gun mechanism). They also learn the size of punishment from the mechanism (if any) and the net earnings for each subject in

[3] The punishment to cost ratio of 3:1 has been employed by many of the previous experiments (e.g. Fehr and Gachter, 2002; Gachter et al., 2008; Herrmann et al., 2008), while some others have employed a 4:1 ratio (Cinyabuguma et al., 2006). For a discussion of the constant ratio versus other punishment regimes see Casari (2005). Previous work has found that a cost to punishment ratio of no lower than 1:3 is necessary to raise public contributions and welfare (Nikiforakis and Normann, 2008; Egas and Riedl, 2008). There is the possibility of earning a negative payoff in the P2P game. Subjects were warned about the possibility of negative payoffs in the instructions and were told that they would never owe money at the end of the experiment; and that at minimum they would be paid \$7. In only 3 cases did a subject earn a negative amount in a period.

their group. At this point, subjects can choose to assign peer-to-peer punishments to their group members. Again, subject *i* chooses an amount of punishment points to assign to player *j*. Each point player *i* assigns costs player *i* $1, and reduces the payoff of player *j* by $3.

Again, own-profit maximizing subjects would assign zero punishment points, leading to predictions identical to the G4H game: subjects hire the gun in the first stage, and fully contribute in the second stage. Average per person earnings would be $40.50 per person. Table 1 summarizes the equilibrium predictions for each of these games. We see that theoretically the gun for hire mechanism whether chosen or imposed (GH, G4H, G4H/P2P) should result in better provision of the public good, and higher average earnings than the linear public goods game (LPG) or the peer-to-peer (P2P) game.

Table 1
Equilibrium predictions.

Game	Public contribution	Punishment points	Total net earnings[a]
LPG: linear public goods	0	na	$16.00
GH: gun hired	5	na	$40.50
G4H: gun for hire	5	na	$40.50
P2P: peer-to-peer	0	0	$16.00
G4H/P2P: gun for hire and peer-to-peer	5	0	$40.50

[a] Total net earnings are earnings minus costs of punishment and plus automatic payments.

4. Procedures

There are two equally valid views of what is the "natural" baseline. The first is that the LPG game is the baseline and the P2P is an intervention. The second takes vigilante justice as an ever present option, and so the baseline should be a game with peer-to-peer (P2P) punishments available. We conduct two sets of experiments using both the LPG, and the P2P games as baselines.

Each session involved 12 subjects and 20 periods: 10 periods of a baseline game (either LPG or P2P) followed by 10 periods of a game with punishment (either GH, G4H, P2P, or P2P/G4H). Each treatment is a set of two games, and there are a total of 5 treatments: (1) LPG–GH, (2) LPG–G4H, (3) LPG–P2P, (4) P2P–P2P, and (5) P2P–G4H/P2P.[4] Each treatment was conducted at least 3 times for a total of 36 subjects per treatment. The LPG–GH treatment was conducted 4 times for a total of 48 subjects. We have a total of 192 subjects. Each session was conducted using z-tree software (Fischbacher, 2007), and lasted under 90 minutes. Subjects earned $28 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.[5] Subjects were given the instructions for the first 10 periods of play, a quiz, and then played that game for 10 periods. All participants had to correctly answer the quiz questions before moving on. This is done again for the last 10 periods. To remove experimenter effects, all sessions were run by the same person. Subjects could earn up to $46 in each period, so they were informed that they would be paid for a single randomly selected period from the 20 periods in the session.[6]

The instructions were written in neutral language by referring to the public good as the "BLUE investment", the private good as the "RED investment", the delegated punishment mechanism as "the computer simulated administrator", and referring to all punishments as "deductions." Full instructions and screen shots are available from the authors in the online Appendix.[7]

5. Results

The "natural" baseline for our experiments is either a world without any punishment options, the LPG game, or a world with only vigilante justice, the P2P game. We will begin by exploring the results for a world without any punishments available in the first 10 periods.

5.1. Baseline: LPG

When we begin in a world with no punishment options we are first interested in testing if our gun for hire mechanism will work when it is exogenously imposed in the GH game. After we have shown that the mechanism works we will show that it is still effective when it is endogenously hired in the G4H game. Lastly we will show that when we compare the endogenously chosen third party mechanism (G4H) to the endogenously chosen vigilante justice (P2P) we find that welfare is greatly improved under the third party mechanism.

5.1.1. Exogenously imposed mechanism: GH

We first use a within subjects comparison to test whether our mechanism can fix the free riding that has built up in the first 10 periods of the LPG game. Looking at Fig. 1 and Table 2 we can see that in Periods 1 to 10 subjects contribute an average of 1.56 tokens per period and that there is a trend toward more free-riding as they repeat the game. When subjects have the gun for hire imposed on them in the last 10 periods of the LPG–GH treatment there is an immediate jump in contributions and the average contribution to the public good rises to 4.57 tokens per period. Earnings increase from $23.82 per period in the LPG game to $37.28 per period in the GH game. This jump is immediate, as can be seen in Fig. 2. Clearly the mechanism

[4] One may be curious why we did not run all permutations of combinations of these games. The reason is we were primarily interested in starting with a baseline world (either LPG or P2P) and then adding on an additional punishment option. So for example if we had run LPG–G4H/P2P then we would have started in a baseline world with no punishment and then added two punishment options. We find this is an interesting manipulation but it was not the focus of this paper.

[5] The use of a random strangers matching protocol should minimize the effect of reputations and contagion because subjects do not know who they are playing with nor if they have played with them before or will play with them in the future. However it is possible that a player may play against the same subject or even within the same group multiple times, and it is also possible for a player to be affected through contagion. We believe these reputation and contagion effects did not have a significant effect on our results, and as a robustness check we have asked whether our results hold when looking at only the first period of play following our baseline games (e.g. Period 11 of the (1) LPG–GH, (2) LPG–G4H, (3) LPG–P2P, (4) P2P–P2P, and (5) P2P–G4H/P2P). The results are of the same sign and generally remain statistically significant in all cases.

[6] 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, and then to roll and announce the outcome on the die out loud.

[7] We included a number of examples in the text of the instructions and in the tests of understanding that we made each subject pass before moving on to actual game play. As pointed out by one of our very helpful reviewers we explicitly mention an example of full contribution to the public good in the GH, G4H and G4H/P2P instructions while we neglect to use this same example in the LPG and P2P instruction sets. This was an unfortunate and unintended oversight, however we do not believe it had any impact on our results. In particular, we have data from another experiment on the same subject pool where subjects played a linear public goods game and the instructions explicitly mention the full contribution example. The text of the instructions read "Example 3: Imagine you invested your 5 tokens this way: 0 in the RED and 5 to the BLUE investment. Also imagine the other group members invest 5, 5, and 5 to the BLUE investment." We found that average public contribution in these LPG games was 1.5 tokens for the first 10 periods for the 80 subjects who had this written in their instructions. In contrast our subjects who had LPG instructions without this example gave an average of 1.7 tokens for the first 10 periods. The difference between these two means is not statistically significant at standard significance levels (if anything it appears the example pushed contributions down). This leads us to believe that the lack of inclusion of this example in the instructions had no effect. Nonetheless, we regret not having been more consistent in our choice of examples in the instructions.

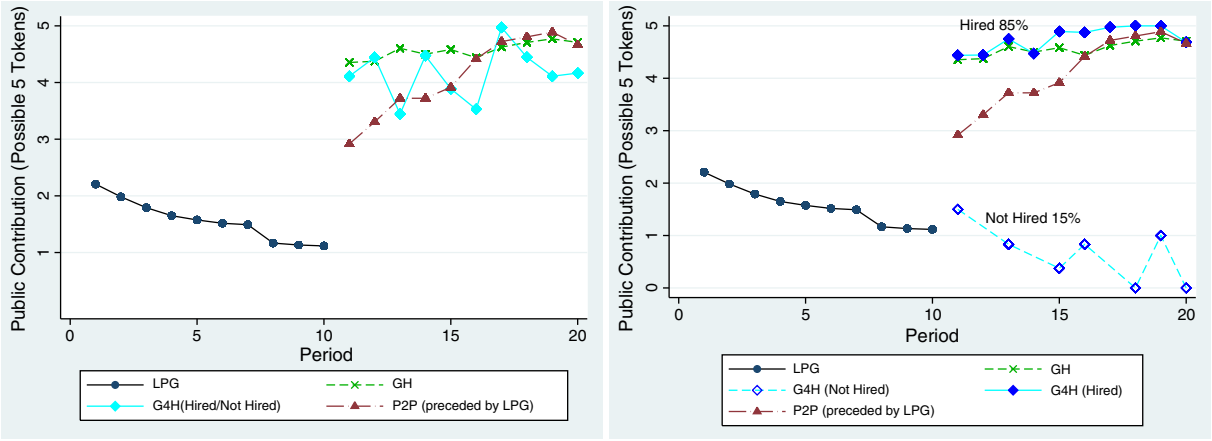

Fig. 1. Contributions to the public good after LPG. Notes: In this figure we show the average per person contribution to the public good by treatment out of a possible 5 tokens for treatments which began with 10 periods of the linear public goods (LPG) game, a game with no punishment mechanism. In the left hand panel the "gun for hire" (G4H) treatment is the average over both when the hired gun mechanism was and was not implemented. In the right hand panel the "gun for hire" (G4H) treatment is divided into when the hired gun mechanism was implemented (solid line) and was not implemented (dashed line).

Table 2
Average earnings per subject with baseline LPG.

Game (periods)	Net earnings (dollars)	Public contribution (5 tokens)	P2P costs (all)	G4H costs (all)
LPG (1–10)	23.82	1.56	na	na
GH (11–20)	37.28	4.57	na	1.06
G4H (11–20)	35.44	4.16	na	1.35
Hired (85%)	38.12	4.74	na	1.57
Not hired (15%)	19.55	0.71	na	na
P2P (11–20)	30.69	4.11	5.86	na

Note: 10 periods of each game per session, 3 groups per session, 4 subjects per group.

has had the desired effect of reducing free-riding and increasing earnings even after punishments have been taken away.

Next we can use between subject comparisons to see if our mechanism performs well both when it is imposed (GH) and when subjects have to pay a cost to implement it (G4H). Although the equilibrium of the G4H game is to implement the gun for hire, subjects may not always immediately realize this fact. For subjects to hire the gun, they must believe that the cost of implementing the delegated punishment mechanism will be outweighed by the gains from reduced free-riding. Subjects appear to believe this — they hire the mechanism 85% of the time in the last 10 periods of LPG–G4H. In fact our gun for hire is over-paid for. There are multiple equilibria for the hiring stage, such that any combination of contribution to the hiring fund that total exactly $2 (8 tokens) is a Nash equilibrium. Yet, we only observe the groups paying exactly $2 a mere 10% of the time in the G4H game, 90% of the time the gun for hire is over-paid for.

We can show that the mechanism improves public contributions both when it is imposed and when it is hired. In the left hand panel of Fig. 1 we can compare average contributions to the public good in the G4H game (note this includes both when the gun has been hired and when it has not) to the GH game. The contributions are slightly higher, but insignificantly, in the GH game.[8] In the right panel of Fig. 1 we can see public contributions in the G4H game divided into when the gun was hired versus when it was not hired. Here we see that when subjects successfully hire the gun they actually contribute more on average than when it was imposed on them, but this difference is not statistically significant.[9] Clearly, when the gun is not successfully hired, which is only 15% of the time, the subjects contribute much less.[10]

Next we ask whether the gun for hire mechanism also improves net earnings. Net earnings are earnings after the costs of hiring and punishment have been deducted. We can see in Fig. 2 and Table 2 that earnings are $37.28 on average in the GH treatment while they are $35.44 in the G4H treatment. These differences in earnings are not statistically significantly different, so it appears that the opportunity for the mechanism alone raises earnings.[11] If we divide the G4H game into when the gun was hired or not hired, we find that when subjects hire the gun their earnings are around $38.12 per person, while if they do not hire, they earn an average of $19.55.

Result 1. *Subjects are willing to pay a cost to submit to a delegated punishment mechanism. In the G4H game the delegated punishment mechanism is implemented 85% of the time, and groups over-pay for this implementation in most cases. Welfare, as measured by average individual net earnings, is similar when the gun has been endogenously hired in the G4H game versus when it has been exogenously imposed in the GH game.*

5.1.2. Endogenously chosen punishments: G4H versus P2P

We have shown that the gun for hire works when it has been either exogenously assigned or when it has been endogenously paid for. We next compare the effectiveness of G4H to P2P. Looking at Table 2, we see that average per-person earnings in Periods 11 to 20 are higher in the LPG–G4H treatment ($35.44 overall: $38.12 when gun is hired and $19.55 when not hired), than in the LPG–P2P treatment ($30.69 in the last 10 periods). Table 3 provides an overview of how average earnings are shaped in each treatment of this experiment. The variable G4H takes the value 1 when subjects are playing the G4H game and zero when they are playing the P2P game after

[8] The difference is not significant using a Kolmogorov–Smirnov test at the session level when comparing public contributions in the GH to all the public contributions in the G4H game (all meaning both when the gun is hired and not hired), $p=0.237$. We use a Kolmogorov–Smirnov test because we only have 4 observations at the session level for the GH and 3 at the session level for the G4H games. Our results differ from those of Sutter et al. (2010) who found higher contributions for endogenously chosen mechanisms.

[9] Kolmogorov–Smirnov test at the session level when the gun has been hired $p=0.265$.

[10] Kolmogorov–Smirnov test at the session level when the gun has *not* been hired in the G4H versus the GH game, $p=0.047$.

[11] For earnings $p=0.237$ using a Kolmogorov–Smirnov test at the session level. We use a Kolmogorov–Smirnov test because we only have 3 observations at the session level for the P2P and G4H games.

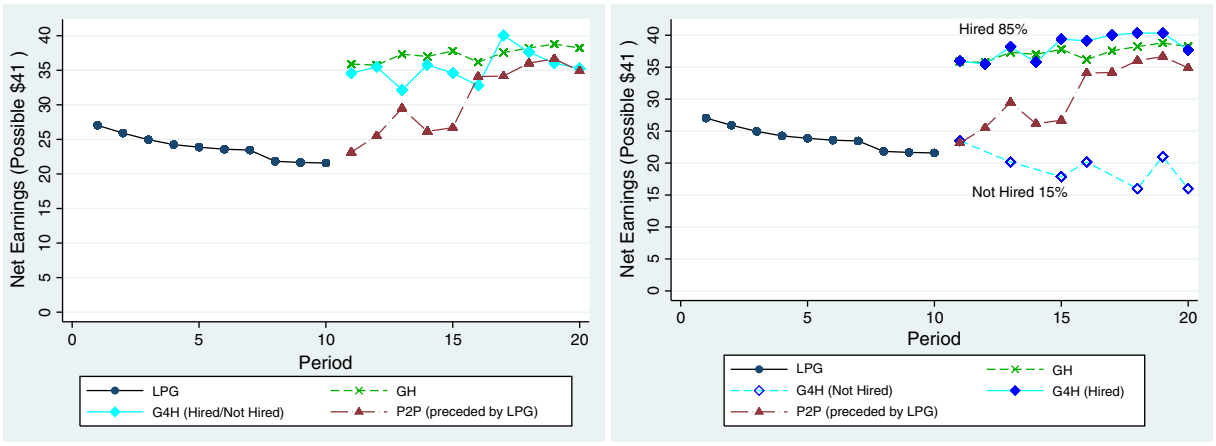

Fig. 2. Average per subject net earnings after LPG. Notes: In this figure we show the average per person earnings after punishment deductions by treatment out of a possible $41 for treatments which began with 10 periods of the linear public goods (LPG) game, a game with no punishment mechanism. In the left hand panel the "gun for hire" (G4H) treatment is the average over both when the hired gun mechanism was and was not implemented. In the right hand panel the "gun for hire" (G4H) treatment is divided into when the hired gun mechanism was implemented (solid line) and was not implemented (dashed line).

Table 3
Determinants of earnings in rounds 11–20 by treatment.

	After LPG	After P2P
G4H	4.76***	
	(1.14)	
G4H/P2P		8.98**
		(3.07)
Period	0.90***	0.49***
	(0.10)	(0.09)
Constant	16.77***	15.25***
	(1.76)	(2.59)
N	720	720
Wald Chi-squared	97.14***	38.03***

Notes: G4H = 1 if subject in the G4H condition in rounds 11–20, and G4H = 0 if subject in the P2P condition in rounds 11–20 after playing LPG in 1–10. Similarly, G4H/P2P = 1 if subject in the G4H/P2P condition in rounds 11–20, and G4H/P2P = 0 if subject in the P2P condition in rounds 11–20. Linear random effects models. Clustered standard errors in parentheses. Standard errors clustered by session. ***$p<.01$, **$p<.05$, *$p<.10$ significance.

Periods 1–10 of LPG. Playing the G4H game instead of the P2P game raises earnings by $4.76 per period on average including when the gun was not hired.[12]

There are two possible reasons for the increased average earnings: increased average giving and decreased average punishment costs. Table 2 shows that average giving was nearly identical in P2P (4.11) and G4H (4.16). However, as Fig. 1 shows, this average masks a great deal of heterogeneity across treatments. When the delegated punishment mechanism is hired, average giving is higher in G4H.[13] As a result, punishment in G4H is small ($1.35 per subject), especially in comparison to P2P ($5.86 per subject). Thus, lower punishments are primarily responsible for the increased efficiency.

The gain in earnings between the two treatments is illustrated graphically in Fig. 2. In the left panel are earnings both when the gun is hired and when it is not hired averaged together. These earnings are almost always higher than earnings with P2P punishments. In the right panel, we see that when the gun is hired, average per subject earnings are always higher than those under P2P punishments.

Result 2. *Welfare, as measured by average individual net earnings, is higher in the G4H treatment than the P2P treatment. When the mechanism is hired, the use of the delegated punishing mechanism both improves public contributions, and lowers costs as compared to allowing peer-to-peer punishments.*

One can see in both Figs. 1 and 2 that the advantage of G4H over P2P diminishes with time, that is earnings and public contributions in the G4H treatment and P2P treatment appear to converge in the last 5 periods of play. In the next section we will see if this is also the case when we start off in a baseline world with P2P punishments.

5.2. Starting from vigilante justice: welfare in Periods 11–20 following P2P in Periods 1–10

It has been argued that peer-to-peer punishment "plays an important role in real life" (Fehr and Gachter, 2000). If such peer-to-peer punishment is indeed natural and often occurring then we should use the P2P game as our baseline rather than the setting without any punishment opportunities. One may wonder if subjects will still hire the third party mechanism when they know that they can use vigilante justice. We found that subjects hired the gun in the G4H/P2P game 72% of the time. Similar to the LPG baseline, 80% of the time the gun is over-paid for in the G4H/P2P game.

Table 4 shows that average per person earnings in Periods 11 to 20 are higher in the G4H/P2P treatment ($31.87 overall: $36.14 when gun is hired and $20.77 when not hired), than they are in the P2P treatment ($22.89).[14] Average net earnings are 40% higher in Periods

[12] The same patterns of significance can be shown in the Kolmogorov–Smirnov test at the session level. For our regression to properly identify effects we must make two assumptions. We must assume that the session is a random variable, which it should be given random assignment of treatments to sessions. We can also show that it passes a Breusch and Pagan Lagrangian multiplier test for random effects. Second, we must also assume no correlation between the session and the observable right hand side variables which in our case are the period, and the treatment dummies. Again this assumption should be met given random assignment of treatment to sessions. We conduct our analysis at the session level (6 sessions per regression: 3 G4H game and 3 P2P game in Regression 1; 3 G4H/P2P game and 3 P2P game in Regression 2) because we have used a strangers matching protocol which means that individuals can play with the same group members multiple times during a session. If we cluster at the session level we are assuming that actions are independent across sessions, which seems a safe assumption given sessions never have the same persons and treatments are assigned randomly. Estimating equations and results with standard errors clustered at the individual level have smaller standard errors and are available in the online appendix.

[13] Comparing session level public contributions in the P2P to when the gun is hired in G4H the public contributions in G4H conditional on hiring are statistically significantly higher than those in P2P using a Kolmogorov–Smirnov test $p = 0.09$.

[14] One might notice that earnings in the last 10 periods of P2P–P2P are $22.89 which is much lower than earnings in the last 10 periods of LPG–P2P which were ($30.69). This is a surprising difference, especially given the theoretical prediction is that in both games the average earnings should be $16. We would like to thank an anonymous referee for pointing this out, and we see this as a fruitful question for future research.

Table 4
Average earnings per subject after P2P in Periods 1–10.

Game (periods)	Net earnings (dollars)	Public contribution (5 tokens)	P2P costs (all)	G4H costs (all)	Total costs
P2P (1–10)	22.36	2.09	4.07	na	4.07
P2P (11–20)	22.89	2.33	4.74	na	4.74
G4H/P2P (11–20): all	31.87	3.67	1.03	1.43	2.45
Hired (72%)	36.14	4.55	0.63	1.97	2.60
Not hired (28%)	20.77	1.37	2.08	na	2.08

Note: 10 periods of each game per session, 3 sessions, 3 groups, 4 subjects per group.

11–20 in the P2P–G4H/P2P treatment than they are in the P2P–P2P treatment, a significant increase.[15]

Additionally, in the regression reported in the second column of Table 3 one can see that the coefficient on the treatment dummy variable G4H/P2P is positive and significant. Playing the G4H/P2P game instead of the P2P game raises earnings by $8.98 per period on average.[16] Fig. 3 shows that contributions to the public good in the last 10 periods of the P2P–G4H/P2P treatment seem to rise over time, while the public contributions stay relatively flat in the P2P–P2P treatment. In the left panel of Fig. 3 we see that averaging over when the gun is hired and not hired subjects have higher per person contributions in the G4H/P2P treatment than in the P2P treatment in every period. This result is even more clear when one looks to the right hand panel of Fig. 3 where contributions have been decomposed into when the gun was hired and when it was not hired. Additionally Fig. 4 shows that earnings trend upwards for the G4H/P2P treatment, while they stay relatively flat in the P2P treatment. Result 3 summarizes.

Result 3. *Subjects are willing to pay a cost to submit to a delegated punishment mechanism even when they know they will have the ability to peer-to-peer punish. In the G4H/P2P game the delegated punishment mechanism is implemented 72% of the time, and groups over-pay for this implementation in most cases. Welfare, as measured by average individual net earnings, is significantly higher in the last 10 periods of the P2P–G4H/P2P treatment than the P2P–P2P treatment.*

5.3. Does delegated enforcement crowd out peer punishment?

We have shown that our G4H mechanism is both implementable and welfare improving when compared to a P2P punishment regime. Next, we show that hiring a gun crowds out the use of peer punishments. If delegated punishment crowds out peer-to-peer punishment, this may be welfare improving. Also, if delegated punishment crowds out peer-to-peer punishment, this will in turn lower any possible motives for peer-to-peer revenge punishments.[17]

Looking back at the Table 4, we can compare the behavior of subjects who could only peer punish in Periods 11–20 (P2P) to those who were also allowed to hire a gun and peer punish (G4H/P2P). When we make this comparison we see that peer punishment costs fall from an average of $4.74 in the P2P game to $1.03 in the G4H/P2P game ($0.63 when the gun is hired, and $2.08 when the gun is not hired). The average use of peer punishment when it is the only option is over four times higher than when peer punishment is available alongside the option for a hired gun ($1.03 versus $4.74). Like previous work, it is interesting that subjects choose to punish at all because this is not the Nash equilibrium of the one shot game. Recall the hypothesis suggested earlier that subjects enjoy punishing, even when it is costly to them, because they experience utility from enforcing justice. If presented with the option to implement a "just" mechanism, people may prefer this mechanism to vigilante justice. Our subjects may believe that the gun for hire is such a "just" mechanism, and so refrain from punishment when it is offered as an option. Unfortunately this paper is unable to test if subjects have this taste for justice, but we believe it is an important area for further research.

Some may wonder if the gun for hire mechanism is simply a less expensive way to punish than the P2P. In the P2P game, it always costs $0.33 to punish another player $1. In the G4H and G4H/P2P games the ratio varies since the punishment depends on size of deviation, and can range as high as $18. Although it may appear that paying only $0.50 per person ($2 in total) to punish $18 is simply a great value, in our experiments the punishment was usually well below this $18 size. On average it cost $0.74 to punish $1 in the G4H game, and $0.51 to punish $1 in the G4H/P2P game, and so punishment was actually cheaper under the P2P mechanism.

Comparing the use of peer punishment by hiring decision in the G4H/P2P treatment alone, we see the average costs of peer punishment fall about 70% when the gun is hired. Fig. 5 shows the average costs per subject of peer punishments in the P2P game, plus the costs by hiring decision in the G4H/P2P game. In both the P2P and G4H/P2P games, the use of peer punishment is trending downward over time. In the P2P game the costs of punishment are always higher than in the G4H/P2P game, whether a gun has been hired or not. One would expect the use of peer punishments to fall when a gun has been hired, but it is especially surprising that the use of peer punishments is lower even when the gun is not successfully hired. Although there is selection into the hiring of the gun, even when averaging over all groups (those that hired and that did not hire), it appears that merely giving the option for the gun for hire, even when that option is not exercised, decreases peer-to-peer punishments. Furthermore, it is noteworthy that in 4 of the 10 periods (after 10 periods of P2P) when the gun is hired peer punishment costs are equal to zero.

Result 4. *Use of peer punishments is over four times higher when the option for a delegated mechanism is not available. When the delegated mechanism is implemented, peer punishment converged to zero by Period 19. Delegated punishment crowds out peer-to-peer punishment, resulting in an overall welfare gain.*

6. Concluding remarks

Much of the previous work on punishment in public goods games has concentrated on asking whether groups can govern themselves through the use of peer-to-peer punishments. This line of inquiry does not allow individuals to collectively agree to concentrate the punishment in a recognized authority. In this paper, we show that subjects willingly pay to delegate punishments in a linear public goods game. We offer a stylized version of delegated punishment in our gun for hire mechanism. The mechanism has the properties that only the largest free rider is punished, the size of the punishment is related to the degree of defection from the other group members' behavior, in equilibrium the mechanism is efficient in the sub-game, and the mechanism is relatively low cost.

[15] These are statistically significantly different from each other using a Kolmogorov–Smirnov test ($p=0.090$) at the session level.

[16] The variable G4H/P2P takes the value 1 when subjects are playing the G4H/P2P game, and zero when they are playing the P2P game after Periods 1–10 of P2P. The same patterns of significance can be shown in the Kolmogorov–Smirnov test at the session level.

[17] We expect that the use of a delegated punishing mechanism would preclude revenge motives in subsequent rounds of punishment, although we have not allowed for multiple rounds of punishment in our current design. When the mechanism is levying fines, it is not possible for an individual to know who to take revenge on. Imagine if your neighbor was leaving garbage in the common areas of your building. You can either speak with your neighbor directly, or ask the superintendent to speak to your neighbor without mentioning your name. If you speak with your neighbor directly they may take offense, and they may "counter-punish" you by stealing your newspaper. On the other hand if your superintendent speaks with your neighbor, there is no way for your neighbor to know that you commissioned the punishment. As a helpful anonymous reviewer pointed out subjects could use peer punishment to exact revenge on group members who helped to pay for the gun for hire mechanism during that period. Although we do not reveal who paid for the mechanism, subjects could punish those they suspected of paying to hire. To test for this in our data we looked at subjects who were punished by the hired gun and who also assigned punishment points. There are only 4 instances of this in all our data, so using the peer punishment for revenge on those who helped to hire the mechanism does not appear to be a widely used strategy.

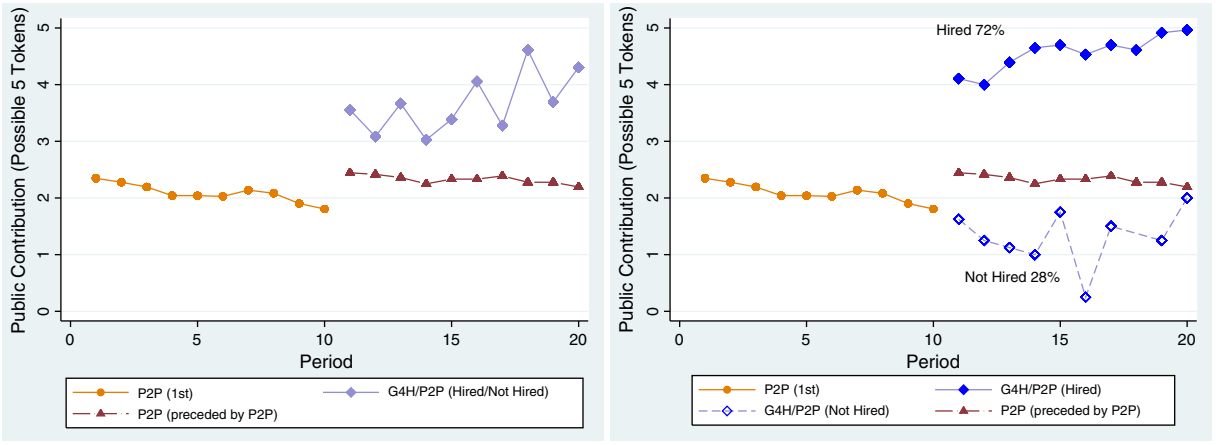

Fig. 3. Contributions to the public good after P2P. Notes: In this figure we show the average per person contribution to the public good by treatment out of a possible 5 tokens for treatments which began with 10 periods of the peer-to-peer (P2P) game, a game with a punishment mechanism. In the left hand panel the "gun for hire/peer-to-peer" (G4H/P2P) treatment is the average over both when the hired gun mechanism was and was not implemented. In the right hand panel the "gun for hire/peer-to-peer" (G4H/P2P) treatment is divided into when the hired gun mechanism was implemented (solid line) and was not implemented (dashed line).

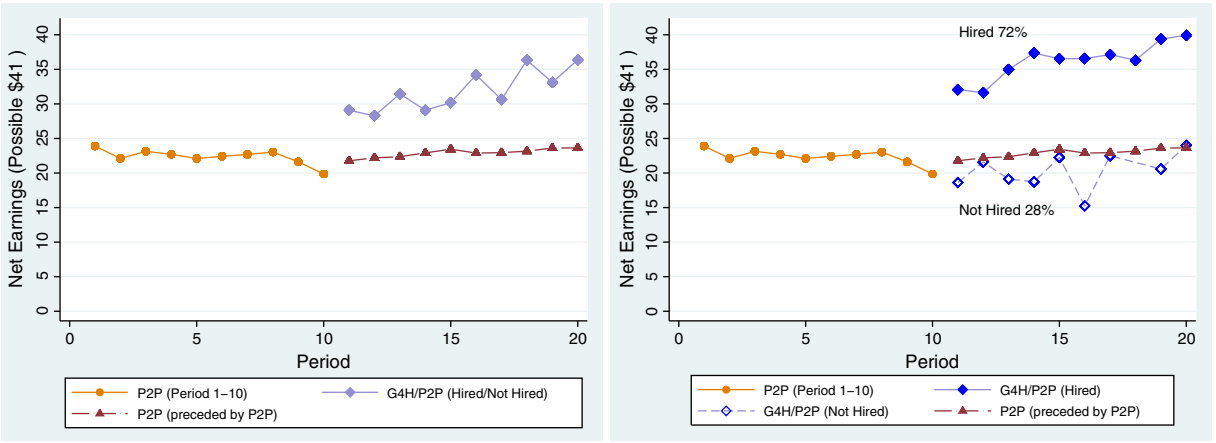

Fig. 4. Average per subject net earnings after P2P. Notes: In this figure we show the average per person earnings after punishment deductions by treatment out of a possible $41 for treatments which began with 10 periods of the peer-to-peer (P2P) game, a game with a punishment mechanism. In the left hand panel the "gun for hire/peer-to-peer" (G4H/P2P) treatment is the average over both when the hired gun mechanism was and was not implemented. In the right hand panel the "gun for hire/peer-to-peer" (G4H/P2P) treatment is divided into when the hired gun mechanism was implemented (solid line) and was not implemented (dashed line).

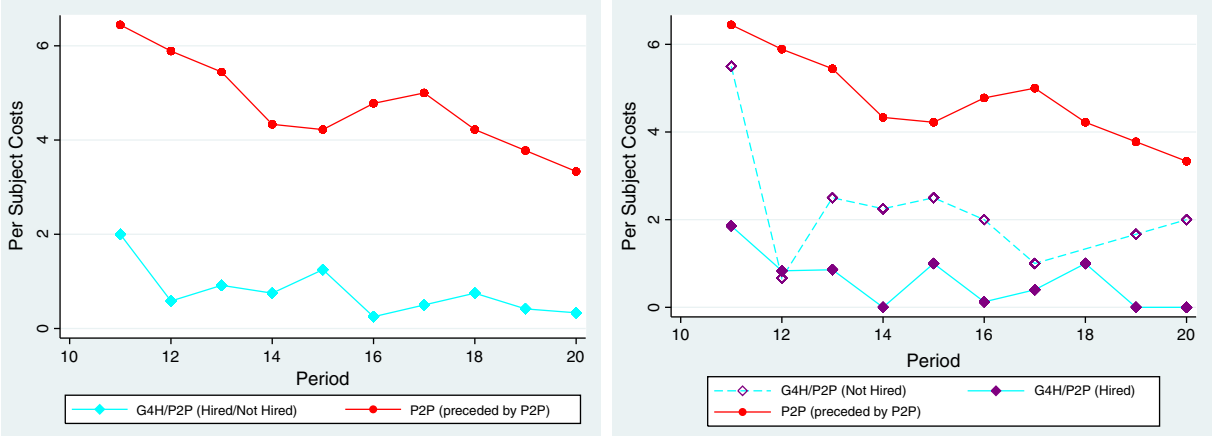

Fig. 5. Average per subject peer punishment costs after P2P in Periods 1–10. Notes: In this figure we show the average per person punishment costs by treatment for treatments which began with 10 periods of the peer-to-peer (P2P) game, a game with a punishment mechanism. In the left hand panel the "gun for hire/peer-to-peer" (G4H/P2P) treatment is the average over both when the hired gun mechanism was and was not implemented. In the right hand panel the "gun for hire/peer-to-peer" (G4H/P2P) treatment is divided into when the hired gun mechanism was implemented (solid line) and was not implemented (dashed line). The reason there is no data point for Period 18 on the non-implemented (dashed) line in the right panel is that the gun was hired by all groups during Period 18 in all sessions.

When given the opportunity to hire a delegated punishment mechanism, we see the mechanism being implemented over 70% of the time in both the linear public good and peer-to-peer baseline worlds. The likely reason that subjects are so willing to submit to a costly outside authority is that they expect monetary gains from reduced free-riding. These expectations are well-founded, as can be seen by the 15% and 40% increase in welfare when comparing a peer-to-peer punishment regime to those with the option of a gun for hire regime. When subjects can only use peer punishments (P2P) the peer punishment costs are over four times those with a delegated mechanism (G4H/P2P). Last and most important, we find that when both punishment methods are available (G4H/P2P), subjects lower their use of peer punishments by 70%. The existence of a delegated punishing mechanism crowds out the use of peer punishments.

To our knowledge, this paper is the first to allow subjects to choose between hiring a costly punishment mechanism and using peer-to-peer punishments. We have shown that players want to hire the delegated mechanism and that the gun for hire mechanism provides a low cost solution to the problem of free-riding. Interestingly, although the delegated mechanism is itself a public good, it does not appear to suffer from the same level of free-riding as observed in the subsequent LPG game. The reason may be that the cost of hiring is fairly low as compared to the potential gains in payoffs. This is analogous to the way we pay taxes or fees to fund delegated punishing mechanisms in general. Often these fees and penalties are small, as in our mechanism. Further research exploring how players react to changes in the cost of implementing the mechanism could be illuminating. Although formally our gun for hire was an external third party, it is clearly an important and desirable next step for research to investigate a more general set of ways individuals can delegate authority. For instance, the recognized authority can be internal to the group, and the enforcers's conformity with enforcement rules a choice variable. This would be most interesting, of course, in the default domain of peer-to-peer punishment. In our experiment the third party always executes punishments exactly as dictated by the mechanism, but if the gun for hire was an actual person one may worry about abuse of power, making it critical to keep the power of the authority relatively weak, as cogently pointed out by Binmore (2005). Work exploring how a human third party authority might abuse power is also an important further research question. The ultimate research goal suggested by our study is to understand how easily small self-governed groups can innovate ways to avoid the inefficiencies of peer-to-peer punishment.

This paper illustrates that under reasonable conditions individuals prefer to pay to be governed by a delegated punishment mechanism rather than use peer-to-peer punishments. The gun for hire mechanism is just one example of a low-cost device that can deter free-riding behavior in a public goods game, improve welfare, and crowd out the use of deleterious peer-to-peer punishments.

In short, when Paladin comes to town, vigilante justice is driven out. Have gun. Will travel.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jpubeco.2012.08.003.

References

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

Bagnoli, M., McKee, M., 1991. Voluntary contribution games: efficient private provision of public goods. Economic Inquiry 29.

Balafoutas, Loukas, Nikiforakis, Nikos, 2011. Norm enforcement in the city: a natural field experiment. Working Paper.

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

Binmore, Ken, 2005. Natural Justice. Oxford University Press.

Bochet, Olivier, Putterman, Louis, 2007. Not just babble: opening the black box of communication in a voluntary contribution experiment. Working Paper.

Botelho, A., Harrison, G.W., Pinto, L.M.C., Rutstrom, E.E., 2007. Social norms and social choice. Working Paper.

Boyd, Robert, Gintis, Herbet, Bowles, Samuel, 2010. Coordinated punishment of defectors sustains cooperation and can proliferate when rare. Science. 328 (5978), 617–620. http://dx.doi.org/10.1126/science.1183665 (30 April).

Casari, Marco, 2005. On the design of peer punishment experiments. Experimental Economics 8 (2), 107–115.

Casari, Marco, Luini, Luigi, 2009. Cooperation under alternative punishment institutions: an experiment. Journal of Economic Behavior and Organization 71 (2), 273–282.

Cinyabuguma, Matthias, Page, Talbot, Putterman, Louis, 2006. Can second-order punishment deter perverse punishment. Experimental Economics 9 (3).

Decker, Torsten, Stiehler, Andreas, Strobel, Martin, 2003. A comparison of punishment rules in repeated public good games: an experimental study. Journal of Conflict Resolution 47 (6), 751–772.

Denant-Boemont, Laurent, Masclet, David, Noussair, Charles N., 2007. Punishment, counterpunishment, and sanction enforcement in a social dilemma experiment. Economic Theory 33 (1).

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

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

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

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

Fehr, Ernst, Gachter, Simon, 2000. Cooperation and punishment in public goods experiments. The American Economic Review 90 (4), 980–994.

Fehr, Ernst, Gachter, Simon, 2002. Altruistic punishment in humans. Nature.

Fischbacher, Urs, 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10 (2).

Gachter, Simon, Renner, Elke, Sefton, Martin, 2008. The long-run benefits of punishment Science 322 (5907), 1510. http://dx.doi.org/10.1126/science.1164744 (5 December).

Gachter, Simon, Renner, Elke, Sefton, Martin, 2008. The long-run benefits of punishment. Science 322 (5907), 1510.

Grechenig, Kristoffel, Nicklisch, Andreas, Thoni, Christian, 2010. Punishment despite reasonable doubt: a public goods experiment with sanctions under uncertainty. Journal of Empirical Legal Studies 7.

Guillen, Pablo, Schwieren, Christiane, Staffiero, Gianandrea, 2007. Why feed the Leviathan? Public Choice 130, 115–128. http://dx.doi.org/10.1007/s11127-006-9075-3.

Gurerk, Ozgur, Irlenbusch, Bernd, Rockenbach, Bettina, 2006. The competitive advantage of sanctioning institutions. Science 312 (5770), 108–111.

Herrmann, Benedikt, Thoni, Christian, Gachter, Simon, 2008. Antisocial punishment across societies. Science 319 (5868), 1362–1367.

Hirschman, Elizabeth, 2000. Heroes, Monsters and Messiahs: Movies and Television Shows as the Mythology of American Culture. Andrews McMeel Publishing.

Ho, Teck-Hua, Camerer, Colin, Weigelt, Keith, 1998. Iterated dominance and iterated best response in experimental "p-beauty contests". The American Economic Review 88 (4), 947–969.

Kosfeld, Michael, Akira Okada, Arno Riedl, 2009. Institution formation in public goods games. American Economic Review 99 (4), 1335–1355. http://dx.doi.org/10.1257/aer.99.4.1335.

Kube, Sebastian, Traxler, Christian, 2011. The interaction of legal and social norm enforcement. Journal of Public Economic Theory 13 (5), 639–660.

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

Masclet, David, Noussair, Charles, Tucker, Steven, Villeval, Marie-Claire, Villeval, Marie-Claire, 2003. Monetary and nonmonetary punishment in the voluntary contributions mechanism. The American Economic Review 93 (1), 366–380.

Masclet, David, Noussair, Charles N., Villeval, Marie Claire, 2011. Threat and punishment in public good experiments. Working Paper.

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

Nikiforakis, Nikos, 2008. Punishment and counter-punishment in public good games: can we really govern ourselves? Journal of Public Economics 92 (1–2), 91–112.

Nikiforakis, N., Engelmann, D., 2011. Altruistic punishment and the threat of feuds. Journal of Economic Behavior & Organization 78 (3), 319–332 (May).

Nikiforakis, Nikos, Normann, Hans-Theo, 2008. A comparative statics analysis of punishment in public good experiments. Experimental Economics 11, 358–369. http://dx.doi.org/10.1007/s10683-007-9171-3.

Noussair, Charles, Tucker, Steven, 2005. Combing monetary and social sanctions to promote cooperation. Economic Inquiry 43 (3), 649–660.

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

Ostrom, Elinor, Walker, James, Gardner, Roy, 1992. Covenants with and without a sword: self-governance is possible. The American Political Science Review 86 (2), 404–417.

Savikhina, Anya, Sheremeta, Roman M., 2010. Visibility of contributions and cost of information: an experiment on public goods. Working Paper.

Sigmund, Karl, De Silva, Hannelore, Traulsen, Arne, Hauert, Christoph, 2010. Social learning promotes institutions for governing commons. Nature.
Steiner, Jakub, 2007. A trace of anger is enough: on the enforcement of social norms. Economics Bulletin 8, 1–4.
Sutter, M., Haigner, S., Kocher, M., 2010. Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. Review of Economic Studies 77.
Tyran, J., Feld, L., 2006. Achieving compliance when legal sanctions are non-deterrent. The Scandinavian Journal of Economics 108, 1467–9442.
Vesterlund, Lise, 2012. Voluntary giving to public goods: moving beyond the linear VCM. Working Paper.
Yamagishi, Toshio, 1986. The provision of a sanctioning system as a public good. Journal of Personality and Social Psychology 51 (1), 110–116.