Journal of Public Economics 144 (2016) 1–13

Contents lists available at ScienceDirect

Journal of Public Economics

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

Information-sensitive Leviathans ☆

Andreas Nicklisch [a], Kristoffel Grechenig [b], Christian Thöni [c,*]

[a] HTW Chur and German Research Foundation, Research Unit "Needs Based Justice and Distributive Procedures", Switzerland
[b] Max Planck Institute for Research on Collective Goods, Bonn, Germany
[c] University of Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 12 July 2015
Received in revised form 10 September 2016
Accepted 15 September 2016
Available online 19 September 2016

JEL classification:
C92
D02
H41

Keywords:
Centralized sanctions
Cooperation
Experiment
Endogenous institutions

ABSTRACT

We study information conditions under which individuals are willing to delegate their sanctioning power to a central authority. We design a public goods game in which players can move between institutional environments, and we vary the observability of others' contributions. We find that the relative popularity of centralized sanctioning crucially depends on the interaction between the observability of the cooperation of others and the absence of punishment targeted at cooperative individuals. While central institutions do not outperform decentralized sanctions under perfect information, large parts of the population are attracted by central institutions that rarely punish cooperative individuals in environments with limited observability.

1. Introduction

Human life in Thomas Hobbes' natural state is lonely, short, and brutal, "a time of war where every man is enemy to every man" (Hobbes, 1651). To redress this grim fate of violence and distrust, people appoint a central authority—a *Leviathan*—to enforce cooperative behavior. People voluntarily delegate their sanctioning power to the Leviathan, in the hope for a more efficient outcome.

In contrast to Hobbes' bleak view, contemporary research suggests that people successfully use *decentralized sanctions* (peer-to-peer punishment) to enforce cooperation (Ostrom et al., 1992; Fehr and Gäter, 2000) and reach efficient outcomes in the long run (Gächter et al., 2008). If human societies are able to organize themselves in a decentralized fashion, one would expect to find many self-governed societies. However, the opposite is the case: We live in a world where centralized sanctions play are very important role, on the national and even on the supranational level.[1] Why did modern societies develop centralized institutions to enforce norms? Under which conditions are people willing to renounce their sanctioning power in favor of a central authority?

We use an experimental approach to these questions, and we analyze a voting by feet mechanism in favor of or against central authorities. We introduce an environment where players ('citizens') participate in a social dilemma. Prior to this, they can vote by feet for one of three institutions: centralized punishment (*CenPun*), decentralized punishment (*DecPun*), and a sanction-free institution (*NoPun*). In *CenPun*, an additional (randomly drawn) subject (the 'authority') can punish the citizens in his institution, while citizens are not allowed to punish each other. The authority's payoff is increasing in the citizens' contributions, and the authority does not have to bear the costs of punishment. In *DecPun* citizens can punish other citizens in the same institution, at their own expenses.

☆ For helpful comments and discussion, we thank the Editor, two anonymous referees, Berno Büchel, Ernst Fehr, Urs Fischbacher, Guillaume Frechette, Simon Gächter, Sonja Köke, Manfred Milinski, Nikos Nikiforakis, Louis Putterman, Arno Riedl, Bettina Rockenbach, Roberto Weber, and the seminar participants at the University of Pennsylvania, the Max Planck Institute (Collective Goods, Bonn), the University of Nottingham, the University of Lausanne, and the University of Cologne. We would also like to thank the Max Planck Society and the German Research Foundation (Ni 1610/1-1) for financial support. An earlier version of this paper circulated under the title "Information-sensitive Leviathans: The Emergence of Centralized Punishment" (Nicklisch et al., 2015).

* Corresponding author.
E-mail address: christian.thoeni@unil.ch (C. Thöni).

[1] Examples are institutions like the European Union, the International Military Tribunal in Nuremberg in 1945/46, and the United Nations Security Council.

http://dx.doi.org/10.1016/j.jpubeco.2016.09.008
0047-2727/

Our analysis builds on three major challenges for self governance which have been identified in the literature: antisocial punishment, revenge, and incomplete information. Antisocial punishment (or perverse punishment) refers to the observation that some subjects target their punishment at cooperative subjects. There is ample evidence that the strength and frequency of antisocial punishment negatively relates to contributions.[2] Related is the problem of retaliation for received punishment. Some studies find that retaliation weakens decentralized punishment institutions because cooperative individuals are less willing to punish free riders (Denant-Boemont et al., 2007; Nikiforakis, 2008; Nikiforakis et al., 2012), while others do not find such a general effect (Kamei and Putterman, 2015). Finally, decentralized punishment can become inefficient in increasing contributions when subjects receive only imperfect information about the contributions of others. Contrary to intuition (but in accordance to the theoretical argument we develop below) subjects tend to punish more when information becomes more noisy (Grechenig et al., 2010; Ambrus and Greiner, 2012).[3]

All three problems are closely related. For instance, less information leads potentially to more punishment of cooperative subjects, which might in turn trigger retaliatory punishment. While in principle one could exogenously vary multiple dimensions of this complex interaction we restrict our design to a manipulation of the informational quality. In terms of the underlying phenomenon (establishing cooperation in groups) we think that it makes sense to see informational conditions as an exogenous characteristic of the environment, while the individual propensity to engage in antisocial punishment or revenge seems endogenous in its nature. Thus, our approach is to vary the quality of information exogenously and study its impact on the relative popularity of the three institutions. More specifically, we introduce three environments differing with respect to the accuracy of information citizens and the authority receive about the contributions of others. In treatment condition ONE, they receive accurate signals about the contributions; in POINT-NINE, they receive signals which are correct in 90% of the cases, while in POINT-FIVE, the signals are correct in 50% of the cases. We measure the popularity of an institution by the fraction of citizens it attracts.

We find that the treatment variation significantly influences institutional choices. In particular, imperfect information lowers the popularity of *DecPun*. We show that the punishment of cooperative citizens significantly influences institutional choices. Finally, with regard to our main research question we find that *CenPun* becomes the most popular institution only when there is imperfect information and at the same time the central authority (the Leviathan) applies a punishment strategy which minimizes the punishment of cooperative citizens. At the same time, revenge motives seem to be less important in our design and cannot explain differences across treatments.

Our study complements and expands recent discussions on the formation of centralized institutions. Dal Bó et al. (2010) compare the effect of a democratically chosen and an exogenously imposed policy intervention aimed at eliminating the attractiveness of free-riding. They find that democratically installed interventions increase cooperation significantly compared to exogenously imposed interventions. Moreover, endogenously introduced regimes with centralized sanctions perform well, even when sanctions are non-deterrent (Tyran and Feld, 2006), or, in some cases, outperform decentralized sanctions (O'Gorman et al., 2009).

Another important aspect of centralized institutions is the way how sanctions are implemented. In contrast to our approach, the majority of articles focus on sanctions that are executed automatically. If both, decentralized and automatically executed centralized punishment are available, the latter seems to crowd out the use of the former (Kube and Traxler, 2011). Markussen et al. (2014) investigate the choice of centralized sanctions through voting, when centralization is costly (and executed automatically). They find that people are particularly responsive to the fixed costs of having a centralized sanctioning scheme in place, more so than they respond to whether or not the sanctioning scheme is fully deterrent. Putterman et al. (2011) allow participants to vote on the rules of an automatically executed sanctioning scheme. The results show that many groups quickly implement sanctions that induce efficient outcomes.

Kosfeld et al. (2009) analyze the choice for automatically executed punishment mechanism which may govern only a subset of players. They show that participants are unwilling to implement equilibrium punishment which allows some players to free-ride. Andreoni and Gee (2012) investigate the formation of centralized sanctions through voting for a sanctioning scheme that punishes only the lowest contributor and find that full contributions are quickly achieved at very low punishment costs. Importantly, these articles focus on sanctions that are executed automatically; that is, once an implemented rule is violated, players are punished with a certain probability while contribution decisions are perfectly observable.[4]

Closer to our approach is Fehr and Williams (2013). They offer citizens the choice between uncoordinated decentralized, coordinated decentralized, or centralized punishment, which is executed by a democratically elected leader. They show that centralization of sanctions leads to high cooperation along with the selection of pro-social leaders who refrain from punishing high contributors. Similarly, Gross et al. (2016) explore the emergence of central punishment authorities under perfect information. They demonstrate that if individuals can transfer their punishment power to others, cooperators empower subjects who have previously indicated their willingness to sanction free-riders. As a consequence, groups with centralized punishment and high cooperation emerge.

Summarizing the previous literature, both centralized as well as decentralized sanctioning sustain cooperation if chosen endogenously. If available, evidence suggests that citizens choose very selectively centralized institutions. That is, effective centralized sanctioning is demanded, but citizens are unwilling to accept centralized punishment that violates their fairness considerations (e.g., allowing some players to free-ride, or punishment targeted at contributors).

In our setting, it is up to the authorities to deliver effective sanctioning. Like in Fehr and Williams (2013), we introduce the authority as a player, who may use punishment in a similar, potentially erroneous or malevolent fashion as his citizens.[5] We do so as we believe that the feature is of particular importance to explain the choice of authorities in earlier societies. That is, we compare centralized and decentralized sanctioning when authorities are not equipped with better mechanisms to guide behavior than citizens (e.g., our authorities are not better informed than citizens, nor do they have more efficient punishment technologies than citizens). Rather, our authorities are autocratic leaders, holding absolute punishment power. Furthermore, like in a feudal society the authority is not appointed by a competitive procedure, but he is merely born into his position.[6]

[2] See e.g. Gächter et al. (2005), Bochet et al. (2006), and Herrmann et al. (2008).

[3] On the other hand, Leibbrandt et al. (2015) provide evidence that antisocial punishment increases when *more* information is provided, i.e., when subjects can identify individual punishers in the group.

[4] See also Sutter et al. (2010), who study endogenous choices between positive and negative sanctioning systems.

[5] This is similar to Carpenter and Matthews (2012), who analyze the effect of third-party punishment for contributions in public good games.

[6] For the effect of democratically appointed leaders see also Hamman et al. (2011) and Corazzini et al. (2014).