

Either All or No One: The Signing of Cooperation Agreements in Public Goods Experiments

Michael Kosfeld (University of Zurich)

Akira Okada (Hitotsubashi University Tokyo)

Arno Riedl (University of Amsterdam)

Extended Abstract

April 2005

When markets fail, the design of appropriate institutions is a key issue for economic analysis and policy. In social dilemma situations, where social welfare maximization typically conflicts with individual payoff maximization, the implementation of a centralized punishment institution that castigates individual behavior if deviant from the welfare maximizing action is a possible solution. The classic example of such an institution is the constitutional state that maintains cooperation of its citizens through enforcement by central authorities (police, courts). Others include the EU Stability and Growth Pact created to enforce budgetary discipline of EU member states, or the Kyoto protocol aiming at a reduction of global CO₂ emissions by means of the implementation of legally binding agreements.

Under the assumption that the institutional solution works, an important question remains: how does it come into existence? Ultimately, if not imposed from the outside, the original set of agents must endogenously form the institution. It is exactly this process of institution formation that we address in this paper.

We analyze — theoretically and experimentally — the endogenous creation of a centralized punishment institution in an n -player public goods game. Under such an institution, individual agents commit themselves to contribute the welfare-maximizing quantity to the public good by signing a binding agreement that transfers the right of punishment in case of defection to a central authority.

The process of institution formation is modeled as a three-stage game: first, each agent decides whether he wants to participate in an institution that — once it is implemented — has the sole right to punish every participating agent who does not contribute the efficient amount to the public good. The institution is costly and punishment is only meted out to participating agents. Hence, non-participants are allowed to free-ride on participants' contributions. In the second stage, agents learn how many of the other agents are willing to participate. The institution is implemented if and only if all agents who are willing to participate agree to its formation. In the final stage, the public goods game is played.

We first show that if agents have standard (selfish) preferences, a unique strict Nash equilibrium exists where exactly $2 \leq s^* \leq n$ agents agree to form the institution and consequently contribute to the public good, while the remaining $n - s^*$ agents do not participate and free-ride. The number of participating agents s^* depends on the payoffs in the public goods game and the cost of the institution. In particular, it may be that s^* is strictly smaller than n , i.e., the unique equilibrium institution has a proper subset of agents who voluntarily commit themselves to cooperation. The institution is implemented because each participating agent still earns a higher payoff than in the unique Nash

equilibrium of the game with no institution, where no one contributes to the public good. It is conceivable, however, that this equilibrium might violate a fairness norm, since agents outside the institution can free-ride on the contribution of the participating agents.

By now there is ample evidence that many individuals reveal a preference for fairness and equity, especially when it comes to social dilemma situations. We therefore analyze the game under the assumption that agents have social preferences and use the Fehr-Schmidt (1999) model to take such preferences into account. In this case, multiple strict Nash equilibria may arise and the equilibrium size of the institution depends on agents' marginal utility loss from payoff inequity. In particular, we show that if each agent's concern for inequity is large enough, no institution with less than all agents participating can be sustained by an equilibrium, that is, only the grand institution ($s^* = n$) remains.

Finally, we present the results of a laboratory experiment designed to investigate the process of institution formation in a four-player public goods game. The experiment goes beyond the existing literature as it connects the classic social dilemma situation with an innovative element of political organization. Based on the model outlined above we implement two main treatments varying the equilibrium size s^* of the institution under the selfish-preference prediction. In the first treatment $s^* = 2$, in the second treatment $s^* = 3$. We also conduct two corresponding control treatments where no institution can be formed and subjects play only the public goods game. In each treatment subjects play 20 periods of the relevant stage game with the composition of groups remaining constant in each period.

Our main experimental findings are as follows.

1. Player successfully form institutions and the number of institutions increases over time. In both treatments, between 70% and 100% of all groups form an institution in final rounds.
2. In both treatments, most institutions that are formed are of full size, i.e., all players participate. This result contrasts with the standard model but is consistent with the assumption that agents have social preferences.
3. The probability that an institution is implemented increases with the number of agents willing to participate. If all agents are willing to participate, the average implementation rate amounts to 0.91 (0.69) in the first (second) treatment. In contrast, when one or two agents say they are not willing to participate the probability for an institution to be implemented falls below 0.38 (0.23) in the first (second) treatment.
4. Compared to the control treatments, the possibility of institution formation significantly increases total contributions to the public good.

Our study produces important insights into the process of institution formation that are relevant for public policy. First, since centralized punishment institutions represent an attractive solution for many social dilemma situations, our finding that subjects show

a general willingness to form such institutions is good news. However, subjects form institutions that differ from those predicted by the standard economic model. In particular, subjects are reluctant to implement a punishment institution that governs only a subset of the agents. This happens, even if the participating agents would earn a higher payoff compared to the non-production of the public good. The result brings to mind the recent discussion of possible consequences of some nations' withdrawal from the Kyoto protocol on other nations' motivation to fulfill the agreement.