Cooperation and Trustworthiness in Repeated Interaction

Tobias Cagala, Ulrich Glogowsky, Veronika Grimm, Johannes Rincke*

February 5, 2015

Public goods provision often involves groups of contributors repeatedly interacting with administrators who can extract rents from the pool of contributions. We suggest a novel identification approach that exploits the sequential ordering of decisions in a panel vector autoregressive model to study social interactions in the laboratory. Despite rent extraction, contributors and administrators establish a stable interaction with cooperation matching the level from a comparable Public Goods Game. In the short run, temporary changes in behavior trigger substantial behavioral multiplier effects. We demonstrate that cooperation breeds trustworthiness and vice versa and that one-time disruptions are particularly damaging in settings with a lack of cooperative attitudes and trust.

JEL codes: C32; C91; C92; H41
Keywords: Cooperation; trustworthiness; rent extraction; methods for laboratory experiments; panel vector autoregressive model

*Cagala: University of Erlangen-Nuremberg (tobias.cagala@fau.de); Glogowsky: University of Erlangen-Nuremberg (ulrich.glogowsky@fau.de); Grimm: University of Erlangen-Nuremberg (veronika.grimm@fau.de); Rincke: University of Erlangen-Nuremberg (johannes.rincke@fau.de). We thank Charles Bellemare, Jacob Goeree, Martin Kocher, Ernesto Reuben, Klaus Schmidt, Dirk Sliwka, Joachim Weimann, and numerous seminar participants for helpful comments. We are grateful for financial support from the Emerging Field Initiative at University of Erlangen-Nuremberg. All errors are our own.

1 Introduction

Pioneered by Isaac *et al.* (1985) and Isaac and Walker (1988), a substantial literature on cooperation in social dilemma situations has emerged. This literature has generated several insights on the impact of institutional environments on the overall level of cooperation (Gächter and Fehr 2000, Andreoni *et al.* 2003, Sefton *et al.* 2007, Gächter *et al.* 2008, Sutter *et al.* 2010, Baldassarri and Grossman 2011) and on the impact of peer effects on individual cooperation decisions (Keser and van Winden 2000, Fischbacher *et al.* 2001, Fischbacher and Gächter 2010).

Most contributions discussing the effects of institutions or peer effects on cooperation from the fact that cooperation often arises in environments where one or more individuals are entrusted with the responsibility of making the public goods available, a role that we naturally label as that of administrators. The fact that administrators control the pool of contributions creates incentive for rent extraction and eventually results in a diminished efficiency of public goods provision. Examples are numerous: taxpayers' gains from tax-compliant behavior depend on the efficiency within the public administration and the level of corruption; the benefits that members of a research team enjoy from scientific success depend on the communication of individual contributions by the principal investigator; members of work teams often face the risk that the team leader may appropriate part of the benefits (bonuses, promotions, etc.) resulting from cooperation among team members.

Studying public goods provision while allowing for the presence of an administrator creates a setting that, in addition to horizontal cooperation, embeds social interactions between the group of contributors and the administrator. The latter layer of interaction has rarely been studied and is, therefore, not well understood.[1]

In this paper, we aim at closing this gap by focusing on two important issues. First, we examine how the presence of an administrator who extracts part of the pool as a private rent affects the overall level and the stability of cooperation relative to a setting with exogenous provision. This links our discussion to the literature studying cooperation in the Public Goods Game. Second, going beyond the overall impact of rent extraction, we study the social interaction between contributors and administrators by analyzing how individual cooperation and rent extraction decisions affect cooperation and rent extraction behavior in subsequent periods. This part of the analysis aims

[1]This holds also true for applied work. For instance, there is little evidence on how reciprocity between taxpayers and government authorities affects the individual's willingness to pay taxes (Luttmer and Singhal, 2014). Studies using survey data typically find positive correlations between trust in government and tax morale (for a review, see OECD 2013), but it is challenging to isolate causal effects with this kind of data. We are not aware of empirical work analyzing the two-way relationship between contributors and administrators in applied settings. However, where researchers have looked at one-directional effects, the evidence seems in line with our main findings. Cullen *et al.* (2014), for instance, find that compliance with federal taxes in U.S. counties positively depends on the degree of political alignment with elected officials.

at understanding how cooperation evolves over time and how temporary disruptions originating from changes in the behavior of contributors and the administrator affect cooperation.

To investigate both topics in an integrated framework, we consider a repeated game that we call the *Public Trust Game*. This game combines the key elements of the *Public Goods Game* (Isaac and Walker 1988) and the *Trust Game* (Berg *et al.* 1995). In particular, we let contributors' payoffs depend on the size of the pool of contributions as in the Public Goods Game but we replace the mechanical distribution of the public good by a decision of an administrator. The administrator decides which part of the public good to keep to herself and which part to return to the group of contributors. This aspect relates our design to the Trust Game. Group members' benefits from *cooperation* depend on the administrator's *trustworthiness*.

Given this framework, it is straightforward to analyze how rent extraction of an administrator affects the overall level of cooperation: we compare the level of cooperation in the Public Trust Game (where provision is endogenous) with the level of cooperation in the Public Goods Game (where provision is exogenous). In contrast, because the repeated interaction between both types of agents leads to a mutual interdependence between cooperation and trustworthiness, studying the interaction between the administrator and the group of contributors is more involved. We suggest an identification approach that accounts for the resulting endogeneity. In particular, we adapt a panel vector autoregressive model to our design and exploit the sequential structure of the game to identify the effects of one-time changes in cooperation (i.e., the size of the pool of contributions) and one-time changes in trustworthiness (i.e., administrators diversion behavior) on cooperation and trustworthiness in subsequent periods. We are not aware of any previous attempts to use similar identification techniques on experimental data. A key property of our approach is that we derive exclusion restrictions directly from the experimental design.[2]

Three sets of findings emerge from our analysis. First, we demonstrate that the level of cooperation in the Public Trust Game is comparable to a standard Public Goods Game with the same efficiency. This can be explained in the spirit of a theory of sequential reciprocity with contributors who perceive the administrator's behavior as neutral. Survey evidence supports this interpretation: on average, contributors in the Public Trust Game perceive the behavior of the administrator as midway between completely satisfactory and completely unsatisfactory.

Second, by studying the repeated interaction among contributors and administrators, we demonstrate that cooperation breeds trustworthiness and vice versa. In

[2]The proposed methods are applicable to a broad family of repeated games where the outcomes of interest are jointly determined autoregressive processes, the resulting time series are stationary, and agents have distinguishable roles.

particular, a one-time increase (decrease) in cooperation triggers a significant increase (decrease) in cooperation and trustworthiness in subsequent periods. Similarly, a one-time increase (decrease) in the trustworthiness positively (negatively) affects future cooperation and trustworthiness. All these responses are, however, of a temporary nature, with behavior eventually converging back to pre-shock levels of cooperation and trustworthiness. One conclusion is that temporary changes in the administrator's trustworthiness have only temporary effects and do not permanently alter the climate for cooperation.

To measure the overall impact of one-time shocks in behavior, we derive multipliers that take feedback effects and all future responses into account. We naturally label these effects *behavioral multipliers*. It turns out that the behavioral multipliers are substantial: the overall impact of a shock in trustworthiness on cooperation is a multiple of the initial impulse, and a similar multiplier boosts the overall impact of contribution shocks on the administrator's trustworthiness. An additional insight resulting from studying impulse responses is that impulses in cooperation are more important to explain the observed level of variation in cooperative behavior than impulses in trustworthiness.

Our third set of findings emerges from studying the individual heterogeneity in baseline attitudes towards cooperation and trust. Exploiting survey data that we collected from the subjects several weeks after the experiment, we show that in groups with less cooperative and less trusting types, the behavioral multipliers are much larger than with more cooperative and more trusting types. This effect is most pronounced among contributors. For instance, the overall response of contributors reporting low levels of trust to one-time changes in their administrator's trustworthiness is almost four times larger compared to groups of contributors reporting high levels of trust. The finding of heterogeneous impulse responses has important implications. In particular, our analysis suggests that one-time disruptions in cooperation or trustworthiness are particularly damaging in settings with a lack of cooperative attitudes and trust.

Our paper contributes to two strands of literature. First, we extend the literature that evaluates the impact of exogenous institutional variations on the level of cooperation. For example, Gächter and Fehr (2000), Anderson and Putterman (2006), and Gächter *et al.* (2008) show that the possibility of peer punishment increases cooperation in Public Goods Games.[3] Baldassarri and Grossman (2011) demonstrate that sanctions by administrators are an effective tool to increase cooperation. In contrast to

[3]Several contributions discuss further aspects of punishment. Contributors make use of punishment even if the group composition changes each period (Fehr and Gächter 2002, Anderson and Putterman 2006). Furthermore, the effectiveness of punishment in fostering cooperation depends on monitoring possibilities (Carpenter 2007), on counter punishment opportunities, and on whether sanctions are monetary or non-monetary (Masclet *et al.* 2003). Reuben and Riedl (2009) find that groups with a distinguished player with a higher marginal per capita return of contributions make ineffective use of costly sanctions.

Baldassarri and Grossman (2011), the administrator in our design decides to extract a rent from the pool of contributions rather than punishing contributors. Interestingly, a mixture of rewards and punishment seems to be most effective (Andreoni *et al.* 2003, Sefton *et al.* 2007). This relates to our study, where contributors may interpret deviations from the expected rate of return (or reference point) induced by the administrator in terms of reward and punishment. More closely related to our study in terms of experimental design is the "team allocator game" studied by Kocher *et al.* (2013). In this game, a distinguished team member has property rights over the benefits from the public good. It turns out that because the distinguished agent uses her allocation power in a way that motivates ordinary agents, cooperation is higher compared to a standard Public Goods Game.

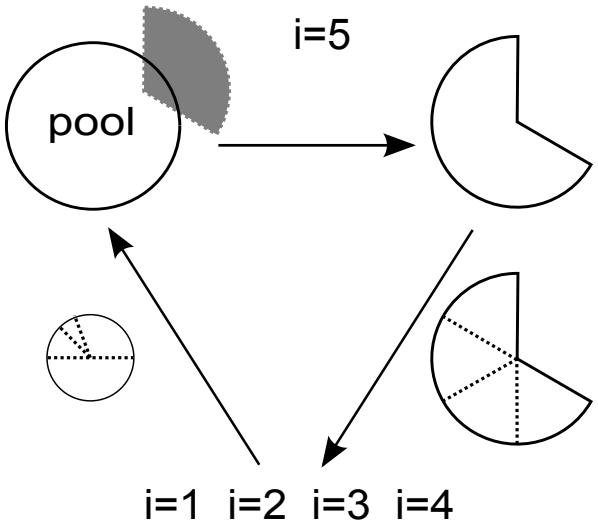
Second, our study adds to the literature on how social interactions affect cooperation. Our finding that cooperation breeds trustworthiness (and vice versa) relates to Keser and van Winden (2000), Fischbacher *et al.* (2001) and Fischbacher and Gächter (2010), who show that many individuals act as conditional cooperators. Bochet *et al.* (2006) and Brosig *et al.* (2003) find that the opportunity to communicate facilitates coordination in the interaction between contributors. Without communication, the presence of a contributor who leads by example increases cooperation (Güth *et al.* 2007). In contrast to the previous literature, we do not study peer interactions but focus on interactions between agents that play inherently different roles in the process of public goods provision.

The article is organized as follows. Section 2 describes the experimental design, Section 3 compares cooperation under exogenous and endogenous provision, Section 4 studies the social interactions, and Section 5 concludes.

2 Experiment

The Public Trust Game (PTG) extends the Public Goods Game (PGG) by introducing an administrator who decides which part of the pool of contributions to keep for herself. Only the remaining part of the pool is used for public goods provision, i.e., equally distributed among the contributors. The provision of public goods, thus, depends on the decision of the administrator. Comparing the PTG to the Trust Game (TG), contributors' (trustors') *cooperation* reflects the collective level of trust, while the part of the pool the administrator (trustee) returns mirrors her *trustworthiness*. Figure 1 summarizes our experimental design.

Figure 1: Experimental Design



Notes: The Figure visualizes the experimental design of the Public Trust Game.
Summary: Subjects interact for 30 periods in groups of 5 agents that consist of four contributors $i = \{1, 2, \ldots, 4\}$ and one administrator $i = 5$. Each period consist of two stages: in the first stage, all contributors choose their individual contribution m_{it} to the public good ($0 \leq m_{it} \leq w = 10$). In the second stage, the administrator decides which value of the pool M_t (tripled sum over contributions) is returned and equally redistributed among contributors R_t and which part of the pool she keeps for herself ($R_t \leq M_t$).

In the following, we discuss the details of our design. Let $i = \{1, 2, \ldots, 5\}$ denote a randomly generated group of 5 agents who interact repeatedly in $T = 30$ periods. We call agents $i = \{1, 2, 3, 4\}$ contributors and agent $i = 5$ the administrator. Each period $t = \{1, 2, \ldots, 30\}$ consists of two stages:

In the first stage, all contributors, endowed with $w_i \equiv w \equiv 10$ tokens, choose their individual contribution $m_{it} = \{0, 1, \ldots, 10\}$ to a public good. The sum of individual contributions is multiplied with the efficiency factor $r = 3$, resulting in the pool $M_t = 3 \sum_1^4 m_{it}$.

In the second stage, the administrator, endowment with $w_5 \equiv 30$ tokens, obtains control over the pool. She has to decide which part of the pool $R_t = \{0, 1, \ldots, M_t\}$ to return to the group of contributors. Whereas this returned part of the pool is equally distributed among the contributors, the administrator keeps the remaining part of the pool to herself.

Diverting resources from the pool changes the efficiency of public goods provision. The true efficiency factor is $\hat{r}_t = (1 - \gamma_t)r$, where $\gamma_t = \frac{M_t - R_t}{M_t} \in [0, 1]$ is the share of the pool kept by the administrator (extraction rate).

While the administrator is making her decision, all contributors indicate their belief about the return $\hat{R}_{it}$. We elicit beliefs in two steps: first, each contributor indicates her belief about the mean contribution of other group members $\hat{m}_{it} = \{0, 1, \ldots, 10\}$. Second, we calculate the individual hypothetical pool $\hat{M}_{it} = 3(m_{it} + 3\hat{m}_{it})$ and elicit con-

tributors' beliefs about the amount the administrator will return $\hat{R}_{it} = \{0, 1, \ldots, \hat{M}_{it}\}$.

At the end of each period, the contributors and the administrator receive information on the endowments of all agents, the size of the pool M_t, the return R_t, and their own profit in period t. Agents' payoffs x_{it} in period t are

$$x_{it} = w - m_{it} + \frac{3}{4}\sum_{j=1}^{4} m_{jt} - \frac{3}{4}\gamma_t \sum_{j=1}^{4} m_{jt}, \quad i = \{1, \ldots, 4\}, \tag{1}$$

$$x_{5t} = w_5 + 3\gamma_t \sum_{j=1}^{4} m_{jt}. \tag{2}$$

Equations (1) and (2) imply that $x_{it} \in [0, 30]$ and $x_{5t} \in [30, 150]$. The administrator, hence, earns at least as much as any contributor. This rules out that contributors can reasonably interpret return rates below one as supportive to the fairness of the payoff allocation.

The design of the PTG provides us with a framework to study the two central topics of our paper. First, we identify the total effect of endogenous public goods provision on the overall level of cooperation by comparing cooperation in the PTG (endogenous provision) to cooperation in the PGG (exogenous provision). We ensure that the efficiency in the PTG and in the PGG is comparable. In particular, we compare the level of cooperation in the PTG with the level of cooperation in a standard four-agent PGG with an efficiency factor that equals the mean efficiency factor $\hat{r} = 2$ in the PTG.[4]

Second, we study the social interaction between contributors and administrators by analyzing how individual cooperation and rent extraction decisions affect cooperation and trustworthiness in subsequent periods by adapting a panel vector autoregressive (PVAR) model to our design. The approach extracts exogenous variation in behavior and exploits these behavioral changes (called shocks or impulses) as *quasi-treatments* to evaluate the causal effects on future values of cooperation and trustworthiness.

Further details of implementation are as follows. The computerized experiment took place between December 2011 and May 2012 in the Laboratory for Experimental Research Nuremberg.[5] In total, 178 students from the University of Erlangen-Nuremberg participated in 6 sessions, generating 18 (22) independent observation in the PTG (PGG). After reading instructions,[6] subjects answered computerized control questions, participated in the PTG and filled out a questionnaire on individual characteristics and game-related issues. The same person led the experiment in all sessions. We invited

[4]We implemented the true efficiency factor based on the actual average extraction rate in the PTG: $\hat{r}_t = (1 - \gamma_t)r = (1 - 0.285) * 3 \approx 2$.

[5]We programmed the experiment with z-Tree (Fischbacher 2007) and recruited subjects with ORSEE (Greiner 2004).

[6]For instructions, see the Appendix.

subjects for a second time to answer survey questions on attitudes towards cooperation and trust. To attenuate the influence of subjects' experience in the PTG on response behavior, we conducted the survey two weeks after the experiment. Sessions lasted approximately 100 minutes; answering the paper-based questionnaire took 30 minutes. In the PTG contributors (administrators) earned €13.4 (€32.8) on average, including a €8.5 show-up fee. Average earnings of contributors in the PGG were €13.3.

3 Level of Cooperation Under Rent Extraction

3.1 Theoretical Considerations

In this section, we discuss the existence of cooperative equilibria in the PTG and show how the presence of a rent extracting administrator influences the overall level of cooperation.[7] Any equilibrium of the one-shot PTG or PGG predicts zero contributions if all agents were rational payoff maximizers and this was common knowledge among them. Also any subgame perfect equilibrium of the finitely repeated game has zero contributions in every period. In contrast, the recent literature has elaborated on various motives that may contribute to explain cooperation and trustworthiness in the repeated (or even one-shot) PTG. In the following, we discuss the impact of two of those motives, namely repeated interaction and reciprocity concerns, on the set of equilibria in our setup.[8]

3.1.1 Infinitely Repeated Interaction

Under repeated interaction with an infinite (or uncertain) horizon, agents face a trade-off between current and future profits. This gives rise to cooperative outcomes if future profits are considered valuable enough.[9] In the PTG, the incentives of contributors to cooperate depend on the individual discount factor, other contributors' behavior, and the level of rent extraction by the administrator.

Let us focus on the conditions under which cooperative equilibria exist.[10] First, there is no equilibrium with no or complete rent extraction. Second, increasing the extraction rate above zero raises the critical discount factor for contributors above the level that sustains cooperation in the repeated PGG. Clearly, because rent extraction reduces the true efficiency factor, it diminishes the scope for cooperation. At the same time, increasing the extraction rate decreases the critical discount factor that prevents

[7] We provide a detailed analysis including the proofs in an online appendix that accompanies the paper.

[8] The Fehr-Schmidt model of inequality aversion (Fehr and Schmidt 1999) predicts that cooperation is harder to sustain in the PTG than in a PGG with the equilibrium MPCR from the PTG.

[9] See Friedman (1971) and the follow up literature on the folk theorem.

[10] We assume for simplicity that extraction rates are similar across all periods.

the administrator from full rent-extraction. This points to a tradeoff in the repeated PTG: the level of anticipated rent extraction affects the incentives to cooperate and, thus, future rent extraction possibilities. As a result, the administrator chooses an intermediate level of rent extraction as long as future profits are valuable enough.

Comparing the infinitely repeated versions of the PTG and the PGG, we find that the critical discount factors that sustain cooperation are identical for both games if we hold the efficiency constant. Hence, for standard preferences the analysis suggests similar levels of cooperation in the PTG and the PGG.

3.1.2 Reciprocity Concerns

Concerns for reciprocity imply that individuals care about the intentions that accompany actions (Rabin 1993). To understand how concerns for reciprocity might affect play in the PTG, we apply Dufwenberg and Kirchsteiger's 2004 theory of sequential reciprocity to our game (see the online appendix for details). Dufwenberg and Kirchsteiger propose a simple model where agent i perceives agent $j's$ action as kind (unkind) if $i's$ payoff is above (below) the average between her lowest and her highest possible material payoff resulting from $j's$ action. Dufwenberg and Kirchsteiger's utility specification implies an incentive for kindness towards others who have been kind to oneself and vice versa. As it turns out, a Sequential Reciprocity Equilibrium of the one-shot PTG with full contributions exists, if agents' reciprocity concerns are strong enough.

In the PTG extraction affects the scope for contributors' kindness. With zero extraction, contributors' decisions do not affect the administrator's payoff, rendering contributors' intentions towards her as neither kind nor unkind. As a result, the administrator cannot gain utility from reciprocating kindness. Therefore, reciprocity concerns can never induce the administrator to refrain completely from rent extraction. Furthermore, there exists a threshold level for the extraction rate: below this threshold, a Sequential Reciprocity Equilibrium with full cooperation exists. If rent extraction exceeds the threshold, i.e. if the administrator is too unkind, full cooperation cannot be sustained. Then, even kind behavior of other contributors cannot compensate for the unkind administrator's behavior and, thus, motivate positive contributions.

Let us finally compare the PTG to the standard PGG without administrator. Because the administrator's kindness provides an additional motive to contribute (besides other contributors' kindness), it is easier to sustain cooperation in the PTG than in the PGG whenever the administrator behaves kindly, and vice versa.