CESifo WORKING PAPERS

6801
2017

December 2017

Public Goods Provision with Rent-Extracting Administrators

Tobias Cagala, Ulrich Glogowsky, Veronika Grimm, Johannes Rincke

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 6801*
Category 13: Behavioural Economics

Public Goods Provision with Rent-Extracting Administrators

Abstract

This paper studies public goods provision in an experiment in which contributors repeatedly interact with rent-extracting administrators. Our main result is that the presence of an administrator reduces contributions but only because rent extraction lowers the MPCR. Analysing the dynamic interactions between the contributors and the administrator, we demonstrate that rent-extraction and cooperation shocks trigger short-run adjustments in the agents' behaviour. However, shocks do not have permanent effects. This explains the long-run resilience of cooperation to rent extraction. We also show that cooperative attitudes and trust are traits that explain permanent differences in the short-run volatility of public goods provision.

JEL-Codes: C320, C910, C920, H410.

Keywords: cooperation, rent extraction, corruption, trustworthiness, public goods, public trust game, panel vector autoregressive model.

Tobias Cagala
Deutsche Bundesbank
Frankfurt am Main / Germany
tobias.cagala@bundesbank.de

Ulrich Glogowsky
University of Munich
Munich / Germany
ulrich.glogowsky@econ.lmu.de

Veronika Grimm
University of Erlangen-Nuremberg
Nuremberg / Germany
veronika.grimm@fau.de

Johannes Rincke
University of Erlangen-Nuremberg
Nuremberg / Germany
johannes.rincke@fau.de

November 27, 2017
We thank Charles Bellemare, Andrea Galeotti (editor), Jacob Goeree, Martin Kocher, Ernesto Reuben, Klaus Schmidt, Dirk Sliwka, Joachim Weimann, two anonymous referees, and numerous seminar participants for helpful comments. We are grateful for the financial support from the Emerging Field Initiative at University of Erlangen-Nuremberg. Ulrich Schneider und Friederike Hertweck provided excellent research assistance. The paper represents the authors' personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank or its staff. All errors are our own.

1 Introduction

Public goods provision is usually conceptualised as a free-riding dilemma (Ledyard 1995; Fischbacher *et al.* 2001; Fischbacher and Gächter 2010). However, free riding is not the only threat to public goods provision: in many situations, poor public accountability is another reason why public services are not provided at a socially desirable level. For example, administrators in charge of public goods provision often misuse public funds for private gains, hereby depriving citizens of vital public services (Reinikka and Svensson 2004, 2005; Olken 2006, 2007). Nevertheless, surprisingly little is known about how rent-extracting administrators affect citizens' voluntary contributions. For instance, we know neither whether rent extraction aggravates the free-riding problem nor whether it triggers a vicious circle in which lower cooperation levels and higher rent extraction reinforce each other, leading to a race to the bottom in public goods provision.

This paper studies how the presence of a rent-extracting administrator affects the level and the dynamics of public goods provision. For that purpose, we propose the *public trust game*, which blends the key elements of the *public goods game* (Isaac and Walker 1988) and the *trust game* (Berg *et al.* 1995). As in the public goods game, the payoffs of the contributors in the public trust game depend on the size of the pool of contributions. However, we replace the mechanical provision of the public good with the decision of an administrator. The administrator decides which part of the pool of contributions to keep for herself and which part to return to the group of contributors. Hence, group members' benefits from cooperation depend on the administrator's rent extraction behaviour (i.e., her trustworthiness).[1] Importantly, in the public trust game, rent extraction reduces the efficiency of public goods provision by lowering the marginal per capita return (MPCR) of an investment in the public good. Our experimental design therefore mirrors many real-world contexts, like that of tax-funded public goods, in which the misuse of public funds reduces the efficiency of public goods provision.

Using this game, we analyse the effect of rent extraction on cooperation and public goods provision. In the first step of our analysis, we employ a treatment-based identification strategy to explore how the presence of a rent-extracting administrator affects contribution behaviour. Comparing the public trust game with standard public goods games, we find that the presence of an administrator reduces contributions but only because rent extraction has an adverse effect on the efficiency of public goods provision (i.e., through lowering the MPCR). Once we control for this loss in efficiency by comparing contributions in our public trust game and a public goods game with the same

[1] While we focus on a dynamic setting, Gächter *et al.* (2004) and Thöni *et al.* (2012) analyse the static relationship between trustworthiness and cooperation descriptively. More direct experimental evidence on the relationship between trustworthiness, trust, and cooperation comes from Karlan (2005), who finds that individuals who contribute to a public good also show higher levels of trust and trustworthiness in the trust game.

efficiency level, we find that the mean contributions are not affected by an administrator who embezzles resources from the pool. We also find that, in the public trust game, cooperation is stable over time. Hence, the presence of a rent-extracting administrator does not erode the established level of cooperation.

While the first step of our analysis "black boxes" the social interactions between the administrator and the contributors, the second step shifts our focus to interactions *within* the groups. In particular, we employ a panel vector autoregressive (PVAR) model and explore how rent-extraction shocks (i.e., exogenous increases in rent extraction) and cooperation shocks (i.e., exogenous decreases in cooperation) affect the subsequent decisions of the contributors and the administrator. Understanding the impact of such shocks is crucial for explaining why cooperation is resilient to rent extraction. For instance, an in-depth knowledge of how the contributors and the administrator interact after an exogenous increase in rent extraction allows us to understand how agents in the public trust game prevent a decline in cooperation through a vicious circle of higher rent extraction and lower contributions.

Using the PVAR model, we derive several findings. First, as expected, contributors cooperate less after an exogenous increase in rent extraction. More generally, we find that conditional cooperation, as documented by Fischbacher *et al.* (2001) and others, extends to cases in which players have different roles: trustworthiness of the administrator breeds cooperation by contributors and vice versa. However, all responses to shocks in the behaviour of the administrator and the contributors are of a temporary nature and do not have permanent effects.[2] Most importantly, after a one-time increase in rent extraction, cooperation and rent extraction eventually converge back to pre-shock values. In a nutshell, the analysis shows that cooperation and rent extraction are mutually dependent but do not reinforce each other. This explains why the presence of a rent-extracting administrator does not set off a downward trend in cooperation.

The second result emerging from the PVAR model refers to the relative importance of rent-extraction and cooperation shocks for explaining the overall variation in agents' behaviour. The analysis reveals that the variation in the behaviour of contributors is mainly explained by shocks in the contributors' instead of the administrator's behaviour. This implies that not only the *level* but also the *volatility* of cooperation is resilient to rent extraction. Taking all the findings on the level and the dynamics of cooperation together, the main insight from our analysis is that, apart from triggering short-run responses and affecting contributions through efficiency, rent-extracting administrators do not matter for cooperative behaviour.

[2]Our finding that shocks fade out indicates a resilience of cooperation that is in line with early evidence from the public goods game. Specifically, Andreoni (1988) finds that, after an unforeseen restart, high initial levels of cooperation are restored. He argues that this finding is inconsistent with pure learning. The evidence from the public trust game suggests that disruptions do not lead to learning effects that permanently affect cooperation.

Although rent-extraction shocks do not have long-lasting effects, understanding their impact on public goods provision is nevertheless essential. That is because unexpected changes in rent extraction make the provision of public goods less reliable in the short run. As a final step in our analysis, we therefore examine whether specific traits of contributors amplify the short-run impact of rent-extraction and cooperation shocks on contribution behaviour. If such traits exist, groups composed of contributors carrying those characteristics suffer from a higher short-run volatility of cooperation and hence from a less reliable provision of public goods. Our third finding is that the cumulative short-run responses to shocks are much larger in groups in which contributors are less cooperative and less trusting than in groups with more cooperative and more trusting contributors. This implies that cooperative attitudes and trust are traits that protect cooperation against the short-term volatility resulting from rent-extraction shocks.

Our paper contributes to several strands of literature. First, by adding an administrator to the public goods game, our work relates to the literature that studies the impact of institutional environments on the overall level of cooperation (e.g., Gächter and Fehr 2000, Andreoni *et al.* 2003; Masclet *et al.* 2003; Anderson and Putterman 2006; Carpenter 2007; Sefton *et al.* 2007; Gächter *et al.* 2008; Sutter *et al.* 2010). Whereas these studies highlight ways to curb the free-riding problem, we shift the focus to the analysis of rent extraction. The addition of an administrator to the public goods game is motivated by the facts that cooperation outside of the laboratory typically involves administrators who are entrusted with the provision of public goods and that the misuse of public funds for private gain is an endemic problem (Olken 2006, 2007). Second, our paper contributes to the experimental literature on corruption. This strand of literature examines either the corruptibility of administrators in games in which only administrators make decisions (e.g., Frank and Schulze 2000) or the bilateral strategic interaction between a briber and an administrator (e.g., Abbink *et al.* 2000); see Armantier and Boly (2008) for a review of the experimental bribery literature. In contrast, our paper focuses on the public goods context and considers corruption taking the form of rent extraction. Third, we add to the literature that studies interactions in public goods games (e.g., Keser and van Winden 2000; Fischbacher *et al.* 2001; Kocher *et al.* 2008; Herrmann and Thöni 2009; Kurzban and Houser 2005; Fischbacher and Gächter 2010). In contrast to previous papers that focus on the conditionality between contributors' decisions, we account for a second and often ignored layer of interaction between the contributors and an administrator.[3] Studying this interaction is essential for understanding the process

[3] We are not the first to focus on public goods games with asymmetric roles. For example, several studies introduce heterogeneity in endowments (e.g., Cherry *et al.* 2005; Buckley and Croson 2006; Sadrieh and Verbon 2006; Reuben and Riedl 2013; Charness *et al.* 2014) or marginal benefits from the public good (e.g, Fisher *et al.* 1995; Palfrey and Prisbrey 1997; Reuben and Riedl 2013). However, to the best of our knowledge, these studies do not provide an in-depth analysis of social interactions, as we do using the PVAR model.

of public goods provision, as disruptions in the administrator's behaviour could affect future cooperation over prolonged intervals of time.[4]

In terms of methods, the use of the PVAR model relates our study to work applying econometric techniques to data from laboratory experiments. For example, previous studies use time-series methods (e.g., Cesarini *et al.* 2009; Fischbacher and Gächter 2010), binary-choice models (e.g., Holt and Laury 2002; Cappelen *et al.* 2007), and finite-mixture regression models (e.g., El-Gamal and Grether 1995; Costa-Gomes *et al.* 2001; Dal Bo and Frechette 2011) to analyse experiments.[5] In contrast to methods that bury interactions by averaging over periods, such as treatment comparisons, the PVAR allows us to analyse the variation within groups directly. This enables us to investigate interactions by tracing the feedback loops triggered by one-time disruptions in the behaviour of the contributors and the administrator. Another valuable feature of the PVAR model is that it still exploits the laboratory setting for causal inference. The identifying assumption in our model is that we correctly specify the sequential order in which the contributors and the administrator make their decisions. The perfect control over the sequence of decision-making in the laboratory guarantees that this assumption is met.

The article is organised as follows: Section 2 describes the experimental design. Section 3 analyses the public trust game and derives our main findings, and Section 4 concludes.

2 The Public Trust Game

In the following, we study a repeated version of the public trust game (PTG). The PTG combines the public goods game (PGG) and the trust game (TG). Specifically, the PTG extends the 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 distributed equally among the contributors. The provision of public goods thus depends on the contributors' contributions (i.e., cooperation) and the administrator's rent extraction (i.e., trustworthiness). Adding a rent-extracting administrator to the PGG captures the role of trustworthy behaviour of administrators in a simple but stylised way. Another attractive feature of our design is that our measure of the administrator's trustworthiness (return to contributors) follows the standard measure of a trustee's trustworthiness in the TG (return to trustor).[6]

[4]Contributors' reactions to rent extraction may differ from their responses to free riding, because framing an action as appropriation (as compared with underprovision) results in stronger responses (Dufwenberg *et al.* 2011, Cox *et al.* 2013; Cox 2015).

[5]Others have estimated structural parameters; see, e.g., Ho *et al.* (1998). Also note a recent paper by Blume *et al.* (2011), who discusses and advocates the application of formal tools of statistical inference to the analysis of social interactions.

[6]Two links to the experimental literature are worth noting. First, contributors' cooperation reflects the collective level of trust. This connects our work to Cassar and Rigdon (2011), who study settings

In the following paragraphs, we discuss the details of our design. Let $i = \{1, 2, \ldots, 5\}$ denote a randomly generated group of five agents who interact repeatedly in $T = 30$ periods. We call agents $i = \{1, 2, 3, 4\}$ the contributors and agent $i = 5$ the administrator. The contributors are endowed with $w_i \equiv w \equiv 10$ tokens, and the administrator is endowed with $w_5 \equiv 30$ tokens. Each period $t = \{1, 2, \ldots, 30\}$ consists of two stages. In the first stage, all the contributors choose their individual contribution $m_{it} = \{0, 1, \ldots, 10\}$ to a public good. The sum of individual contributions is multiplied by the efficiency factor $r = 3$, resulting in the pool $M_t = 3\sum_1^4 m_{it}$.

In the second stage, the administrator 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. The returned part of the pool is distributed equally among the contributors and reflects the value of the public good; the administrator keeps the remaining part.[7] Expropriating 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 extraction rate (i.e., the share of the pool kept by the administrator). While the administrator makes her decision, the contributors indicate their beliefs about the return $\hat{R}_{it}$. We elicit beliefs as point predictions in two steps: first, each contributor indicates her belief about the mean contribution of the 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 contributors' beliefs about the amount that the administrator would return $\hat{R}_{it} = \{0, 1, \ldots, \hat{M}_{it}\}$ from this hypothetical pool.[8]

At the end of each period, all the agents in a group are informed about the group members' endowments, 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:

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

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

Two features result from the payoff structure. First, equation (1) shows that contributor

with multiple trustors that allow the trustee to discriminate between trustors. Second, there are other experimental games that include administrators in settings that differ from our experiment (see, e.g., Baldassarri and Grossman 2011; Kocher *et al.* 2013; Cagala *et al.* 2017).

[7] An alternative way of modelling trustworthiness via the administrator's choice of rent extraction is to let the administrator exert costly effort to prevent decay of the pool of contributions. While our design relates more to the issue of rent extraction by administrators, a design with administrators providing effort would rather address the problem of slack among bureaucrats.

[8] We follow Gächter and Renner (2010) and do not incentivise the elicitation of beliefs. This helps to attenuate the effect of belief elicitation on behaviour in the PTG. In particular, if beliefs are incentivised, then subjects can hedge the risk of contributing despite a low return rate by stating that the return rate will be low. Put differently, hedging affects stated beliefs and contribution decisions. In Section 3.2.2, we show that there is no evidence of any systematic bias in beliefs.

i's payoff depends on the behaviour of the administrator and of the other contributors. Therefore, not only expectations about free riding but also trust in the administrator could influence i's contribution decision.[9] Second, 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 the possibility of contributors reasonably interpreting return rates smaller than one as supporting the fairness of the payoff allocation. Also note that an individual's total payoff is the sum over her period-specific payoffs.

The sequence of events within the PTG was as follows: After reading the instructions (see the Online Appendix), the subjects answered computerised control questions, participated in the PTG, and completed a questionnaire on their individual characteristics and game-related issues. The same person led the experiment in all the sessions. We also invited 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 behaviour, we conducted the survey two weeks after the experiment. Attrition between the experiment and the survey was negligible: only three participants did not show up to answer the paper-based questionnaire.

The computerised experiment took place in the Laboratory for Experimental Research Nuremberg. We programmed the experiment with z-Tree (Fischbacher 2007) and recruited subjects with ORSEE (Greiner 2015). We obtained data for 18 independent groups for the PTG. For comparison we also implemented PGGs with two different efficiency factors. This procedure gave us a total of 44 independent PGG groups (see Subsection 3.1 for details). In total 266 students from the University of Erlangen-Nuremberg participated in the experiment.[10]

The sessions lasted for about 100 minutes, and answering the paper-based questionnaire took about 30 minutes. The contributors (administrators) in the PTG earned € 13.4 (€ 32.8) on average, including a € 2.5 show-up fee. The average earnings of the contributors in the PGG were € 13.2 (PGG2) and € 18.5 (PGG3), respectively. We paid the participants an additional fee of € 6 for answering the paper-based questionnaire.

3 Analysing the Public Trust Game

3.1 Level of Cooperation

We begin the analysis of the PTG by discussing how the presence of rent-extracting administrators influence the level of cooperation. As part of this analysis, we contrast the

[9]The fact that trust can influence contributors' decisions distinguishes our setting from multi-person dictator games (e.g. Oxoby and Spraggon 2008; Dasgupta 2011; Barr *et al.* 2015), in which trust plays no role.

[10]The 18 PTG sessions and 30 of the PGG sessions took place between December 2011 and May 2012. In response to a suggestion by a referee, we conducted 14 additional PGG sessions in May 2017.

contribution behaviour in the PTG with that in standard four-agent PGGs without an administrator.[11] As the PGG models an institutional environment without rent extraction, it allows us to test whether the presence of an administrator affects contributions.

We compare the PTG to two different PGGs. The first public goods game (PGG3) implements the efficiency factor $r = 3$, which corresponds to the efficiency level in the PTG without rent extraction. Comparing the contributions in the PGG3 with the contributions in the PTG allows us to quantify how the presence of rent-extracting administrators affects the contribution behaviour. The second public goods game (PGG2) implements an efficiency factor equal to the mean efficiency factor $\hat{r} = 2$ in the PTG.[12] This allows us to test whether a rent-extracting administrator further influences contributions once we control for the negative impact of rent extraction on efficiency.

Figure 1 shows the evolution of public goods provision in the PTG and the two different PGGs over time. Table 1 reports the corresponding descriptive statistics. The grey (blue) bars in Figure 1 represent the mean *contribution rates* (mean *return rates*) in the PTG. The individual contribution rates are $\overline{m}_{it} = 100\frac{m_{it}}{w} = 10m_{it}$, while the return rates are $\overline{R}_t = 100\frac{R_t}{4wr} = \frac{5}{6}R_t$.[13] The difference between the contribution rates and the return rates in the PTG corresponds to the share of the constant upper limit of the pool that is not returned to the contributors. The dashed lines depict the mean contribution rates in the two PGGs (red: $r = 2$; blue: $r = 3$). To account for the typical start-game and end-game effects, Table 1 presents the descriptive statistics for three intervals: the learning interval (periods 1 to 3), the main interval (periods 4 to 27), and the end-game interval (periods 28 to 30). In the following, we focus on the behaviour in the main interval. Our conclusions are unaffected if we include all the periods.

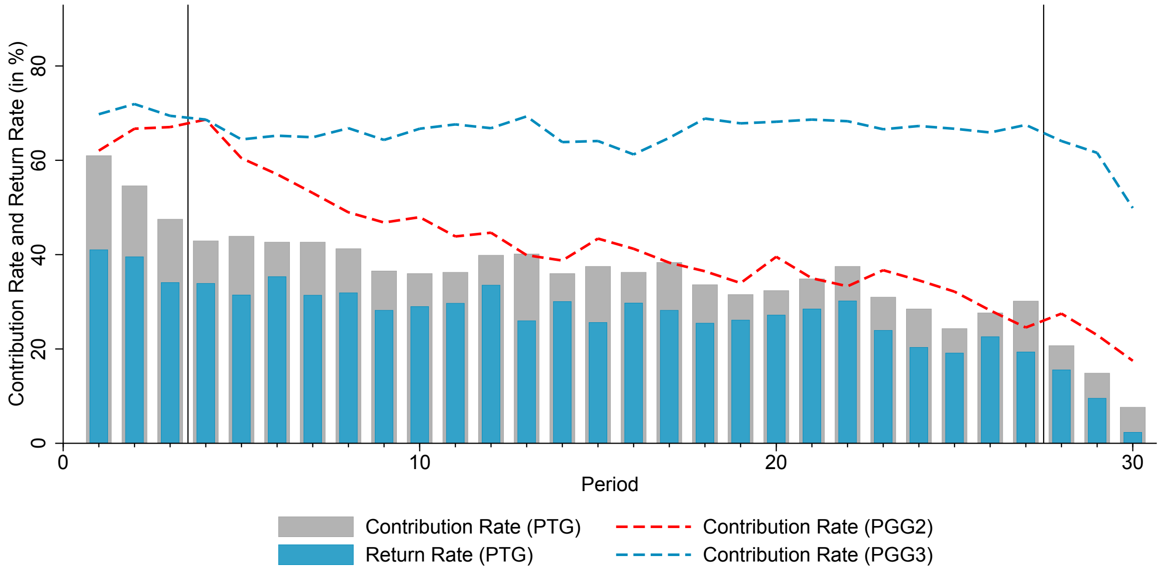
Based on Figure 1 and Table 1, we can establish the following main result. Comparing the mean contribution rates across treatments in the main interval, we find similar values for the PTG and the PGG2 (35.9% and 42.0%, respectively); the corresponding value for the PGG3 is much higher (66.4%). Regressing the group-average contribution rates on indicator variables for the different games and using robust standard errors, we cannot reject the null hypothesis of equal contribution rates in the PTG and the PGG2 (p-value$= 0.313$). However, we can reject the null hypothesis that the mean contribution

[11] The PGGs comprised only the four contributors as players. Endowments were identical to contributors' endowments in the PTG. The pool was fully and equally distributed to contributors.

[12] We calculate the true efficiency factor based on the average extraction rate in the PTG: $\hat{r} = (1-\gamma)r = (1-0.285)\cdot 3 \approx 2$. The overall difference in contributions between the PTG and the PGG incorporates all the effects that originate from the dynamic interaction between the contributors and the administrator in the PTG.

[13] The multiplication by $\frac{5}{6}$ standardises the return rate to lie in the same interval as the contribution rate, i.e. $[0;100]$. We relate returns to the maximum possible pool size $4wr$ instead of the actual pool, because the latter could result in a spurious relationship between the contribution and the return rate. If an administrator does not respond to shocks to contributions and always returns the same amount, a positive shock in contributions would lower the return relative to the actual contributions. In this case dividing by the actual returns would suggest a spurious negative relationship between the contributions and the relative returns.

Figure 1: Mean Contribution and Mean Return Rates Over Time



Notes: The Figure contrasts behaviour in the public trust game with that in two different public goods games. The grey bars display mean contribution rates in the public trust game and the blue bars show mean return rates. The red dashed line shows mean contribution rates in the public goods game with $r = 2$ (PGG2). The blue dashed line displays mean contribution rates in the public goods game with $r = 3$ (PGG3).

Table 1: Descriptive Statistics

Variable	*Learning Interval* Mean	*Learning Interval* Std.Dev.	*Main Interval* Mean	*Main Interval* Std.Dev.	*Main Interval* Median	*Main Interval* Min.	*Main Interval* Max.	*End-Game Interval* Mean	*End-Game Interval* Std.Dev.
Contribution Rate PTG	54.4	17.6	35.9	23.5	36.3	0	90.0	14.4	17.8
Return Rate PTG	38.2	20.3	27.8	22.2	25.0	0	83.3	9.1	12.8
Contribution Rate PGG2	65.3	23.4	42.0	27.2	40.0	0	100.0	22.7	20.1
Contribution Rate PGG3	70.4	19.5	66.4	26.6	67.5	0	100.0	58.5	28.0

Notes: Observational unit: group n in period t. The total number of observations is 540 in the PTG ($N = 18$, $T = 30$), 660 in the PGG2 ($N = 22$, $T = 30$), and 660 in the PGG3 ($N = 22$, $T = 30$).

rates in the PTG and the PGG3 are equal (p-value< 0.001). The same holds true for a comparison between the two PGGs (p-value< 0.001).[14] Hence, an exogenous reduction in the efficiency of public goods provision has the same diminishing effect on cooperation as an equivalent drop in the efficiency level that is due to a rent-extracting administrator. Our first main finding from the analysis of the PTG is therefore that the presence of a rent-extracting administrator reduces contributions relative to a public goods game with a multiplier of $r = 3$. The effect is substantial: the average contribution rate in the PTG is 30.5 percentage points lower than in the PGG3. However, rent extraction leaves con-

[14] Besides running regressions with group-average contribution rates, we also perform Mann-Whitney U tests at the group-level and regressions using individual contribution rates in each period (instead of the average contribution rates), with standard errors clustered at the group-level. In both testing procedures, we cannot reject the null of equal mean contribution rates between the PTG and the PGG2 (p-values$>$ 0.300), and we can reject the null for the comparison between the PTG and the PGG3 (p-value< 0.001).

tributors with fewer public goods in exchange for their contributions and hence reduces the marginal per capital return from the public good. Once we control for this loss in efficiency by comparing the PTG with the PGG2, there is no evidence of a systematic difference in the average contribution levels. We conclude that, besides reducing efficiency, the presence of a rent-extracting administrator does not matter for the level of cooperation.[15]

We also note that, apart from the typical start-game and end-game effects, the administrators and contributors in the PTG manage to establish a form of interaction that is stable over time. Various panel tests of stationarity confirm this descriptive observation for the contribution rates and return rates in the PTG; for details, see Table A1 in the Appendix. Hence, the presence of a rent-extracting administrator in the PTG does *not* lead to a vicious circle of higher rent extraction and lower contributions. In fact, with respect to the stability of cooperation, the PTG is more similar to the PGG3 than to the PGG2.[16] Taken together, comparing the PTG with the PGGs establishes the following main result:

RESULT 1: *The presence of a rent-extracting administrator reduces cooperation in the PTG relative to a standard PGG without an administrator. However, the level of cooperation in the PTG is not different from the level in a PGG in which the exogenously set MPCR mirrors the loss in efficiency due to rent extraction. Moreover, the observed levels of cooperation and rent extraction in the PTG are stable over time.*

In the following, we focus on the dynamics of the PTG and analyse the interactions between the contributors and the administrator. In particular, we examine how the contributors respond if they experience a rent extraction shock (i.e., an exogenous increase in rent extraction) and whether such responses trigger further behavioural responses from both types of players. Understanding the impact of such shocks helps us to gain a better understanding of why public goods provision is resilient when the administrator repeatedly interferes with the cooperation among contributors. Before analysing the data, we set out a conceptual framework to capture the dynamics of the PTG.

[15] In the Online Appendix, we discuss the existence of cooperative equilibria in the PTG that result in similar levels of cooperation as in a standard PGG. Specifically, we show that our findings on the overall level of cooperation are consistent with a model of sequential reciprocity as long as contributors perceive the behaviour of the administrator as neutral (i.e., neither kind nor unkind).

[16] The literature shows that cooperation is more stable when the multipliers are higher. See, for example, Isaac *et al.* (1984) for early evidence with a multiplier of $r = 3$.

3.2 Dynamics of Cooperation and Rent Extraction

3.2.1 Conceptual Framework

Subsequently, we build on the fundamentals of decision-making from the learning literature (see Fudenberg and Levine (1998) for a survey) to delineate a conceptual framework that describes how agents in the PTG adjust their decisions over time. We assume that agents view their actions in the stage game as the object of choice. Each contributor selects her contribution m_{it} according to a decision rule that links their contributions to their beliefs about the public goods return $\hat{R}_{it}$, where

$$m_{it} = f(\hat{R}_{it}, \theta_i) + s_{it}. \tag{3}$$

Eq. (3) consists of a deterministic and a stochastic component. The deterministic component of the decision rule $f(\cdot)$ reflects i's best response to the expected behaviour of other agents, where θ_i captures individual time-invariant characteristics and $\hat{R}_{it}$ is i's belief about the public goods return. Because the return R_t depends on contributors' decisions on contributions and the administrator's decision on provision, we implicitly assume that each contributor builds $\hat{R}_{it}$ based on her expectations about the behaviour of both types of agents. The random variable s_{it} represents the stochastic component of the decision rule, which reflects disturbances (i.e., shocks) in contributions on an individual level. It captures random errors in decision-making that lead to deviations from the best responses.

Similar to the contributors, the administrator chooses the return R_t, contingent on a deterministic component $g(\cdot)$ and a stochastic component v_t that reflects disturbances in the administrator's behaviour:

$$R_t = g(M_t, \phi) + v_t, \tag{4}$$

with ϕ capturing the time-invariant administrator characteristics and $M_t = r\sum_{i=1}^{4} m_{it}$. Because the administrator has complete information on the pool size prior to her decision, her decision rule depends on the realisation of M_t. According to (3) and (4), changes in contributors' beliefs about the return directly induce changes in the pool size and indirectly (i.e., via the pool) affect the return.

Importantly, the form of belief updating by contributors drives the dynamics in our game. We follow the approach of Fischbacher and Gächter (2010) and assume that contributors form their beliefs in period t on the basis of their beliefs in period $t-1$ and on past realisations.[17] In particular, we assume a canonical learning rule that incorporates

[17] A natural (but more complicated) alternative would be a framework that models contributors who build beliefs by additionally incorporating the expected consequence of (their own) actions on the future behaviour of other agents.

the concept of adaptive expectations:

$$\hat{R}_{it} = h(\hat{R}_{it-1}, \Delta R_{it-1}, \theta_i), \tag{5}$$

with $h(\cdot)$ being a function of the lagged belief $\hat{R}_{it-1}$, the lagged error of expectation $\Delta R_{it-1} = R_{t-1} - \hat{R}_{it-1}$, and the time-invariant individual characteristics θ_i.

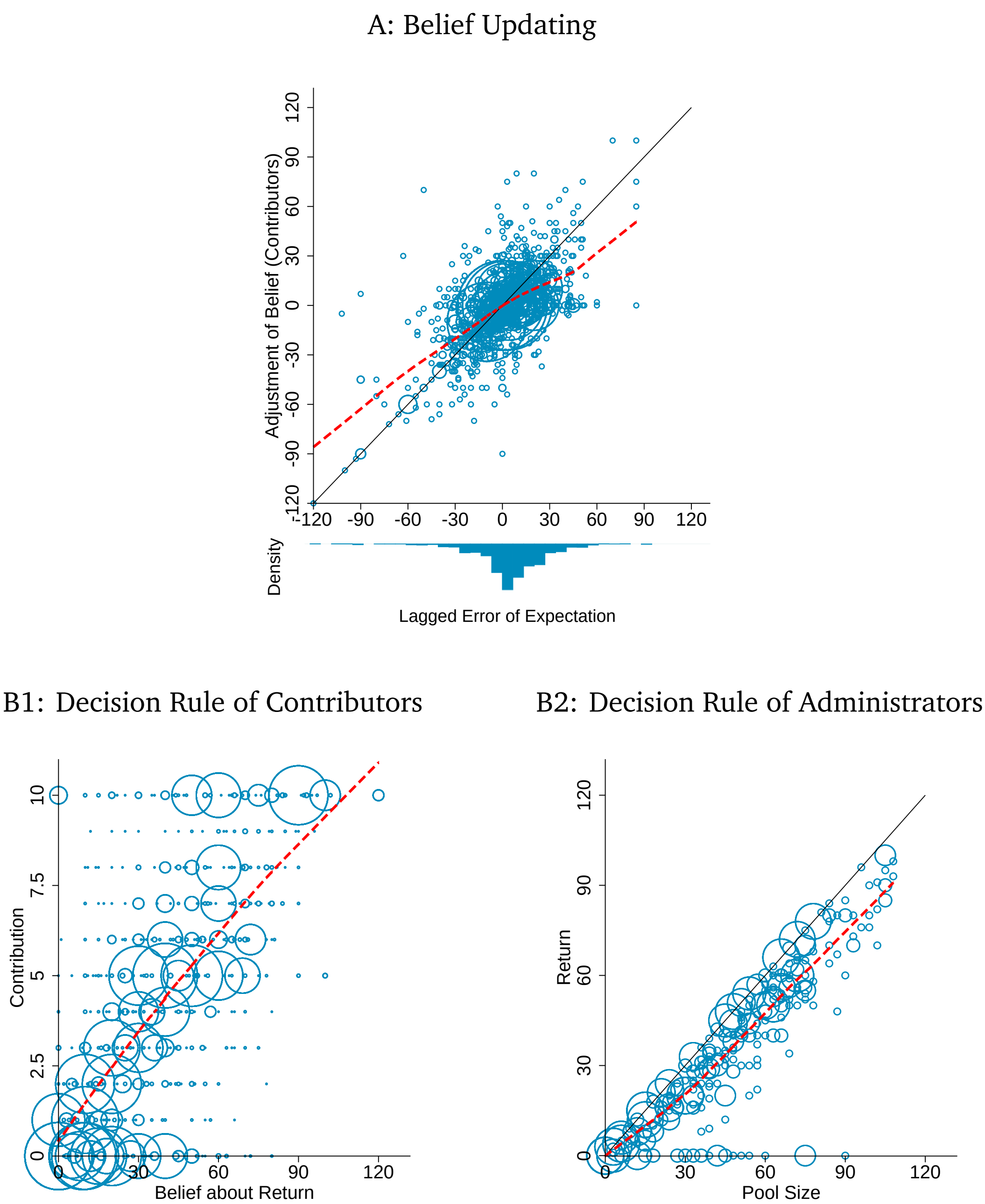
Our framework fully describes the dynamic interactions between the administrator and the contributors: while (3) and (4) represent agents' decisions in the stage game, (5) models the belief-updating mechanism that introduces dynamics into the system. The administrator conditions her behaviour on the observed behaviour of the group of contributors. In contrast, the contributors choose contributions based on their beliefs about the public goods return and hence indirectly condition their behaviour on the behaviour of all the other agents in the previous periods. On an aggregated level, the decision rules and the belief-updating mechanism imply that the group-level cooperation and the administrator's trustworthiness are mutually dependent processes. Importantly, because of this conditionality in behaviour, disturbances can erode the established levels of public goods provision over time.

3.2.2 Decisions and Belief Updating: Descriptive Analysis

Because dynamic behavioural reactions to disturbances require some conditionality in behaviour, our first step is to determine descriptively whether the administrator's and the contributors' decisions are mutually dependent. More precisely, we descriptively examine decision rules and belief formation before studying them in the integrated framework of the PVAR model. This analysis is directly linked to the conceptual discussion in Subsection 3.2.1.

Figure 2 collects the results of the descriptive analysis. Panel A refers to contributors' learning rule. The lower part of Panel A displays a histogram of the lagged error of expectation (the difference between the lagged return and the lagged belief). The histogram shows that the distribution is fairly symmetric around zero, suggesting that contributors' mean beliefs match the mean return well. Given the length of the interaction and the stability of cooperation and trustworthiness, this observation is not surprising. Assuming a linear version of the belief-updating equation (5), the upper part of Panel A plots contributors' belief adjustments (the difference between the current and the lagged belief about the return rate) against the lagged error of expectation. The data show a strong positive association. Moreover, the scatter plot is centered on the origin: contributors do not adjust their beliefs if they predicted the return correctly in the previous period. The fitted values (red dashed line) reveal that the relationship between the belief adjustments and the lagged error of expectation is close to linear, with a slope slightly smaller

Figure 2: Decision Rules and Belief Updating



Notes: The Figure shows scatter plots and fitted values (red dashed lines) for belief-updating (Panel A) and decision rules (Panels B1 and B2). The upper part of Panel A plots contributors' adjustment of beliefs (differences between current and lagged beliefs) against the lagged error of expectation (differences between the lagged return rates and the lagged beliefs). The lower part of Panel A shows the distribution of errors of expectation. Panel B1 plots individual contributions against individual beliefs about the return. Panel B2 plots administrators' returns against group mean contributions. We derive fitted values from non-parametric smoothing with locally weighted regressions. Hence, the fitted values follow the data and are not restricted to a globally linear relationship.

than one.[18] This suggests that contributors form beliefs consistent with a simple linear adaptive belief-updating rule.

We complete the descriptive evidence by studying decision rules. Panel B1 illustrates

[18] We derive all the fitted values in Figure 2 by non-parametric smoothing with locally weighted regressions.

the mapping of contributors' individual contributions to beliefs. The scatter plot shows a strong correlation between contributions and beliefs about the return rate. Again, the fitted values (red dashed line) suggest a linear relationship. Turning to the administrator's decision rule, Panel B2 shows a scatter plot of the return R_t against the pool M_t. The plot reveals that, as suggested by decision rule (4), administrators' choices are strongly correlated with same-period contributions. The plot also demonstrates that, although administrators keep part of the pool for themselves in most cases, they tend to make a large share of the pool available to contributors. Taken together, the descriptive analysis supports all three parts of our simple learning model.

3.2.3 Identifying and Estimating Dynamic Interactions

This section introduces a simple time series model to identify the effects of one-time changes (i.e., disturbances or shocks) in cooperation and rent extraction on behaviour in the PTG. While our main interest lies in understanding the impact of rent-extraction shocks on cooperation, we discuss both types of disturbances. This lays the ground for a comprehensive analysis of the dynamics in the PTG that, inter alia, allows us to determine the relative importance of either type of shock.

Econometric Model We use the conceptual framework outlined in Subsection 3.2.1 to derive a panel vector autoregressive model that captures the dynamic interaction between the contributors and the administrator in the PTG. We assume that $f(\cdot)$, $g(\cdot)$, and $h(\cdot)$ are additively separable and linear in all the arguments. Furthermore, as suggested by the descriptive analysis, we consider contributors who have standard adaptive expectations. This leads to the following set of equations:

$$m_{nit} = \alpha_1 \theta_{ni} + \alpha_2 \hat{R}_{nit} + s_{nit} \tag{6}$$

$$R_{nt} = \beta_1 \phi_n + \beta_2 M_{nt} + v_{nt} \tag{7}$$

$$\hat{R}_{nit} = \varphi_1 \theta_{ni} + \hat{R}_{nit-1} + \varphi_2 (R_{nt-1} - \hat{R}_{nit-1}), \tag{8}$$

where (6) is the decision rule of group n's contributor i, (7) is the decision rule of group n's administrator, and (8) describes contributor i's belief-updating rule in period t. Importantly, the equations reflect the sequential decision structure of the game. Because contributors move first, contributor i's decision depends on her belief about the return rather than the actual return in period t. In contrast, as the second mover, the administrator conditions her decision on the actual pool of contributions in period t.

From equations (6) to (8), it is straightforward to derive a structural form panel vector autoregressive model (see the Appendix)

$$\overline{M}_{nt} = \overline{\tau}_n + \rho_1 \overline{M}_{nt-1} + \rho_2 \overline{R}_{nt-1} + \overline{u}_{nt} \tag{9}$$

$$\overline{R}_{nt} = \overline{\varrho}_n + \rho_3 \overline{M}_{nt-1} + \rho_4 \overline{R}_{nt-1} + \beta_2 \overline{u}_{nt} + \overline{v}_{nt}, \tag{10}$$

where $\overline{M}_{nt} = \frac{5}{6} M_{nt}$ is the contribution rate at the group-level, $\overline{R}_{nt} = \frac{5}{6} R_{nt}$ is the return rate, $\overline{\tau}_n$ and $\overline{\varrho}_n$ are group fixed effects, and $\overline{u}_{nt}$ and $\overline{v}_{nt}$ are error terms. The parameters ρ_1 to ρ_4 capture the dynamics, while β_2 measures the contemporaneous impact of $\overline{M}_{nt}$ on $\overline{R}_{nt}$. The absence of a contemporaneous effect of the return rate on the contribution rate follows directly from the fact that administrators decide after contributors in the same period.

In a nutshell, we model the contribution and the return rates as crossed processes and therefore allow for (intertemporal) relations in both directions: cooperation can affect rent extraction and vice versa. Furthermore, the PVAR model decomposes the variation in the group-level contribution rate $\overline{M}_{nt}$ and the return rate $\overline{R}_{nt}$ into deterministic (non-random) components that explain variation by past realisations of $\overline{M}_{nt}$ and $\overline{R}_{nt}$ and exogenous (random) components captured by the error terms.[19] Our conceptual framework suggests interpreting the error terms as aggregated random errors in decision-making conditional on past decisions. Hence, $\overline{u}_{nt}$ and $\overline{v}_{nt}$ capture (positive or negative) disruptions in cooperation and rent extraction, respectively. In our empirical analysis, we use this random variation to identify the effects of exogenous one-time behavioural changes on future levels of cooperation and rent extraction.

To visualise the disruptions, Figure A1 in the Online Appendix depicts the realisations of shocks for one of the groups, while Figure A2 shows the distributions of shocks for all the groups. We note that the distributions of both $\overline{u}_{nt}$ and $\overline{v}_{nt}$ are fairly symmetric. We also note that the PVAR model is consistent with the assumed belief-updating process (see Subsection 3.2.1). In fact, we can recover the updating parameter φ_2 from our estimates.

Identification and Estimation of Model Parameters The error terms only represent exogenous changes in the corresponding decisions if they are uncorrelated across equations. We follow Sims (1980) and assume uncorrelatedness by imposing restrictions on the contemporaneous relationships between $\overline{M}_{nt}$ and $\overline{R}_{nt}$. In particular, as can be seen in Eq. (9), we restrict the effect of $\overline{R}_{nt}$ on $\overline{M}_{nt}$ to zero such that the return rate has no contemporaneous effect on the contribution rate. Importantly, this restriction is known to be valid, as it follows directly from the sequence of decision-making in our experimental design: because the administrator decides on the return after contributors have made their decision, the return cannot affect the contribution behaviour in the same period.

[19] Estimating a PVAR(1) model minimises the Schwarz Bayesian information criterion.

While the PVAR model is a standard tool in time series applications, we are not aware of any previous attempts to apply it to a context in which an experimental design guarantees the validity of the key identifying assumption concerning the contemporaneous effects. All the assumptions that we need to make in addition to a treatment-based identification strategy can be tested (stationarity, number of lags) or are frequently used in econometric models (additivity and linearity).

We recover the PVAR parameters by estimating the reduced-form PVAR model with a least square dummy variable estimator (LSDV)[20] and then computing the Cholesky factorisation of the reduced-form PVAR variance-covariance matrix of the residuals.[21] Using the estimated PVAR coefficients, it is straightforward to recover the remaining structural parameters in the system of equations (6) to (8) as:

$$\hat{\varphi}_2 = 1 - \hat{\rho}_1$$
$$\hat{\alpha}_2 = \frac{\hat{\rho}_2}{4r\hat{\varphi}_2}.$$

3.2.4 Dynamic Interactions: Evidence

Structural Parameters We find that the structural parameters of the decision-rule equations $\hat{\alpha}_2 = 0.052$ and $\hat{\beta}_2 = 0.820$ are positive and significantly different from zero (p-values < 0.001; Delta method). This reveals positive conditionality in the behaviour of administrators and contributors. Our estimate of $\hat{\varphi}_2 = 0.911$ is not significantly different from one (p-value $= 0.227$). The structural approach hence confirms the strong association between past expectations errors and belief adjustments. In fact, the estimate of $\hat{\varphi}_2$ implies that contributors fully incorporate their expectation error from the previous period when updating their beliefs.

Impulse Responses To visualise the contributors' and the administrator's behaviour in response to one-time disturbances, we utilise *impulse response functions* (IRFs). Building on the PVAR model estimates, an IRF simulates how contribution rates and return rates evolve after an exogenous disturbance in either the contribution rate or the return rate. The disturbances represent quasi-treatments and capture changes in the current value of one of the errors. For ease of exposition, our IRFs evaluate positive shocks (i.e., increases in the contribution rate and return rate). For negative shocks, the magnitude of responses is identical; only the signs of the effects change. We evaluate the significance of impulse responses using 95 per cent confidence intervals based on a double-bootstrapping resampling scheme with 10,000 repetitions.

[20]Because of the length of our panel (24 periods after excluding start-game and end-game periods), the Nickel bias (Nickell 1981) is a minor concern.

[21]Cagala and Glogowsky (2014) provide a Stata code to estimate panel vector autoregressive models.

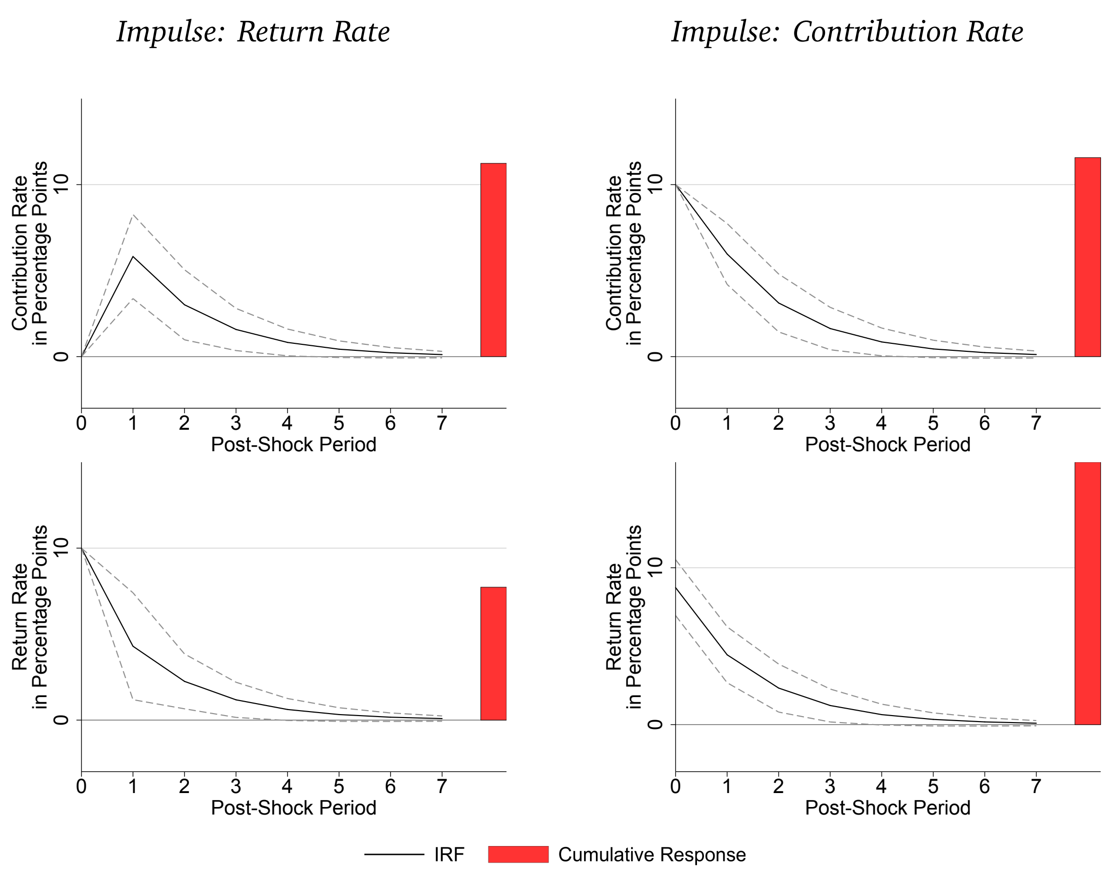
Figure 3 shows the IRFs for an impulse in the return rate (left-hand panels) and for an impulse in the contribution rate (right-hand panels). The upper panels show the responses of the contribution rate, while the lower panels depict the reactions of the return rate. Studying the IRFs is instructive, as they highlight how cooperation and rent-extraction shocks affect the behaviour of contributors and administrators, respectively. To show this, let us consider the left-hand panels, depicting the responses to a 10-percentage-point impulse in the return rate in step 0. As the contributors decide before the administrator, the contemporaneous response in the contribution rate is zero (upper panel). From step one onwards, the responses encompass not only the effect of the initial decision but also the indirect effects working through the feedback effects in the system. In step 1, the contributors respond to the return-rate impulse by raising their contribution rate by 5.8 percentage points relative to its initial value. The initial impulse and the resulting increase in the contribution rate trigger further behavioural responses. For example, the return rate increases by 4.3 percentage points in step 1 (lower panel). However, the increase in the return rate in step 1 is lower than the initial shock in step 0, leading to a lower increase in the contribution rate in step 2 in comparison to the increase in step 1. As a result, cooperation and rent extraction converge back to pre-shock values after a few periods.

Turning to a 10-percentage-point impulse in the contribution rate (right-hand panels), we note a contemporaneous increase in the return rate by 8.7 percentage points (lower panel). In step 1, the rise in the return rate, together with the initial impulse, increases the contribution rate by 6.0 percentage points relative to its initial value. We note that the responses to a return-rate impulse are significantly different from zero (5% level) for four after-shock periods. Similarly, we find that a shock to the contribution rate has significant effects on the contribution rate (return rate) for four (three) subsequent periods.

From the IRFs, we can derive the overall responses as the cumulative effects of a given impulse (depicted as red bars in Figure 3). These overall responses cumulate the contemporaneous and future responses of a given impulse, including all the feedback effects, and can naturally be labelled “behavioural multipliers”.[22] For return and contribution-rate shocks, we find substantial behavioural multipliers. For instance, a 10-percentage-point rise in the return rate causes an 11.2 percentage point increase in the contribution rate relative to its initial level (upper left-hand panel). Similarly, a one-time increase in the contribution rate of 10 percentage points adds 11.6 percentage points to the contribution rate once all the indirect effects are taken into account (upper right-

[22] We calculate the multipliers by cumulating all the post-impulse responses that are statistically significant at the 5% level. We exclude the impulse in the calculation of the overall responses. The overall response of the return rate includes the contemporaneous response (the response of trustworthiness to same-period cooperation). Calculating behavioural multipliers relates our study to the empirical social multiplier literature (e.g., Glaeser and Scheinkman 2000; Glaeser *et al.* 2003).

Figure 3: Impulse Responses and Behavioural Multipliers



Notes: The Figure shows IRFs (solid line) with 95 % confidence bands (dashed lines) and cumulative responses (red bars). The cumulative response is the sum over significant post-impulse return (contribution) rates. The first-row (second-row) IRFs show the response of the contribution rate (return rate) to standardised 10-percentage-point impulses. Left hand side (right hand side) panels show return- (contribution-) rate impulses.

hand panel). With respect to the responses in the return rate, we find that a one-time change in the group-level contributions triggers an overall impact on the return rate of 16.4 percentage points (lower right-hand panel).[23]

The previous analysis establishes two main findings. First, the IRFs document that disturbances trigger significant short-run adjustments in the behaviour of both types of players. The resulting dynamics of the PTG amplify the direct effects of disturbances: in the short run, more cooperation by the administrator (i.e., less rent extraction) triggers more cooperation among the contributors. Conversely, higher contributions induce more cooperative behaviour of the administrator. Hence, trustworthiness breeds cooperation and vice versa.[24] This finding connects our paper to the literature on asymmetric public goods games (e.g., Fisher *et al.* 1995; Palfrey and Prisbrey 1997; Cherry *et al.* 2005;

[23]The IRFs also allow us to determine the overall impact of an impulse on the payoffs. Evaluating the same impulses as before, we find that contributors suffer a net loss in the case of an individual one-time increase in cooperation. The same holds true for administrators: a one-time increase in the return rate leads to a net payoff loss. This confirms the interpretation of impulses as random errors in decision-making that lead to deviations from the best responses.

[24]The fact that administrators respond to cooperation shocks relates our work to Berg *et al.* (1995), who study the trust game and show that trust breeds trustworthiness if the subjects can observe past behaviour (Berg *et al.* 1995).

Sadrieh and Verbon 2006; Reuben and Riedl 2013; Charness *et al.* 2014; Reuben and Riedl 2013); it implies that conditional cooperation, as documented by Fischbacher *et al.* (2001) and others, extends to cases in which one type of player interferes with public goods provision.

Second, the IRFs show that all the behavioural responses to disturbances are of a temporary nature and fade out over time. Once the system has fully adjusted to a given disturbance, both types of players behave as they did before the shock. Hence, one-time rent-extraction shocks do not cause permanent changes in individuals' attitudes towards cooperativeness. Combining both results, the analysis reveals that cooperation and rent extraction are mutually dependent but do not reinforce each other. This explains why the presence of a rent-extracting administrator leaves the established level of cooperation unaffected and does not lead to a vicious circle of higher rent extraction and lower contributions. We can summarise our results:

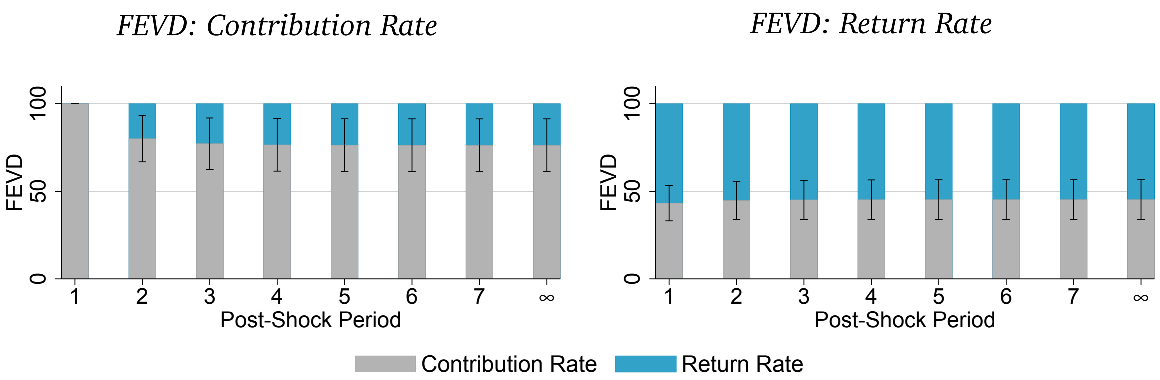
RESULT 2a: *After cooperation and rent-extraction shocks, both types of players interact in a way that amplifies the direct effects of shocks. The cumulative responses indicate that conditional cooperation extends to cases in which players have different roles: trustworthiness breeds cooperation and vice versa.*

RESULT 2b: *In the PTG, all responses to cooperation and rent-extraction shocks are of a temporary nature: shocks trigger transitory changes but do not have permanent effects. This implies that one-time changes in rent extraction do not erode the established level of cooperation.*

Forecast Error Variance Decompositions In the following, we analyse the relative importance of impulses in cooperation and trustworthiness to explaining the observed variation in both behavioural dimensions. The question that we ask is whether the overall variation in cooperation and trustworthiness, and hence in public goods provision, stems mainly from disturbances in contribution rates or disturbances in return rates (or both).

We answer this question using *forecast error variance decompositions* (FEVDs), a standard tool in multiple time series analysis; see Lütkepohl (2007) for an introduction. A FEVD shows the fraction of the forecast error variance that is due to different impulses in a specific post-shock period. If the horizon tends to infinity, a FEVD specifies the fraction of the variance in the dependent variable that the different impulses explain. To enable valid comparisons, we consider standardised 10-percentage-point contribution-rate and return-rate impulses. This allows us to evaluate the relative importance of one-time changes to the variation while holding the magnitude of the impulses constant. A FEVD

Figure 4: Decomposing the Variation in Cooperation and Trustworthiness



Notes: The Figure shows FEVDs (grey and blue bars) with 95% confidence bands (spikes) for contribution rates (left panel) and return rates (right panel).

thus quantifies the importance of a standardised impulse in the contribution rate relative to a standardised impulse in the return rate to the explanation of the total variance.

Figure 4 shows the FEVDs. Two key observations emerge. First, for standardised impulses in the contribution and return rate, the variation in contributions is mainly explained by impulses in contributors' behaviour (left panel): in the long run, 76.3% of the variance in the contribution rate is explained by contribution impulses. Second, the forecast error variance decomposition is more symmetric for the return rate (right panel): standardised return-rate impulses explain 54.8% of the long-run variation in the return rate, while contribution-rate impulses explain 45.2%.[25] Our third result is:

RESULT 3: *The variation in contributions is mainly explained by disturbances in contributors' behaviour. Hence, contributors are rather influenced by cooperation than by rent-extraction shocks. In contrast, disturbances in contributors' and administrators' behaviour have similar weights in explaining the variation in the return rate.*

The importance of Result 3 becomes clear once we link it to Result 1, which states that, beyond its effect working through efficiency, rent extraction leaves mean contributions unaffected. Result 3 shows that this finding extends from mean contributions to the *volatility* of the contribution behaviour. Both findings together imply that the presence of a rent-extracting administrator does not fundamentally affect the decisions of contributors.

[25] We find similar results when we account for the empirical impulse size using residual mean squared error (RMSE) impulses (interpretable as sample impulses) instead of 10-percentage-point impulses. We note that FEVDs evaluate standardised shocks; that is, the direct effect of shocks on the payoff of contributors does not differ between contribution-rate shocks and return-rate shocks. This implies that our FEVD results are not likely to be driven by our design choice to have several contributors interacting with one administrator (and the corresponding larger scope for contributors to induce variation in cooperation and trustworthiness).

3.2.5 Impact of Attitudes Towards Cooperation and Trust

In the previous sections, we have explored the short-run dynamics of the PTG. One of the key insights from this analysis is that cooperation is resilient to disturbances in the administrator's behaviour: after the contributors experience a rent-extraction shock, their contributions converge back to pre-shock levels. Despite the finding that rent-extraction shocks do not have permanent effects, understanding their impact on public goods provision is nevertheless essential. That is because unexpected changes in rent extraction make the provision of public goods less reliable in the short run. The final step in our analysis of the short-run dynamics is therefore to examine whether specific traits of contributors amplify the impact of rent-extraction and cooperation shocks. If such traits exist, groups composed of contributors carrying those characteristics suffer from a less reliable provision of public goods. Subsequently, we focus on traits that are closely related to our outcomes: the contributors' attitudes towards cooperation and trust. Using sample splits, we investigate the heterogeneity in IRFs and FEVDs regarding these attitudes.

We measure the subjects' attitudes by means of a survey that we conducted two weeks after the experiment. The time lag between the experiment and the survey attenuates any potential impact of the subjects' experience in the experiment on their survey responses. The survey includes standard measures of attitudes from the World Values Survey. We elicit generalised trust with the following question: "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?". The response options are: "most people can be trusted", "cannot be too careful", and "I don't know". To measure cooperative attitudes, we ask subjects to indicate on a scale from one to 10 whether different forms of free-riding or non-cooperative behaviour "can always be justified" or "never be justified". We then construct an internally consistent index using the items "cheating on taxes if you have a chance" and "not paying the fare in public transport" (Cronbach's alpha 0.754) and rescale the index to lie between zero (fully non-cooperative) and 100 (fully cooperative).

We split the sample as follows. For each of the groups, we average over the contributors' stated attitudes. Both for trust and cooperative attitudes, we then contrast groups of contributors in the lowest tertile on the attitudes scale with groups of contributors in the highest tertile. The analysis thus distinguishes between groups with trusting and non-trusting contributors and groups with cooperative and non-cooperative contributors, respectively. As we focus on the behaviour of contributors, we do not consider the heterogeneity in terms of the administrators' attitudes. For completeness, we report the corresponding results in the Online Appendix.

The splitting procedure results in a pronounced between-subsample heterogeneity in the contributors' attitudes.[26] In groups of non-trusting contributors (lowest tertile), only 27.8% of the contributors state that they generally trust in others. This contrasts to a value of 83.3% for the trusting contributors (highest tertile). Regarding attitudes towards cooperation, the heterogeneity is similarly pronounced: the groups of non-cooperative contributors (lowest tertile) report a mean index value of 50.0, while the cooperative types (highest tertile) report a mean of 74.2.

Subsequently, we discuss the findings for the two sample splits.[27] Figure 5 shows the IRFs and the behavioural multipliers for the non-trusting contributors (Panel A) and the trusting contributors (Panel B). Again, we differentiate between contribution and return-rate impulses. However, because we consider the heterogeneity among the contributors, we focus on responses of the contribution rate. We report the IRFs and the FEVDs for the return rate in the Online Appendix (see Figures A3, A4, and A5). Figure 5 reveals a distinct heterogeneity in the impulse response functions: for the groups with the non-trusting contributors, the responses to contribution and to return-rate shocks are stronger and more persistent. This implies that cooperation is more resilient to one-time changes in the groups with the more trusting contributors; that is, the convergence to the initial cooperation levels is faster. The difference in the IRFs translates into behavioural multipliers that are more than three times larger for the non-trusting types than for the trusting types.

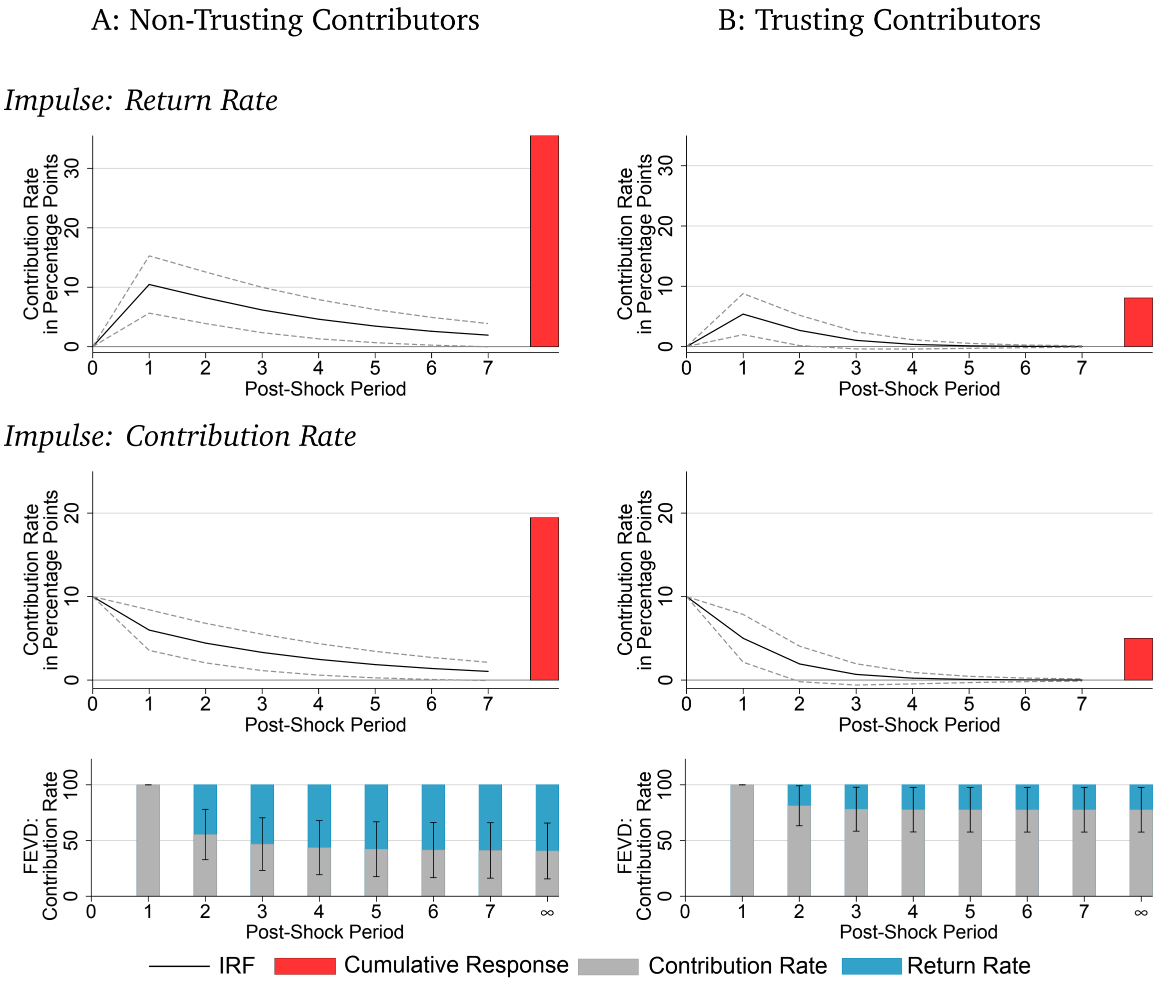
The FEVDs for the contribution rate, displayed in the lower part of the figure, reinforce the interpretation that the heterogeneity in the IRFs and the overall responses is, in fact, related to the differences in the contributors' trust: comparing the FEVDs of the trusting and the non-trusting contributors reveals that the share of the variation in the contribution rate that is driven by impulses in the return rate is much larger for the non-trusting than for the trusting contributors.

We complete the analysis of the heterogeneous responses by presenting evidence on the heterogeneity in the contributors' cooperative attitudes. Figure 6 shows the results for the non-cooperative (Panel A) and the cooperative groups (Panel B). Again, the figure reveals a pronounced difference in the behavioural responses: in the groups in which contributors are non-cooperative, the contribution-rate and the return-rate shocks trig-

[26] All subsamples comprise six groups. Due to ties, a simple split by tertiles led to an unequal number of groups between subsamples. To split the sample symmetrically, we used responses to additional survey questions (trust in strangers, acceptance of claiming and receiving social benefits to which one is not entitled, acceptance of taking bribes) to break the ties.

[27] We ran a further robustness check to insure us against the concern that differences between subsamples could be driven by different average contribution rates. If, for example, non-cooperative contributors had a lower average contribution rate, there would be less scope for downward adjustments of contributions after a negative return rate shock, resulting in a reduced multiplier. To run the test, we repeated the subsample analysis after dividing the behavioural multipliers by the average contribution rates in the respective subsample. We find that the heterogeneity in rescaled multipliers is very similar to the heterogeneity without rescaling. We also confirmed that in all subsamples, the distributions of shocks are symmetric.

Figure 5: Heterogenous Responses – Trusting vs. Non-Trusting Contributors

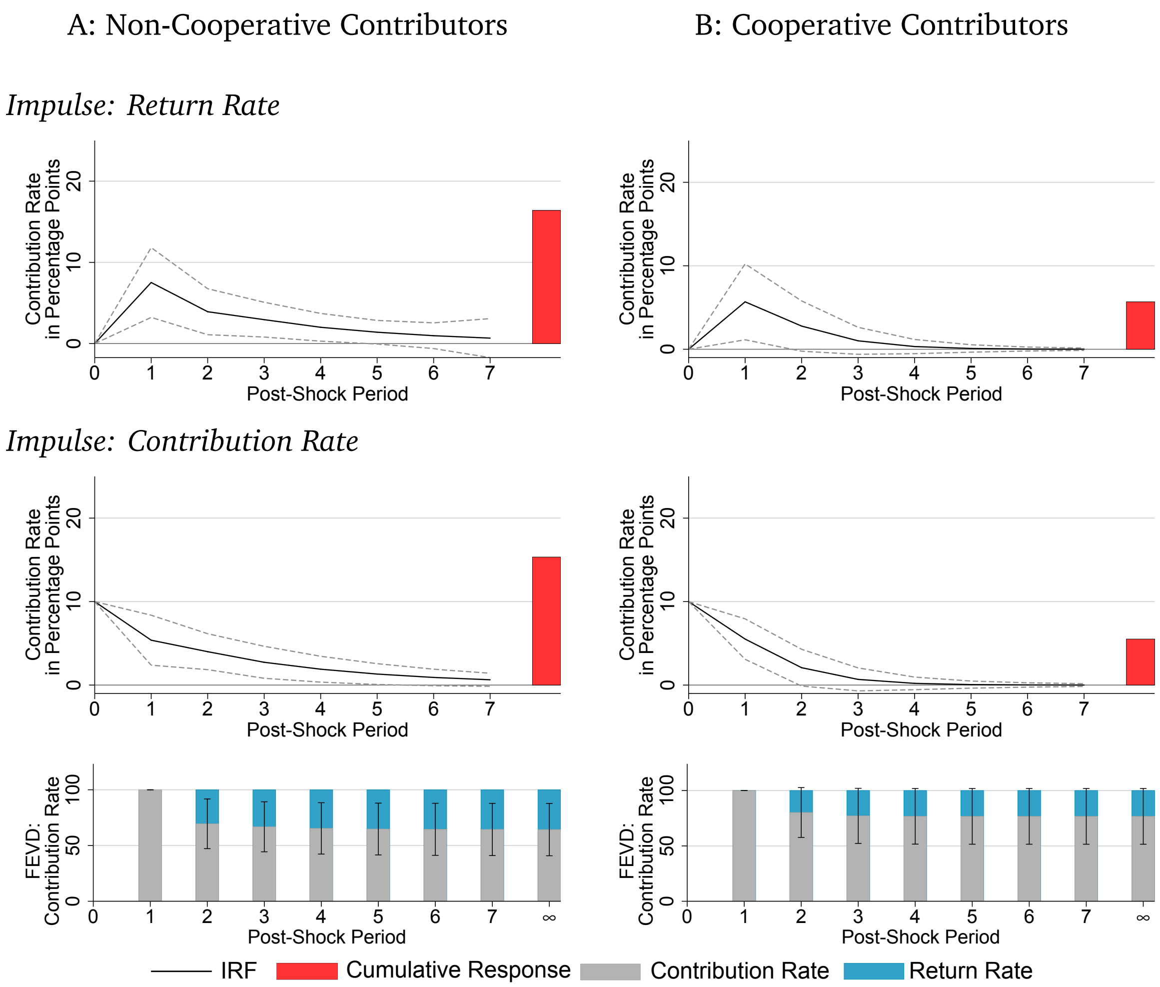


Notes: The first- and second-row panels show IRFs (solid line) with 95% confidence bands (dashed lines) and cumulative responses (red bars). The third-row panels display FEVDs (grey and blue bars) with 95% confidence bands (spikes) for contributors. The cumulative response is the sum over significant post-impulse contribution rates. The first-row (second-row) IRFs are for standardised 10-percentage-point return-rate (contribution-rate) impulses. The left-hand (right-hand) panels are for non-trusting (trusting) types. The classification into types is based on survey responses to a question on general trust. The groups in the lowest (highest) tertile of the distribution of the share of contributors who generally trust in others are classified as non-trusting (trusting) type groups.

ger more persistent responses and lead to much stronger behavioural multipliers. Taken together, the Figures 5 and 6 suggest that trust and cooperative attitudes protect cooperation against disruptions. Our final results is:

RESULT 4: *The behavioural multipliers are much larger if the contributors are less cooperative and less trusting than if they are more cooperative and more trusting. Hence, cooperative attitudes and trust are traits that explain permanent differences in the short-run volatility of public goods provision.*

Figure 6: Heterogenous Responses – Cooperative vs. Non-Cooperative Contributors



Notes: The first- and second-row panels show IRFs (solid line) with 95% confidence bands (dashed lines) and cumulative responses (red bars). The third-row panels display FEVDs (grey and blue bars) with 95% confidence bands (spikes) for contributors. The cumulative response is the sum over significant post-impulse contribution rates. The first-row (second-row) IRFs are for standardised 10-percentage-point return-rate (contribution-rate) impulses. The left-hand (right-hand) panels are for non-cooperative (cooperative) types. The classification into types is based on survey responses to questions regarding the acceptance of free-riding. The groups in the highest (lowest) tertile of the distribution of the cooperativeness index are classified as non-cooperative (cooperative) type groups.

4 Conclusion

Public goods provision often involves groups of individuals repeatedly interacting with administrators, who can extract private rents from the pool of contributions. Our focus is on understanding how the presence of a rent-extracting administrator affects the level and the dynamics of public goods provision. We analyse a repeated game that blends the trust game and the public goods game. In the resulting public trust game, an administrator decides how much of the pool of contributions to return to the group of contributors and how much to keep for herself.

Our main result is that the presence of a rent-extracting administrator reduces contributions but only because rent extraction reduces the efficiency of public goods provision.

Comparing the public trust game with a public goods game that exogenously decreases efficiency such that it matches the reduced efficiency level in the public trust game, we do not find any difference in the average contributions. We also show that cooperation in the public trust game is stable over time. Hence, contrary to the intuition that rent-extraction triggers a vicious circle of lower cooperation and higher rent extraction, the presence of a rent-extracting administrator does not erode the established level of cooperation.

In the second step of our analysis, we shift our focus to the social interactions between the administrator and the contributors. In particular, we employ a panel vector autoregressive model and explore how rent-extraction shocks (i.e., exogenous increases in rent extraction) and cooperation shocks (i.e., exogenous decreases in cooperation) affect the subsequent decisions of the contributors and the administrator. Understanding the impact of such shocks is crucial for explaining why cooperation is so resilient.

The analysis of the interactions generates the following insights. First, we demonstrate that contributors who experience an exogenous increase in rent-extraction reduce their contributions. More generally, we show that conditional cooperation extends to cases in which players have different roles: trustworthiness of the administrator breeds cooperation by contributors and vice versa. However, all responses to shocks in the behaviour of the administrator and the contributors are of a temporary nature and do not have permanent effects. Hence, public goods provision eventually converges back to the pre-shock values. The main point to take-away from this analysis is thus that cooperation and rent extraction are mutually dependent but do not reinforce each other. These findings explain why cooperation is so resilient. Second, our analysis also shows that the variation in the contribution behaviour is hardly explained by disturbances in the administrator's behaviour. We conclude that not only the level but also the volatility of cooperation is resilient to rent extraction. Taking all the findings on the level and the dynamics of public goods provision together, the main insight from our paper is that, apart from triggering short-run response and affecting contributions through efficiency, the presence of a rent-extracting administrator does not matter for cooperative behaviour.

A final result refers to the short-term volatility of public goods provision. We demonstrate that in groups composed of contributors with more cooperative attitudes and higher trust, rent-extraction shocks induce less short-term volatility in cooperation. As a result, groups where contributors carry those characteristics benefit from a more reliable provision of public goods.

Our results apply to a broad range of settings in which the provision of public goods depends on the behaviour of potentially corrupt agents, who act at a higher hierarchical level. A natural application is tax-financed public goods provision when taxpayers interact with a rent-extracting bureaucracy. In this setting, the level of public goods provided depends not only on taxpayers' compliance with the tax law but also on the behaviour of bureaucrats. Taken the behaviour in the PTG at face value, we predict that bureaucratic corruption reduces tax compliance through its effect on the efficiency of public goods provision. Moreover, we expect that rent-extraction shocks significantly reduce taxpayers' compliance in the short run but do not trigger permanent changes in tax compliance. Concerning the short-run responses to rent-extraction shocks, our results also suggest that the detrimental effects on tax compliance should be larger in communities with lower trust and less cooperative attitudes.

References

ABBINK, K., IRLENBUSCH, B. and RENNER, E. (2000). The moonlighting game: An experimental study on reciprocity and retribution. *Journal of Economic Behavior & Organization*, **42** (2), 265–277.

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

ANDREONI, J. (1988). Why free ride? Strategies and learning in public goods experiments. *Journal of Public Economics*, **37** (3), 291–304.

—, HARBAUGH, W. and VESTERLUND, L. (2003). The carrot or the stick: Rewards, punishments, and cooperation. *American Economic Review*, **93** (3), 893–902.

ARMANTIER, O. and BOLY, A. (2008). *Can corruption be studied in the lab? Comparing a field and a lab experiment*. CIRANO Working Papers 2008s-26, CIRANO.

BALDASSARRI, D. and GROSSMAN, G. (2011). Centralized sanctioning and legitimate authority promote cooperation in humans. *Proceedings of the National Academy of Sciences*, **108** (27), 11023–11027.

BARR, A., BURNS, J., MILLER, L. and SHAW, I. (2015). Economic status and acknowledgement of earned entitlement. *Journal of Economic Behavior & Organization*, **118** (C), 55–68.

BERG, J., JOHN, D. and KEVIN, M. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, **10** (1), 122–142.

BLUME, L., BROCK, W., DURLAUF, S. and IOANNIDES, Y. (2011). Identification of social interactions. *Handbook of Social Economics*, vol. 1, North-Holland, pp. 853 – 964.

BUCKLEY, E. and CROSON, R. (2006). Income and wealth heterogeneity in the voluntary provision of linear public goods. *Journal of Public Economics*, **90** (4-5), 935–955.

CAGALA, T. and GLOGOWSKY, U. (2014). *Panel vector autoregressions for Stata (xtvar)*. Software package available at https://ideas.repec.org/c/boc/bocode/s457944.html.

—, —, GRIMM, V., RINCKE, J. and CUEVA, A. T. (2017). Does corruption affect the private provision of public goods?, mimeo.

CAPPELEN, A., HOLE, A., SORENSEN, E. and TUNGODDEN, B. (2007). The pluralism of fairness ideals: An experimental approach. *American Economic Review*, **97** (3), 818–827.

CARPENTER, J. (2007). Punishing free-riders: How group size affects mutual monitoring and the provision of public goods. *Games and Economic Behavior*, **60** (1), 31–51.

CASSAR, A. and RIGDON, M. (2011). Trust and trustworthiness in networked exchange. *Games and Economic Behavior*, **71** (2), 282–303.

CESARINI, D., DAWES, C., JOHANNESSON, M., LICHTENSTEIN, P. and WALLACE, B. (2009). Genetic variation in preferences for giving and risk taking. *The Quarterly Journal of Economics*, **124** (2), 809–842.

CHARNESS, G., COBO-REYES, R. and JIMÉNEZ, N. (2014). Identities, selection, and contributions in a public-goods game. *Games and Economic Behavior*, **87** (C), 322–338.

CHERRY, T. L., KROLL, S. and SHOGREN, J. F. (2005). The impact of endowment heterogeneity and origin on public good contributions: Evidence from the lab. *Journal of Economic Behavior & Organization*, **57** (3), 357–365.

COSTA-GOMES, M., CRAWFORD, V. P. and BROSETA, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, **69**, 1193–1235.

COX, C. (2015). Decomposing the effects of negative framing in linear public goods games. *Economics Letters*, **126** (C), 63–65.

COX, J., OSTROM, E., SADIRAJ, V. and WALKER, J. (2013). Provision versus appropriation in symmetric and asymmetric social dilemmas. *Southern Economic Journal*, **79** (3), 496–512.

DAL BO, P. and FRECHETTE, G. R. (2011). The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review*, **101** (1), 411–29.

DASGUPTA, U. (2011). Do procedures matter in fairness allocations? Experimental evidence in mixed gender pairings. *Economics Bulletin*, **31** (1), 820–829.

DUFWENBERG, M., GÄCHTER, S. and HENNIG-SCHMIDT, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior*, **73** (2), 459–478.

EL-GAMAL, M. A. and GRETHER, D. M. (1995). Are people bayesian? Uncovering behavioral strategies. *Journal of the American Statistical Association*, **90**, 1137–1145.

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

— and GÄCHTER, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, **100** (1), 541–56.

—, — and FEHR, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, **71** (3), 397–404.

FISHER, J., ISAAC, R. M., SCHATZBERG, J. W. and WALKER, J. M. (1995). Heterogenous demand for public goods: Behavior in the voluntary contributions mechanism. *Public Choice*, **85** (3-4), 249–266.

FRANK, B. and SCHULZE, G. G. (2000). Does economics make citizens corrupt? *Journal of Economic Behavior & Organization*, **43** (1), 101–113.

FUDENBERG, D. and LEVINE, D. (1998). *The theory of learning in games*. The MIT Press.

GÄCHTER, S. and FEHR, E. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, **90** (4), 980–994.

GÄCHTER, S., HERRMANN, B. and THONI, C. (2004). Trust, voluntary cooperation, and socio-economic background: Survey and experimental evidence. *Journal of Economic Behavior & Organization*, **55** (4), 505–531.

— and RENNER, E. (2010). The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics*, **13** (3), 364–377.

GÄCHTER, S., RENNER, E. and SEFTON, M. (2008). The long-run benefits of punishment. *Science*, **322** (5907), 1510.

GLAESER, E., SACERDOTE, B. and SCHEINKMAN, J. (2003). The social multiplier. *Journal of the European Economic Association*, **1** (2-3), 345–353.

— and SCHEINKMAN, J. (2000). *Non-market interactions*. NBER Working Papers 8053, National Bureau of Economic Research, Inc.

GREINER, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, **1** (1), 114–125.

HERRMANN, B. and THÖNI, C. (2009). Measuring conditional cooperation: A replication study in russia. *Experimental Economics*, **12** (1), 87–92.

HO, T.-H., CAMERER, C. and WEIGELT, K. (1998). Iterated dominance and iterated best response in experimental "p-beauty contests". *American Economic Review*, **88**, 947–969.

HOLT, C. and LAURY, S. (2002). Risk aversion and incentive effects. *American Economic Review*, **92** (5), 1644–1655.

ISAAC, M. and WALKER, J. (1988). Group size effects in public goods provision: The voluntary contributions mechanism. *Quarterly Journal of Economics*, **103** (1), 179–99.

ISAAC, R. M., WALKER, J. M. and THOMAS, S. H. (1984). Divergent evidence on free riding: An experimental examination of possible explanations. *Public Choice*, **43**, 113–149.

KARLAN, D. (2005). Using experimental economics to measure social capital and predict financial decisions. *American Economic Review*, **95** (5), 1688–1699.

KESER, C. and VAN WINDEN, F. (2000). Conditional cooperation and voluntary contributions to public goods. *Scandinavian Journal of Economics*, **102** (1), 23–39.

KOCHER, M., CHERRY, T., KROLL, S., NETZER, R. J. and SUTTER, M. (2008). Conditional cooperation on three continents. *Economics Letters*, **101** (3), 175–178.

—, MATZAT, D. and RIEWE, G. (2013). The team allocator game: Allocation power in public good games. *mimeo*.

KURZBAN, R. and HOUSER, D. (2005). Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. *Proceedings of the National Academy of Sciences of the United States of America*, **102** (5), 1803–1807.

LEDYARD, J. O. (1995). Public goods: A survey of experimental research. In A. Roth and J. Kagel (eds.), *A Handbook of Experimental Economics*, Princeton, NJ: Princeton University Press, pp. 111–194.

LÜTKEPOHL, H. (2007). *New introduction to multiple time series analysis*. Springer Publishing Company.

MASCLET, D., NOUSSAIR, C., TUCKER, S. and VILLEVAL, M.-C. (2003). Monetary and non-monetary punishment in the voluntary contributions mechanism. *American Economic Review*, **93** (1), 366–380.

NICKELL, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, **49** (6), pp. 1417–1426.

OLKEN, B. A. (2006). Corruption and the costs of redistribution: Micro evidence from Indonesia. *Journal of Public Economics*, **90** (4-5), 853–870.

— (2007). Monitoring corruption: Evidence from a field experiment in Indonesia. *Journal of Political Economy*, **115** (2), 200–249.

OXOBY, R. and SPRAGGON, J. (2008). Mine and yours: Property rights in dictator games. *Journal of Economic Behavior & Organization*, **65** (3-4), 703–713.

PALFREY, T. R. and PRISBREY, J. E. (1997). Anomalous behavior in public goods experiments: How much and why? *American Economic Review*, **87** (5), 829–846.

REINIKKA, R. and SVENSSON, J. (2004). Local capture: Evidence from a central government transfer program in Uganda. *Quarterly Journal of Economics*, **119** (2), 679–705.

— and — (2005). Fighting corruption to improve schooling: Evidence from a newspaper campaign in Uganda. *Journal of the European Economic Association*, **3** (2-3), 259–267.

REUBEN, E. and RIEDL, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, **77**, 122–137.

SADRIEH, A. and VERBON, H. A. (2006). Inequality, cooperation, and growth: An experimental study. *European Economic Review*, **50** (5), 1197–1222.

SEFTON, M., SHUPP, R. and WALKER, J. (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry*, **45** (4), 671–690.

SIMS, C. (1980). Macroeconomics and reality. *Econometrica*, **48** (1), 1–48.

SUTTER, M., HAIGNER, S. and KOCHER, M. (2010). Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *Review of Economic Studies*, **77** (4), 1540–1566.

THÖNI, C., TYRAN, J.-R. and WENGSTROM, E. (2012). Microfoundations of social capital. *Journal of Public Economics*, **96** (7-8), 635–643.

Appendix

Derivation of the Panel Vector Autoregressive Model

In the following, we derive the structural form panel vector autoregressive model.[28]

Consider the decision rules (6) and (7) and the belief-updating rule (8). By plugging in (8) into (6) and applying some simple transformations, we get:

$$m_{nit} = \tau_{ni} + \rho_0 R_{nit-1} + \rho_1 m_{nit-1} + u_{nit} \tag{11}$$

where

$$\begin{aligned}
\tau_{ni} &= \alpha_2 \varphi_1 \theta_{ni} + \alpha_1 \varphi_2 \theta_{ni}, \\
\rho_0 &= \alpha_2 \varphi_2, \\
\rho_1 &= (1 - \varphi_2), \\
u_{nit} &= s_{nit} + (\varphi_2 - 1) s_{nit-1}.
\end{aligned}$$

Eq. (11) explains a contributor's contribution in t by the own lagged contribution and the lagged return. Although beliefs are not directly included, beliefs enter into (11) through the decision variables. A contributor's decision is hence in line with her underlying belief formation process. As a consequence of the transformation, u_{nit} is moving-average autocorrelated. We assume that s_{nit} is $AR(1)$ such that the $MA(1)$ and the $AR(1)$ autocorrelation neutralise each other. This results in a situation where $cov(u_{nik}, u_{nij} | R_{nit-1}, m_{nit-1}, \tau_{ni}) = 0$ for $k \neq j$.

Using (11), the distributional assumption $u_{nit} \overset{iid}{\sim} \mathcal{N}(\mu_{ni}, \sigma^2_{ni})$, and the definition $M_{nt} = r \sum_{i=1}^{4} m_{nit}$, we can derive the pool M_{nt} as

$$M_{nt} = \tau_n + \rho_1 M_{nt-1} + \rho_2 R_{nt-1} + u_{nt}, \tag{12}$$

where

$$\begin{aligned}
\tau_n &= r \sum_{i=1}^{4} \tau_{ni}, \\
\rho_1 &= (1 - \varphi_2), \\
\rho_2 &= 4 r \alpha_2 \varphi_2, \\
u_{nt} &= r \sum_{i=1}^{4} u_{nit}.
\end{aligned}$$

[28] Fernández-Villaverde et al. (2007) discuss under which general conditions a model in state space representation transforms into a vector autoregressive model.

and $u_{nt} \overset{iid}{\sim} \mathcal{N}(\sum_{i=1}^{4} \mu_{ni}, \sum_{i=1}^{4} \sigma_{ni}^2)$. Combining (12)) and (7) gives

$$R_{nt} = \varrho_n + \rho_3 M_{nt-1} + \rho_4 R_{nt-1} + \beta_2 u_{nt} + v_{nt}, \tag{13}$$

where

$$\begin{aligned} \varrho_n &= \beta_1 \phi_n + \beta_2 \tau_n, \\ \rho_3 &= \beta_2 \rho_1, \\ \rho_4 &= \beta_2 \rho_2, \end{aligned}$$

and $v_t \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$. Multiplying (12) and (13) with 5/6 gives the panel vector autoregressive model summarised by (9) and (10).

References for Appendix

FERNÁNDEZ-VILLAVERDE, J., RUBIO-RAMÍREZ, J., SARGENT, T. and WATSON, M. (2007). Abcs (and ds) of understanding vars. *American Economic Review*, **97** (3), 1021–1026.

Table A1: Panel Unit Root Tests

Variable	*Breitung (B)*	*Levin-Lin-Chu (LLC)*	*Im-Pesaran-Shin (IPS)*
Contribution Rate	-2.28**	-3.56***	-3.76***
Return Rate	-1.45*	-2.01**	-2.84***

Notes: Observational unit: group n in period t; Number of observations: 396 ($N = 18$, $T = 24$); Models contain two lags selected by AIC and HQIC, panel-specific means, and exclude linear time trends; B and LLC assume common autoregressive parameters for all series, IPS relaxes the assumption of common autoregressive parameters; $H0$ of B, LLC, and IPS: All series contain a unit root; $H1$ of B and LLC: All series are stationary; $H1$ of IPS: The fraction of panels that are stationary is nonzero. The tests reject the non-stationarity of contribution rates (aggregated to group-level) and return rates in the PTG. Cooperation and trustworthiness are stable over time. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$