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

Conference Paper

Cooperation and Trustworthiness in Repeated Interaction

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Public Goods, No. E03-V3 (Revised Version: February 2015)

Provided in Cooperation with:
Verein für Socialpolitik / German Economic Association

Suggested Citation: Cagala, Tobias; Glogowsky, Ulrich; Grimm, Veronika; Rincke, Johannes (2015) : Cooperation and Trustworthiness in Repeated Interaction, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Public Goods, No. E03-V3 (Revised Version: February 2015)

This Version is available at:
http://hdl.handle.net/10419/107597

Cooperation and Trustworthiness in Repeated Interaction

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

February 5, 2015

Abstract

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 abstract 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.

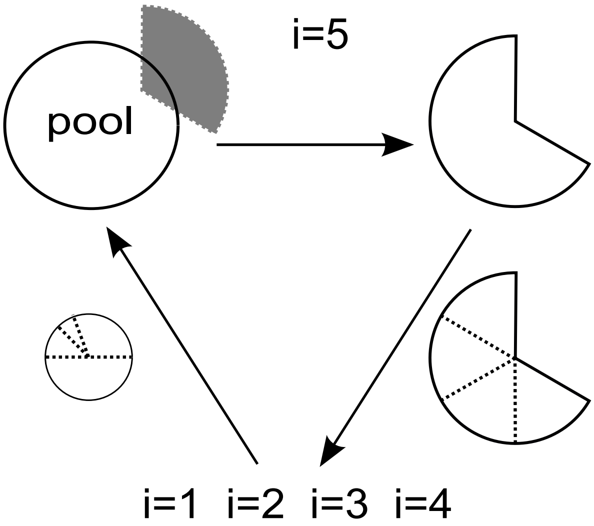
Second, our study adds to the literature on how social interactions affect cooperation. Our finding that cooperation breeds trustworthiness (and vice versa) relates to Keser and van Winden (2000), Fischbacher *et al.* (2001) and Fischbacher and Gächter (2010), who show that many individuals act as conditional 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 contributors' 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

$$\begin{aligned} x_{it} &= w - m_{it} + \frac{3}{4}\sum_{j=1}^{4} m_{jt} - \frac{3}{4}\gamma_t \sum_{j=1}^{4} m_{jt}, \quad i = \{1, \ldots, 4\}, && (1) \\ x_{5t} &= w_5 + 3\gamma_t \sum_{j=1}^{4} m_{jt}. && (2) \end{aligned}$$

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.

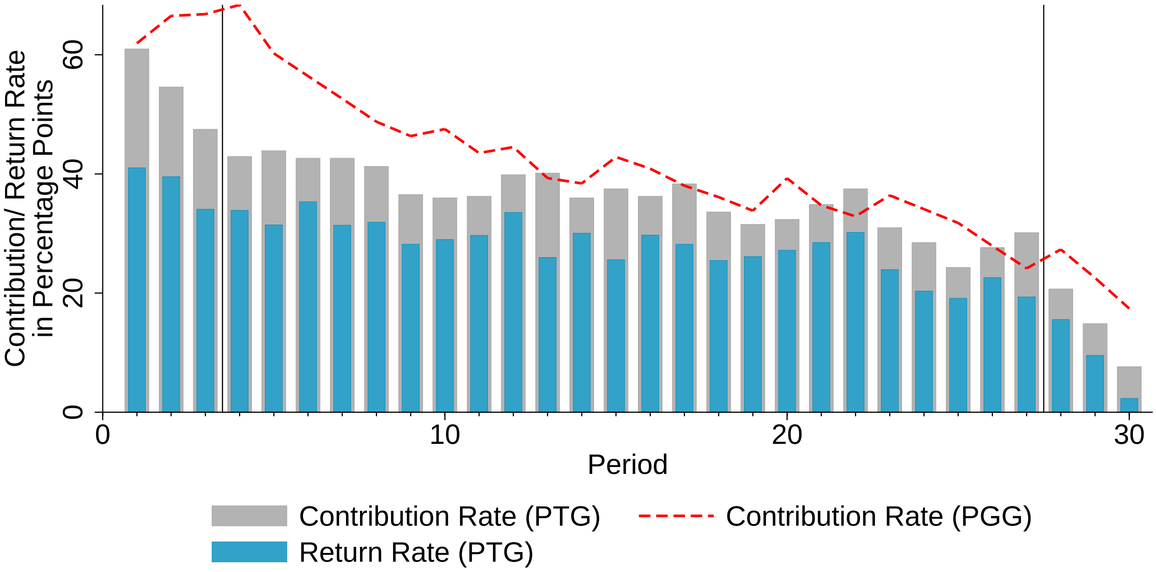
In the PTG extraction affects the scope for contributors' kindness. With zero extraction, contributors' decisions do not affect the administrator's payoff, rendering contributors' intentions towards her as neither kind nor unkind. As a result, the administrator cannot gain utility from reciprocating 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.

3.2 Level of Cooperation and Stability: Results

Figure 2 shows how group-level cooperation and trustworthiness in the PTG evolve over time, and contrasts this with contribution behavior in a standard PGG. The grey (blue) bars show mean *contribution rates* (mean *return rates*) in the PTG. Individual contribution rates are $\overline{m}_{it} = 100\frac{m_{it}}{w} = 10m_{it}$, while return rates are $\overline{R}_t = 100\frac{R_t}{4wr} = \frac{5}{6}R_t$. The dashed line depicts the development of the mean contribution rate in the PGG. The difference between contribution rates and return rates in the PTG corresponds to the share of the constant upper limit of the pool the administrator does not return to contributors.[11]

Figure 2: Mean Contribution Rates and Mean Return Rates Over Time



Notes: The Figure shows mean contribution rates in the Public Goods Game (dashed line) and mean contribution rates (grey bars) and return rates (blue bars) in the Public Trust Game over all groups in each period.
Summary: The level of cooperation in the Public Trust Game and the Public Goods Game is comparable. Contribution rates and return rates are relatively stable. Panel tests for stationarity reject the null hypothesis of overall non-stationarity of contribution rates and return rates in the Public Trust Game.

Two observations follow from Figure 2. First, apart from the typical start-game and end-game effects, cooperation and trustworthiness in the PTG are quite stable.[12] To statistically test for the stability of cooperation and trustworthiness, we draw on several panel tests for stationarity (see Table A2 in the Appendix). The unit root tests strongly reject the null hypothesis of overall non-stationarity of contribution rates (aggregated to group-level) and return rates in the PTG. This is evidence that cooperation and trustworthiness are stable over time.

The second observation is that the endogenous provision has little effect on the overall level of cooperation. Comparing the mean contribution rates across treatments,

[11] Table A1 in the Appendix presents descriptive statistics on the series depicted in Figure 2.

[12] In the following, we restrict our sample to observations from period 4 to period 27 (main interval).

we find a value of 35.9% in the PTG and a value of 39.9% in the PGG (with the same exogenously implemented mean efficiency). Using a nonparametric Mann-Whitney U test, we cannot reject the null hypothesis that contribution rates in both treatments are drawn from the same population (p-value= 0.638).[13]

In the context of sequential reciprocity, this finding is consistent with contributors perceiving the behavior of the administrator as neutral. This interpretation is supported by an observation from the survey on game related issues: on a Likert scale ranging from 1 (very dissatisfied) to 10 (very satisfied), the average rating of contributors regarding their satisfaction with the administrator's behavior was 4.6. This suggests that contributors perceive the behavior of the administrator as rather neutral.

RESULT 1: *In the PTG cooperation and trustworthiness are stable over time. Moreover, the endogenous provision has little effect on the overall level of cooperation compared to a PGG with the same efficiency.*

4 Dynamics of Cooperation and Rent Extraction: Social Interactions

4.1 Conceptual Framework

In what follows, we borrow fundamentals of decision-making from the learning literature to delineate a conceptual framework describing how agents adjust their decisions over time.[14]

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 own contributions to beliefs about the public goods return $\hat{R}_{it}$,

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

Eq. (3) consists of a deterministic and a stochastic component. The deterministic component of the decision rule $f(\cdot)$ produces i's best response to the expected behavior 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 behavior of both types of agents. The random variable s_{it} represents the stochastic component of

[13] All nonparametric tests reported in this paper use group-level data.

[14] Fudenberg and Levine (1998) summarize the literature on learning in repeated games.

the decision rule. It captures random errors in decision-making that lead to deviations from best responses.

Similar to the contributors' decision rule, the administrator chooses the return R_t contingent on a deterministic component $g(\cdot)$ and a stochastic component v_t

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

with ϕ capturing 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 realization 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. 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 realizations.[15] In particular, we assume a canonical learning rule that incorporates 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 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 behavior on observed behavior of the group of contributors. In contrast, contributors choose contributions based on beliefs about the public goods return and, hence, indirectly condition their behavior on the behavior of all other agents from previous periods. On an aggregated level, the decision rules and the belief updating mechanism lead to group-level cooperation and the administrator's trustworthiness being mutually dependent processes.

4.2 Decisions and Belief Updating: Descriptive Analysis

It is useful to descriptively consider decision rules and belief formation before studying them in the integrated framework of the PVAR model. This links our analysis directly to the theoretical discussion in section 4.1 and shows how contributors and

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

administrators interact in the repeated game.

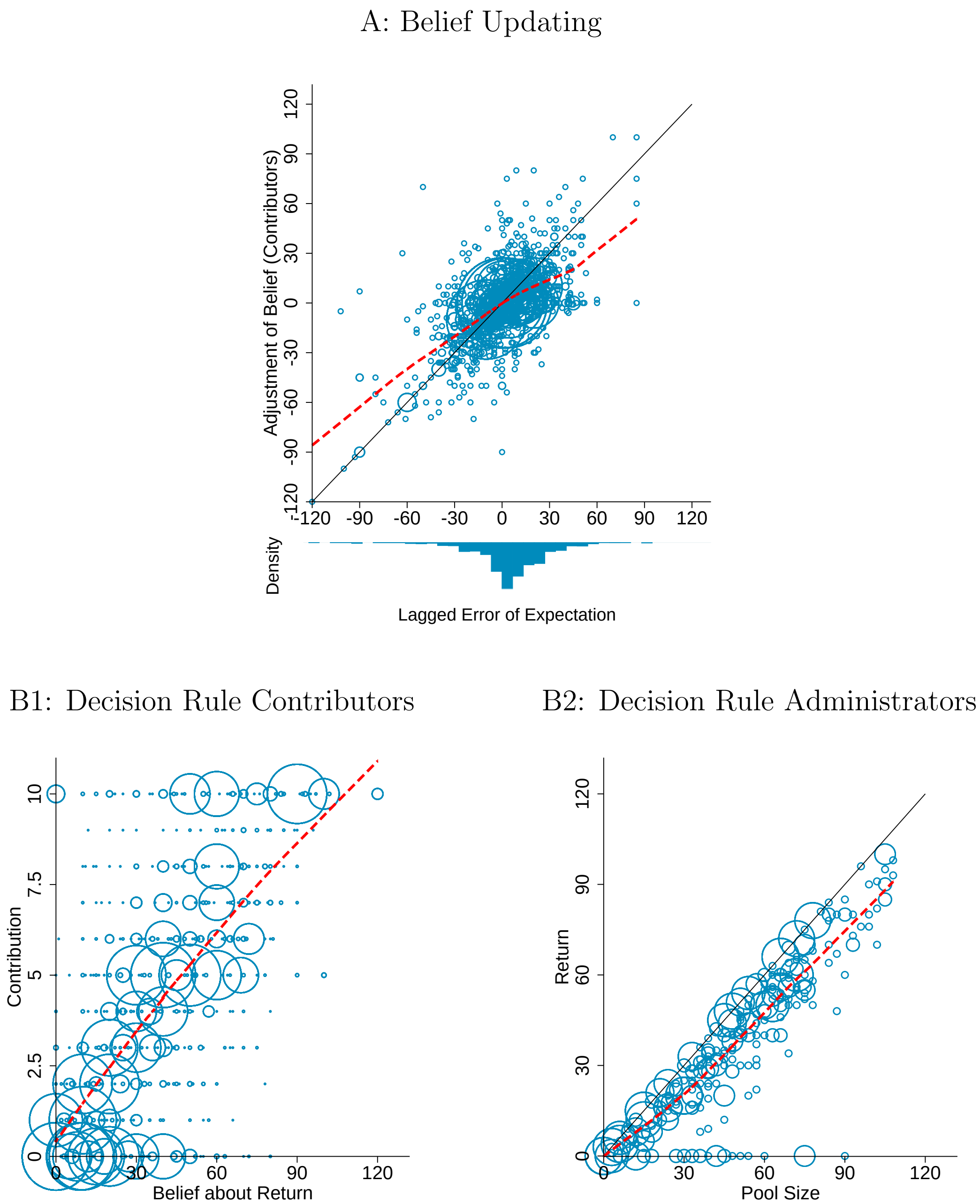
Figure 3 collects the results of the descriptive analysis. Panel A refers to contributors' learning rule and displays a scatter plot of belief adjustments together with a histogram of the lagged error of expectation (difference between lagged return and lagged belief) in the lower part. The histogram shows that the distribution is fairly symmetric around zero, suggesting that contributors' mean beliefs match up well with mean return realizations. Following the belief updating equation (5), the upper part of panel A plots contributors' belief adjustments (difference between current and lagged belief on the return rate) against the lagged error of expectation. The data show a strong positive association. Moreover, the scatter plot is centered around the origin: contributors do not adjust their beliefs if they correctly predicted the return in the previous period. The fitted values (red dashed line) reveal that the relationship between belief adjustments and the lagged error of expectation is close to linear, with a slope slightly smaller than one.[16] This suggests that contributors form beliefs using a simple linear adaptive belief updating rule.

We complete the descriptive evidence of decision-making in the PTG by studying decision rules. Panel B1 illustrates 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 choices 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 to themselves in most of the 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.

[16] We derive all fitted values in this section by nonparametric smoothing with locally weighted regressions.

Figure 3: Decision Rules and Belief Updating



A: Belief Updating

B1: Decision Rule Contributors

B2: Decision Rule Administrators

Notes: The figure shows scatter plots and fitted values (red dashed lines) for belief updating (Panel A) and decision rules (Panels B1 & 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 lagged return rates and 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 nonparametric smoothing with locally weighted regressions.

Summary: Contributors form fairly accurate beliefs about the return rate using a simple linear adaptive belief updating rule. There is a strong positive association between contributions and beliefs that is close to linear. Similarly, a strong association exists between administrators' decisions about returns and same-period contributions. Again, the association is close to linear. Taken together, the figure supports all three parts of our simple learning model.

4.3 Identifying and Estimating Dynamic Interactions

This section introduces a simple method to identify the effects of one-time changes in cooperation and trustworthiness on the future climate for cooperation in the PTG.

4.3.1 Econometric Model

We use the conceptual framework from Section 4.1 to derive a panel vector autoregressive model that captures the dynamic interaction between cooperation and trustworthiness in the PTG. We assume that $f(\cdot)$, $g(\cdot)$, and $h(\cdot)$ are additively separable and linear in all 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.

From equations (6) to (8), it is straightforward to derive a simple structural form panel vector autoregressive model (see the Appendix)[17]

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

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

where $\overline{M}_{nt} = \frac{5}{6}M_{nt}$ is the contribution rate at 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 $\overline{R}_{nt}$. In a nutshell, 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) realizations of $\overline{M}_{nt}$ and $\overline{R}_{nt}$ and exogenous (random) components captured by the error terms.[18] Our conceptual framework suggests to interpret the error terms as aggregated random errors in decision-making conditional on past decisions. In our empirical analysis, we use this random variation to identify the effects of exogenous one-time behavioral changes on future levels of cooperation and trustworthiness.

Because we model the contribution rate and the return rate as crossed processes, we

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

[18] Estimating a PVAR(1) model minimizes the Schwarz Bayesian Information Criterion.

allow for (intertemporal) relations in both directions: cooperation can affect trustworthiness and vice versa. Moreover, the PVAR model is consistent with the underlying belief updating process. In fact, we can recover the updating parameter φ_2 from our estimates.

4.3.2 Identification and Estimation of Model Parameters

The error terms represent exogenous changes in the corresponding decisions only if they are uncorrelated across equations. We follow Sims (1980) and ensure uncorrelatedness by imposing restrictions on the contemporaneous relationships between $\overline{M}_{nt}$ and $\overline{R}_{nt}$. In particular, we restrict the effect of $\overline{R}_{nt}$ on $\overline{M}_{nt}$ to zero. This restriction directly follows from our experimental design. More specifically, the sequence of decision-making in the PTG (administrator decides after contributors have made their decision) ensures that the administrator's decision cannot affect contribution behavior in the same period. We are not aware of any previous attempts to use similar identification techniques on experimental data.

We recover the PVAR parameters by estimating the reduced form PVAR model with a least square dummy variable estimator (LSDV)[19] and then computing the Cholesky factorization of the reduced-form PVAR variance-covariance matrix of the residuals.[20] Using the estimated PVAR coefficients, it is straightforward to recover the remaining structural parameters from the system of equations (6) to (8) as

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

4.4 Dynamic Interactions: Evidence

4.4.1 Structural Parameters

We find that the structural parameters of the decision rule equations $\hat{\alpha}_2 = 0.0519$ and $\hat{\beta}_2 = 0.820$ are both positive and significant (p-values < 0.001).[21] This reveals positive conditionality in the behavior 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 lagged errors of expectations and adjustments of beliefs. In fact, the estimate of $\hat{\varphi}_2$ implies that contributors fully incorporate their error in expectation from the previous period when updating their

[19] 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.

[20] Cagala and Glogowsky (2014) provide stata code and documentation to estimate panel vector autoregressive models.

[21] We use the delta method to calculate the corresponding standard errors.

beliefs.

4.4.2 Impulse Responses and Behavioral Multipliers

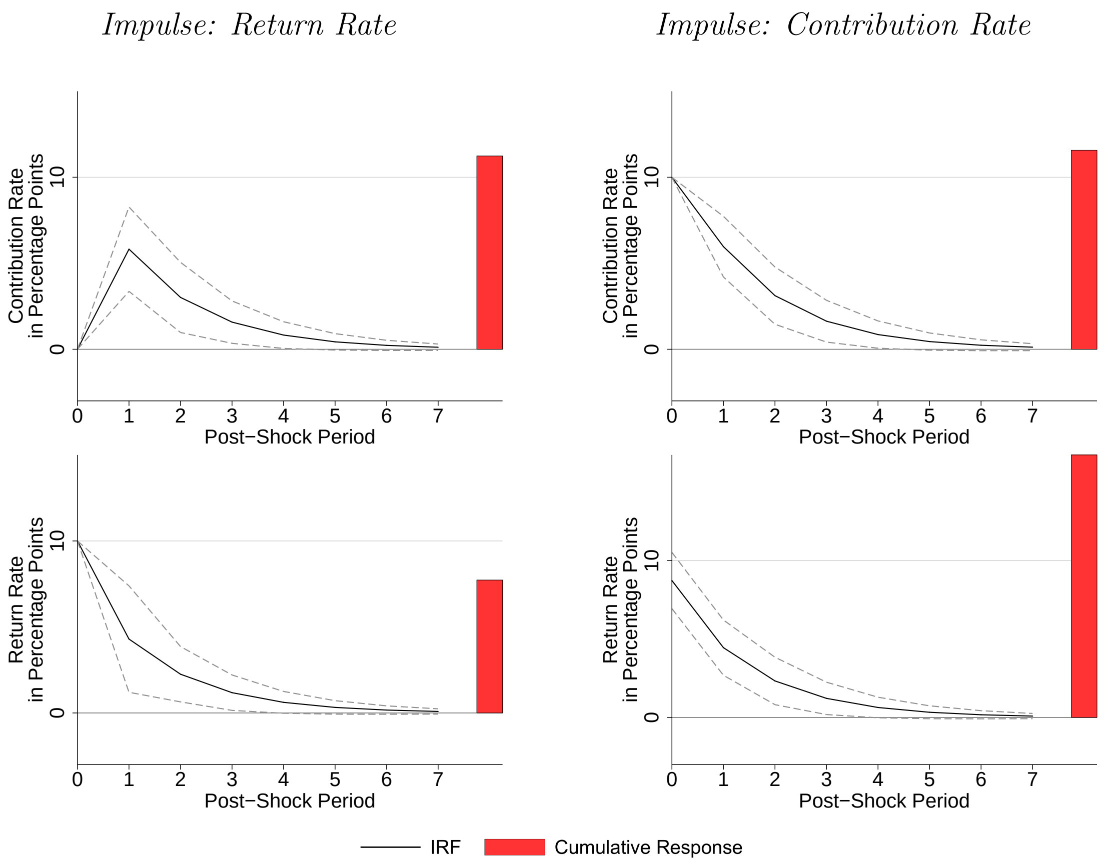
We utilize *impulse response functions* (IRFs) to visualize contributors' and administrators' behavior in response to one-time behavioral changes. Building on the PVAR model estimates, an IRF simulates the development of contribution rates and return rates after an exogenous change in either the contribution rate or the return rate (quasi-treatment) captured by an increase in the current value of one of the errors. The significance of these responses is evaluated using 95-percent confidence intervals.[22]

Figure 4 shows 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 responses of the contribution rate, while the lower panels depict responses of the return rate. Figure 4 unravels the behavioral dynamics in cooperation and trustworthiness in three distinct dimensions which we will discuss in turn.

First, the IRFs illustrate the conditionality between cooperation and trustworthiness. To see this, let us begin by considering the left-hand panels, showing responses to a 10 percentage point impulse in the return rate in step 0. As contributors decide prior to the administrator, the contemporaneous response in the contribution rate is zero (upper panel). From step 1 onwards, the responses encompass not only the effect of the initial decision, but all indirect effects working through the complex feedbacks in the system. In step 1, contributors respond to the return rate impulse by raising the contribution rate by 5.8 percentage points relative to its initial value. This increase in the contribution rate, along with the initial impulse, triggers further behavioral responses, including a step-1 addition to the return rate of 4.3 percentage points (lower panel). Turning to a ten 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 expansion of the return rate, together with the initial impulse, increases the contribution rate by 6.0 percentage points relative to its initial value. To summarize, the IRFs clearly show a positive response of contributors in the aftermath of an increase in trustworthiness and a qualitatively similar behavioral response of administrators' trustworthiness to an increase in cooperation. Hence, the positive conditionality in the PTG goes both ways.

[22] We construct these intervals based on a double bootstrap re-sampling scheme with $10,000$ repetitions.

Figure 4: Impulse Responses and Behavioral Multipliers


Impulse: Return Rate

Impulse: Contribution Rate

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 standardized ten percentage-point impulses. Left hand side (right hand side) panels show return (contribution) rate impulses.
Summary: Cooperation and trustworthiness are mutually interdependent. The responses of contributors and administrators to impulses in cooperation and trustworthiness last over several periods. Cooperation and trustworthiness eventually converge back to initial values, implying that one-time decisions do not permanently change the climate for cooperation or trustworthiness. The cumulative responses (red bars) imply strong behavioral multiplier effects.

The evidence on the response of trustworthiness to contribution rate impulses is in line with evidence for the trust game suggesting that trust breeds trustworthiness (Berg *et al.* 1995). Our results also suggest that cooperation is not only conditional to other contributors' behavior as in (Fischbacher *et al.* 2001), but that conditionality in cooperative behavior extends to the behavior of outside actors like the administrator in the PTG.

Second, Figure 4 shows that despite the presence of substantial temporary effects, the behavioral responses to impulses eventually fade out over time. We note that the responses to a return rate impulse are significantly different from zero (5% level) for 4 subsequent periods. Similarly, we find that a shock in the contribution rate has significant effects on the contribution rate (return rate) for 4 (3) subsequent periods. This leads us to the conclusion that one-time changes in behavior do not permanently alter the climate for cooperation and trustworthiness.

Third, from the IRFs, we can derive overall responses as cumulative effects of a given impulse (depicted as red bars in Figure 4).[23] These overall responses cumulate contemporaneous and future responses of a given impulse including all feedback effects and are naturally labeled behavioral multipliers effects. We find that for return rate shocks and for contribution rate shocks, the behavioral multipliers are considerably larger than zero and substantial in size. For instance, a 10 percentage point increase in the return rate causes contributor responses that add up to a 11.2 percentage point expansion of the contribution rate relative to its initial level (upper left-hand panel). Similarly, a one-time increase in the contribution rate adds an additional 11.6 percentage points to the contribution rate once all indirect and feedback effects are taken into account (upper right-hand panel). As regards responses in the return rate, we find that a one-time change in group-level contributions triggers an overall impact in the return rate of 16.4 percentage points (lower right-hand panel).[24]

RESULT 2: *Cooperation and trustworthiness are strongly mutually interdependent, suggesting that trust breeds trustworthiness and vice versa. Despite substantial temporary effects, the behavioral responses to impulses eventually fade out over time. One-time changes in behavior, hence, do not permanently alter the climate for cooperation and trustworthiness. Behavioral multipliers that cumulate the responses over time are substantial.*

4.4.3 Forecast Error Variance Decompositions

In the following, we analyze the relative importance of impulses in cooperation and trustworthiness for explaining the variation we observe in both behavioral dimensions. To enable valid comparisons, we consider 10-percentage-point impulses in both dimensions. Such standardized impulses allow us to evaluate the relative importance of one-time changes for the variation while holding the magnitude of impulses constant. For this part of the analysis, we use the standard tool of *forecast error variance decompositions* (FEVD). A FEVD shows the fractions of the forecast error variance that are due to the different standardized impulses at a given horizon. If the horizon tends to infinity, the FEVD shows the fraction of the variance in the dependent variable that the different standardized impulses explain. In our context, an FEVD quantifies the importance of a standardized impulse in the contribution rate relative to that of

[23] We calculate the overall response by cumulating all post-impulse responses which are statistically significant at the 5% level. We exclude the impulse itself in the calculation of overall responses. The overall response of the return rate includes the contemporaneous response (response of trustworthiness to same-period cooperation).

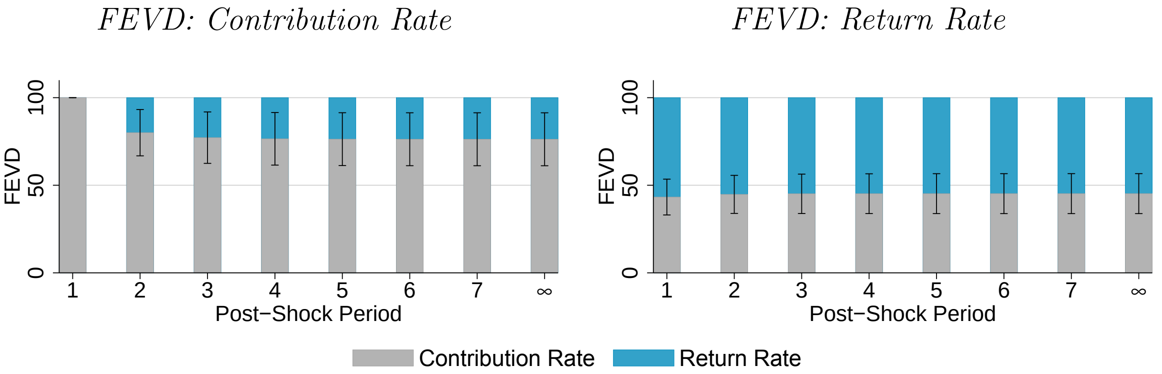
[24] The IRFs also allow us to determine the overall impact of an impulse on payoffs. Evaluating the same impulses as before, we find that contributors suffer a net loss in case of an individual one-time increase in cooperation. The same holds true for administrators: a one-time increase in the return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

a standardized impulse in the return rate for explaining the total variance.

Figure 5 summarizes the evidence on FEVDs. Two key observations emerge. First, for standardized impulses of the same size in the contribution rate and the return rate, the variation in contributions is mainly explained by impulses in contributors' behavior (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): standardized return rate impulses explain 54.8% of the long-run variation in the return rate, while contribution rate impulses explain 45.2%. Consequently, the degree of independence of contributors' decisions from the decisions of administrators is much higher than vice versa. Taken together, the FEVD implies that the observed level of variation in cooperative behavior is mainly explained by impulses in cooperation rather than impulses in trustworthiness.[25]

RESULT 3: *The variation in cooperative behavior can be explained by impulses in cooperation rather than by impulses in trustworthiness.*

Figure 5: 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).

Summary: Variation in cooperation is mainly driven by standardized impulses in cooperation. Variation in trustworthiness is equally driven by standardized impulses in trustworthiness and cooperation.

4.4.4 Impact of Baseline Attitudes Towards Cooperation and Trust

The next step of our analysis is to investigate the heterogeneity in impulse responses in terms of baseline attitudes towards cooperation and trust. The purpose of the analysis is to shed light on how baseline attitudes shape the responses of contributors and administrators to impulses in cooperation and trustworthiness. We elicit subjects' baseline attitudes towards cooperation and trust by means of a survey that we conducted two weeks after subjects took part in the experiment. The time lag between

[25] We find similar results when we account for the empirical impulse size by using residual mean squared error (RMSE) impulses (interpretable as sample impulses) instead of 10-percentage-point impulses.

the experiment and the survey attenuates any potential impact of subjects' experience in the experiment on survey responses.

The survey elicits baseline attitudes using questions that are taken from the World Value Survey. The trust question reads: "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?", with the response options "most people can be trusted", "cannot be too careful", "I don't know". The survey part aiming at cooperative attitudes lists several items that involve some form of free-riding, or non-cooperative behavior, and asks whether, on a scale of 1 to 10, they "can always be justified" or "never be justified". We then construct an internally consistent index (Cronbach's alpha 0.754) from the items "cheating on taxes if you have a chance" and "not paying the fair in public transport". We rescale the index to lie between zero (fully non-cooperative) and 100 (fully cooperative).

To study the heterogeneity in terms of baseline attitudes, we perform sample-splits and derive impulse response functions and behavioral multipliers within subsamples. The sample splits are based on (group means of) attitudes and contrast administrators (groups of contributors) in the lowest tertile with administrators (groups of contributors) in the highest tertile for each split. The analysis, thus, distinguishes between groups with trusting and non-trusting contributors, cooperative and non-cooperative contributors, and cooperative and non-cooperative administrators, respectively.[26]

The splitting procedure results in a pronounced between-subsample heterogeneity in baseline attitudes.[27] In groups with non-trusting contributors (lowest tertile), only 23.3% of contributors state that they generally trust in others. This contrasts to 83.3% in the sample of trusting contributors (highest tertile). As regards attitudes towards cooperation, the heterogeneity is similarly pronounced: groups with non-cooperative contributors (lowest tertile) have a mean index value of 50.0, while cooperative types (highest tertile) have a mean of 74.2. For administrators, we find similar differences in terms of attitudes towards cooperation: non-cooperative administrators have mean index values of 43.3 as compared to 75.0 among cooperative administrators.

We discuss the findings for the three sample splits in turn. We begin with Figure 6, which shows impulse response functions and behavioral multipliers for contributors while differentiating between non-trusting (Panel A) and trusting types (Panel B). We again differentiate between contribution and return rate impulses. However, as we consider heterogeneity among contributors here, we focus on responses related to the contribution rate. For completeness, we additionally report IRFs and the FEVD for the return rate in the Online Appendix (see Figures A1, A2 and A3). Figure 6 re-

[26] In the PTG, administrators do not respond to any direct signal about the trustworthiness of other agents. We, therefore, do not consider the heterogeneity in terms of trust among administrators.

[27] The sample with non-trusting (trusting) contributors comprises of 5 (6) groups. The sample with non-cooperative (cooperative) contributors' comprises of 6 (6) groups. For administrator-attitude splits, the sample of non-cooperative (cooperative) administrators comprises of 6 (8) groups.

veals a distinct heterogeneity in impulse responses by contributor type: in groups with non-trusting contributors, responses to contribution rate and return rate impulses are much stronger and more persistent. This implies that in groups with more trusting contributors, cooperation is much more resilient to one-time changes in cooperation and trustworthiness in the sense that contributions converge back to initial values much faster than with less trusting contributors. The difference in IRFs translates into behavioral multipliers that are more than three times larger among non-trusting types as compared to trusting types. In short, Figure 6 suggests that trust among contributors is a personal trait that protects established forms of cooperation against shocks taking the form of one-time disruptions in group-level cooperation or the trustworthiness of administrators.

Although we split the sample underlying Figure 6 by attitudes towards trust alone, it could be that other individual characteristics correlated with these attitudes drive the observed heterogeneity in outcomes. The FEVDs for the contribution rate, displayed in the lower part of the figure, reinforce the view that the heterogeneity in IRFs and overall responses is indeed due to differences in contributors' trust: comparing the FEVDs of trusting and non-trusting contributors reveals that for non-trusting contributors, the share of the variation in the contribution rate that is driven by standardized impulses in the return rate is much larger than for trusting contributors.[28]

Our findings regarding differences between trusting and non-trusting contributors are corroborated by evidence on the heterogeneity among contributors in terms of cooperative attitudes. Figure 7 displays the results for a similar sample split as before, the only difference being that we now split the sample into groups with non-cooperative (Panel A) and cooperative (Panel B) contributor types. Again, the figure reveals strong differences in behavioral responses: in groups of non-cooperative contributors, one-time changes in cooperation and trustworthiness trigger more persistent responses in cooperation and lead to much stronger cumulative responses.

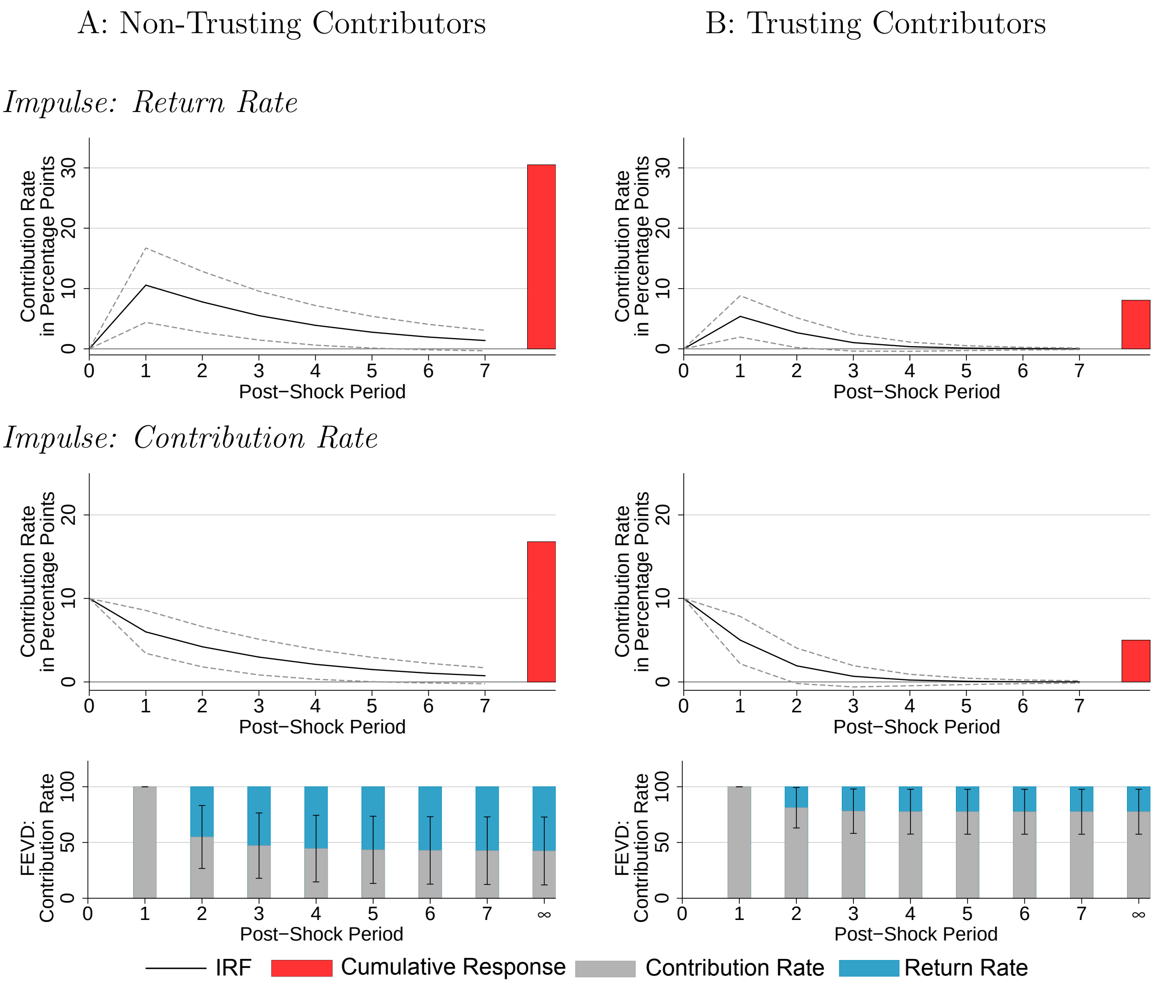
We complete the analysis of heterogeneous impulse responses by considering the impact of cooperative attitudes among administrators. Figure 8 presents the findings. It follows the layout of the previous figures but focuses on return rate responses based on a sample split into groups with non-cooperative (Panel A) and cooperative (Panel B) administrators. As regards impulses in the return rate, we do not find much of a difference in responses between cooperative and non-cooperative administrators. In fact, the IRFs look quite similar, and the responses in the first post-shock period are insignificant for both types. The similarity does not come as a surprise as administrators' attitudes towards cooperation should have no direct effect on their reaction to an

[28] We do not observe these differences between FEVDs for trusting and non-trusting types if we use RMSE-impulses. The reason is that in groups with non-trusting types the contribution rate RMSE is relatively large, making one-time changes in cooperation the main driver of the overall variation in the non-standardized FEVD.

impulse in their past behavior. Turning to the lower panel, we find that administrators respond very differently to contribution rate impulses, depending on their attitudes towards cooperation: for non-cooperative administrators, the impact of one-time changes in the contribution rate is much more persistent and cumulatively more pronounced (Panel A) as for cooperative administrators (Panel B). For instance, considering a one-time drop in the contribution rate this finding implies that the overall downward adjustment of the return-rate depends on attitudes towards cooperation among the administrators. Administrators who are themselves non-cooperative respond much stronger and more persistently to disruptions in group-level cooperation.

RESULT 4: *The behavioral multipliers are much larger in groups with less cooperative and less trusting types than with more cooperative and more trusting types.*

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


A: Non-Trusting Contributors

B: Trusting Contributors

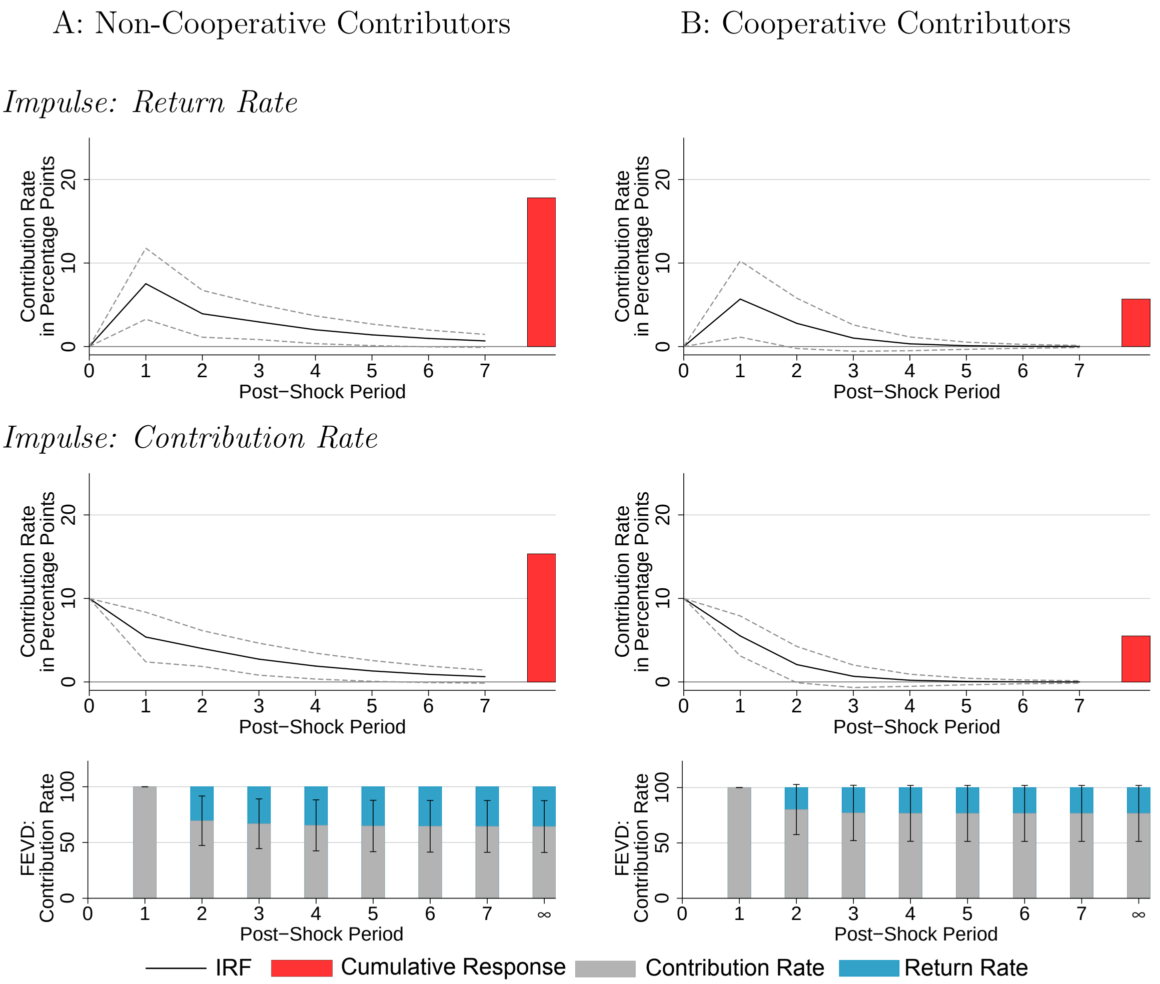
Impulse: Return Rate

Impulse: Contribution Rate

Notes: First- and second-row panels show IRFs (solid line) with 95% confidence bands (dashed lines) and cumulative responses (red bars). 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 standardized ten percentage point return rate (contribution rate) impulses. 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. 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.

Summary: Non-trusting contributors show a longer and more persistent response to impulses compared to trusting contributors. The variation in contribution rates among non-trusting types is mainly driven by return rate impulses, while the variation among trusting types is mostly driven by impulses in the contribution rate.

Figure 7: Heterogenous Responses – Cooperative vs. Non-Cooperative Contributors



A: Non-Cooperative Contributors

B: Cooperative Contributors

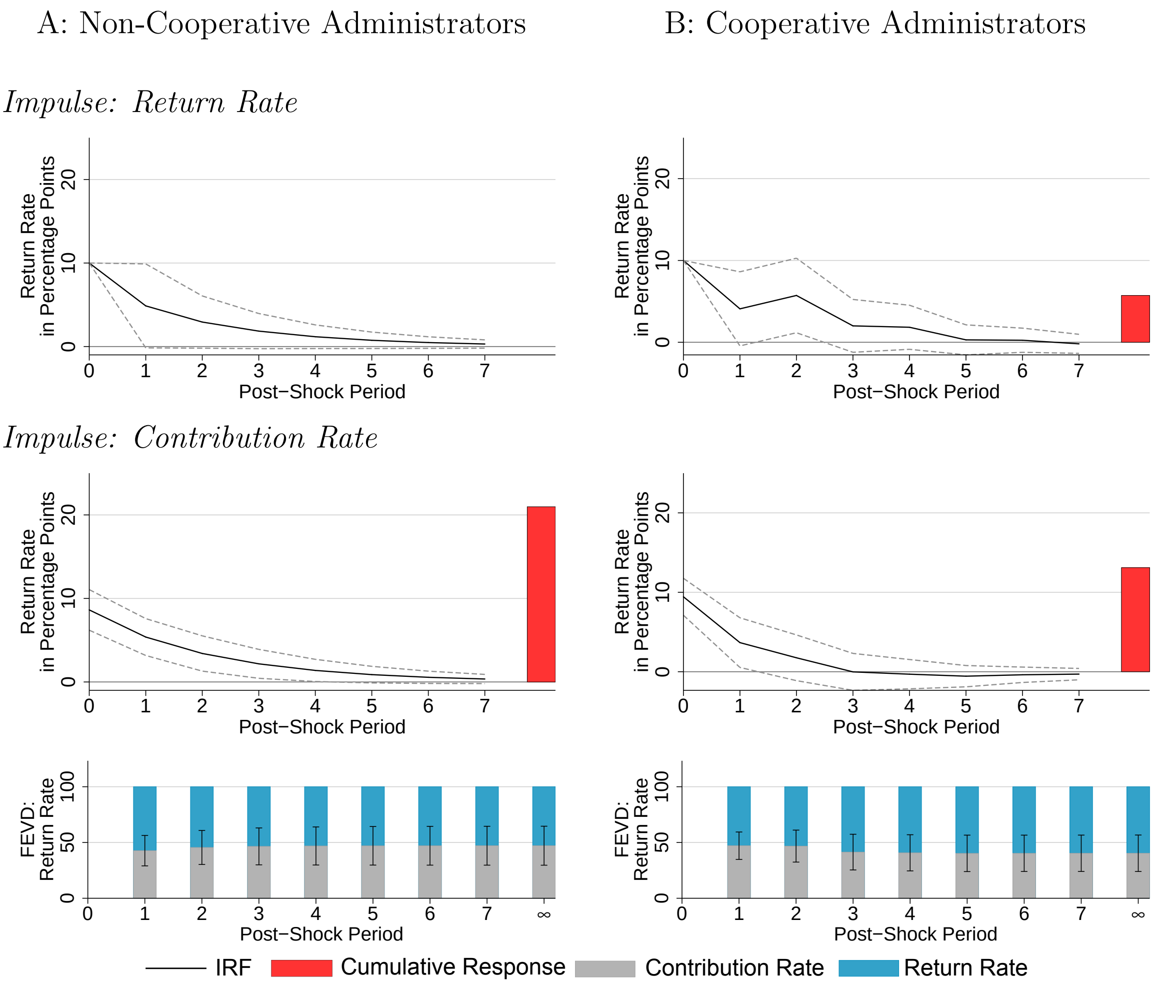
Impulse: Return Rate

Impulse: Contribution Rate

Notes: First- and second-row panels show IRFs (solid line) with 95% confidence bands (dashed lines) and cumulative responses (red bars). 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 standardized ten percentage point return rate (contribution rate) impulses. 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. Groups in the highest (lowest) tertile of the distribution of the cooperativeness index are classified as non-cooperative (cooperative) type groups.

Summary: Non-cooperative contributors show a longer and more persistent response to impulses.

Figure 8: Heterogenous Responses – Cooperative vs. Non-Cooperative Administrators



A: Non-Cooperative Administrators

B: Cooperative Administrators

Impulse: Return Rate

Impulse: Contribution Rate

Notes: First- and second-row panels show IRFs (solid line) with 95% confidence bands (dashed lines) and cumulative responses (red bars). Third-row panels display FEVDs (grey and blue bars) with 95% confidence bands (spikes) for administrators. The cumulative response is the sum over significant post-impulse return rates. The first-row (second-row) IRFs are for standardized ten percentage point return rate (contribution rate) impulses. 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. Administrators in the highest (lowest) tertile of the distribution of the cooperativeness index are classified as non-cooperative (cooperative) types.

Summary: Non-cooperative administrators show a longer and more persistent response to contribution rate impulses.

5 Conclusion

Public goods provision often involves groups of individuals repeatedly interacting with administrators who can extract private rents from the pool of contributions. To study how group members contribution behavior (i.e., cooperation) and rent extraction decisions by an administrator (i.e., trustworthiness) interact over time, we use key elements of the Public Goods Game and the Trust Game and combine them into the Public Trust Game.

As in other context involving social interactions, isolating causal effects poses a methodological challenge (Sobel, 2002). We overcome this problem by exploiting the sequential structure of our game in a panel vector autoregressive model. Analyzing the time series originating from the dynamic interaction between contributors and administrators, we establish several novel findings. First, contributors and administrators manage to establish a form of interaction that is stable over time and leads to a level of cooperation that is not statistically different from the level obtained in a standard Public Goods Game with the same efficiency. Second, we demonstrate that cooperation breeds trustworthiness and vice versa: one-time increases (decreases) in cooperation trigger significant increases (decreases) in cooperation and trustworthiness in subsequent periods. Similarly, one-time increases (decreases) in trustworthiness positively (negatively) affect future cooperation and trustworthiness. We derive behavioral multipliers that measure the overall impact of such one-time changes in behavior and demonstrate that the multipliers in the Public Trust Game are substantial. All these responses to one-time changes in behavior are of a temporary nature, with behavior in both dimensions eventually converging back to initial levels. We conclude that random errors in decision-making leading to deviations from best responses do not permanently alter the climate for cooperation. Finally, we study the heterogeneity in responses to shocks by subjects' baseline attitudes towards cooperation and trust and find that temporary disruptions in cooperation and trustworthiness are particularly damaging in settings with a lack of cooperative attitudes and trust.

From a methodological point of view, we contribute to the literature by demonstrating how panel vector autoregressive models can be applied to model decision-making in the laboratory. Our methods can be applied 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.

We believe that our results speak to a broad range of settings where the provision of public goods depends on the behavior of agents who act at a higher hierarchical level compared to the group of contributors. A natural application is public goods provision when taxpayers interact with the bureaucracy. In this setting, the level of public goods provided does not only depend on taxpayers' compliance with the tax law, but also on the administrative efficiency of the bureaucracy. Among other things, our findings suggest that taxpayers and bureaucrats can establish a form of interaction that is stable over time, but that one-time changes in bureaucratic efficiency can be expected to have strong effects on taxpayers' compliance behavior. The evidence from the Public Trust Game also suggests that in societies with a lack of cooperative attitudes and trust, the behavioral multipliers of one-time changes in the behavior of taxpayers and bureaucrats will lead to much stronger and longer-lasting deviations from established levels of tax compliance and bureaucratic efficiency compared to societies with high levels of cooperative attitudes and trust.

In this paper, we analyze a setting where contributors and administrators repeatedly interact and can adjust their behavior in every period. This naturally restricts our analysis to the effects of one-time changes in behavior. It remains for future research to explore how the interplay between cooperation and trustworthiness is affected by permanent changes in behavior of contributors and administrators, respectively.

References

Anderson, C. M. 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., Harbaugh, W. and Vesterlund, L. (2003). The carrot or the stick: Rewards, punishments, and cooperation. *American Economic Review*, **93** (3), 893–902.

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.

Berg, J., John, D. and Kevin, M. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, **10** (1), 122–142.

Bochet, O., Page, T. and Putterman, L. (2006). Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior & Organization*, **60** (1), 11–26.

Brosig, J., Weimann, J. and Ockenfels, A. (2003). The effect of communication media on cooperation. *German Economic Review*, **4** (2), 217–241.

Cagala, T. and Glogowsky, U. (2014). *Panel vector autoregressions for Stata (xtvar)*. Software package available at www.wirtschaftspolitik.rw.uni-erlangen.de/Software/XTVAR.zip.

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

Cullen, J., Turner, N. and Washington, E. (2014). The politics of tax evasion. *mimeo*.

Dufwenberg, M. and Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, **47** (2), 268–298.

Fehr, E. and Gächter, S. (2002). Altruistic punishment in humans. *Nature*, **415** (6868), 137–140.

— and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, **114** (3), 817–868.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., Sargent, T. J. and Watson, M. W. (2007). Abcs (and ds) of understanding vars. *American Economic Review*, **97** (3), 1021–1026.

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.

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

Friedman, J. W. (1971). A non-cooperative equilibrium for supergames. *Review of Economic Studies*, **38** (113), 1–12.

Fudenberg, D. and Levine, D. K. (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.

—, Renner, E. and Sefton, M. (2008). The long-run benefits of punishment. *Science*, **322** (5907), 1510.

Greiner, B. (2004). The online recruitment system orsee 2.0 - a guide for the organization of experiments in economics, Working Paper Series in Economics 10, University of Cologne, Department of Economics.

Güth, W., Levati, M. V., Sutter, M. and van der Heijden, E. (2007). Leading by example with and without exclusion power in voluntary contribution experiments. *Journal of Public Economics*, **91** (5-6), 1023–1042.

Isaac, R. M., McCue, K. F. and Plott, C. R. (1985). Public goods provision in an experimental environment. *Journal of Public Economics*, **26** (1), 51–74.

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

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

Luttmer, E. F. P. and Singhal, M. (2014). Tax morale. *Journal of Economic Perspectives*, **28** (4), 149–68.

Masclet, D., Noussair, C., Tucker, S. and Villeval, M.-C. (2003). Monetary and nonmonetary 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.

OECD (2013). *Tax and development: What drives tax morale?* Organization for Economic Co-operation and Development.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, **83** (5), 1281–1302.

Reuben, E. and Riedl, A. (2009). Public goods provision and sanctioning in privileged groups. *Journal of Conflict Resolution*, **53** (1), 72–93.

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

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

Sobel, J. (2002). Can we trust social capital? *Journal of Economic Literature*, **40** (1), 139–154.

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

Appendix

Derivation of the Panel Vector Autoregressive Model

Consider the decision rules (6) and (7) and the belief updating rule (8). By plugging in (8) into (6) and by 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 neutralize 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 $u_{nit} \overset{iid}{\sim} \mathcal{N}(\mu_{ni}, \sigma_{ni}^2)$, 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

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

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 summarized by (9) and (10).

Table A1: Descriptive statistics

Variable	*Learning Interval*		*Main Interval*					*End-Game Interval*	
	Mean	Std.Dev.	Mean	Std.Dev.	Med.	Min.	Max.	Mean	Std.Dev.
Contribution Rate PTG	51.4	17.6	35.9	23.5	36.2	0	90.0	14.4	17.8
Return Rate PTG	38.2	20.3	27.8	22.2	25.0	0	83.3	9.13	12.8
Contribution Rate PGG	64.7	17.6	39.9	21.9	37.5	0	90.0	21.3	15.4

Notes: Observational unit: group n in period t; Total number of observations PTG: 540 ($N = 18$, $T = 30$); Total number of observations PGG: 450 ($N = 15$, $T = 30$).

Summary: The mean contribution rate in the learning interval is significantly larger than the mean contribution rate in the main interval (one-sided Wilcoxon signed rank test, $p = 0.000$) which is in turn significantly larger than the mean contribution rate in the end-game interval ($p = 0.000$). Furthermore, the mean return rate in the main interval is significantly smaller than the mean return rate in the learning interval ($p = 0.002$) and significantly larger than the mean return rate in the end-game interval ($p = 0.000$).

Table A2: 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 ($N = 18$, $T = 22$); Models contain 2 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; $^{***}p$ <0.01, $^{**}p$ <0.05, $^{*}p$ <0.1

Summary: 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.

Cooperation and Trustworthiness in Repeated Interaction

Online Appendix

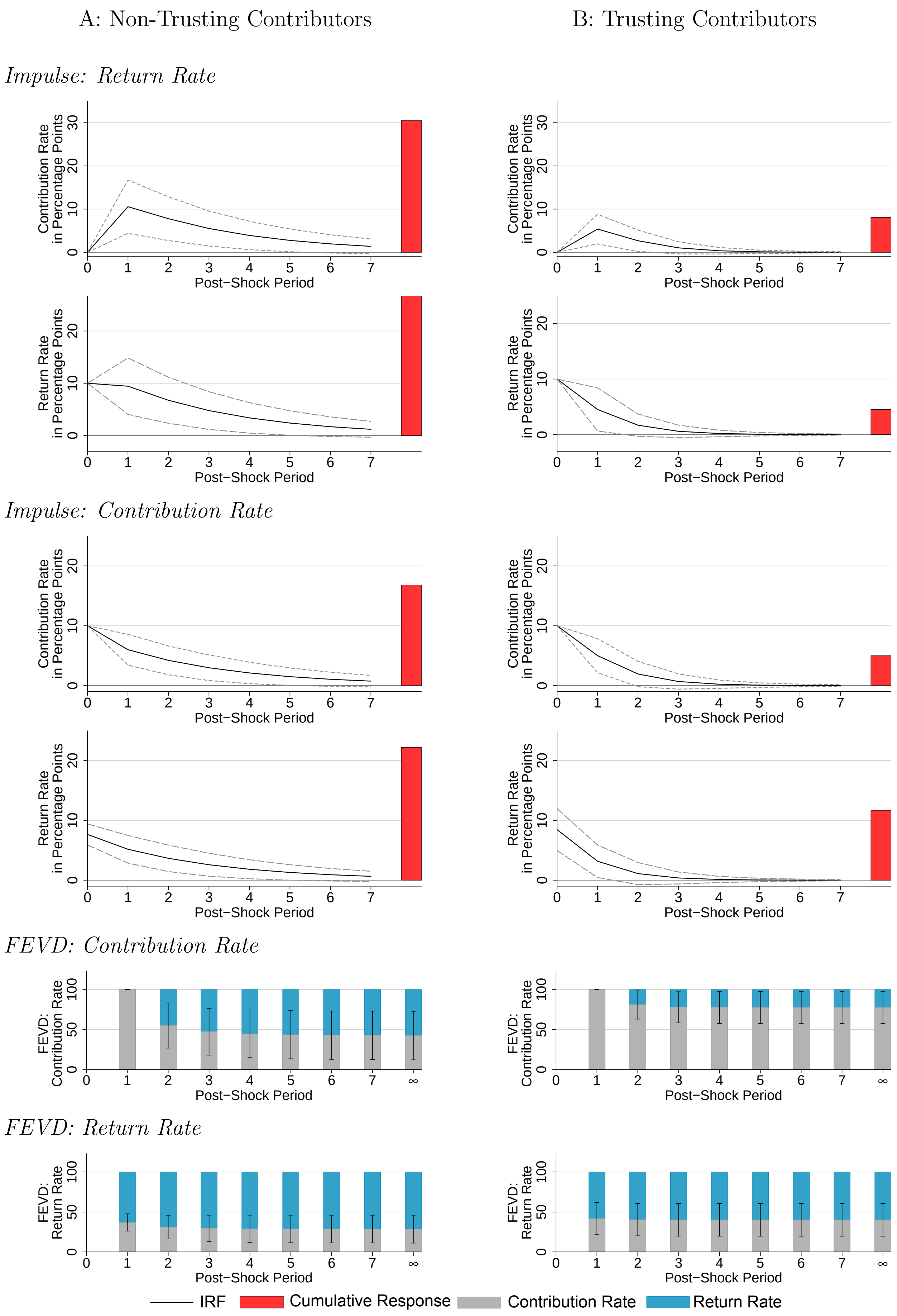
Not for Publication

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

January 29, 2015

*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)

Figure A1: Heterogenous Responses – Trusting vs. Non-Trusting Contributors



A: Non-Trusting Contributors

B: Trusting Contributors

Impulse: Return Rate

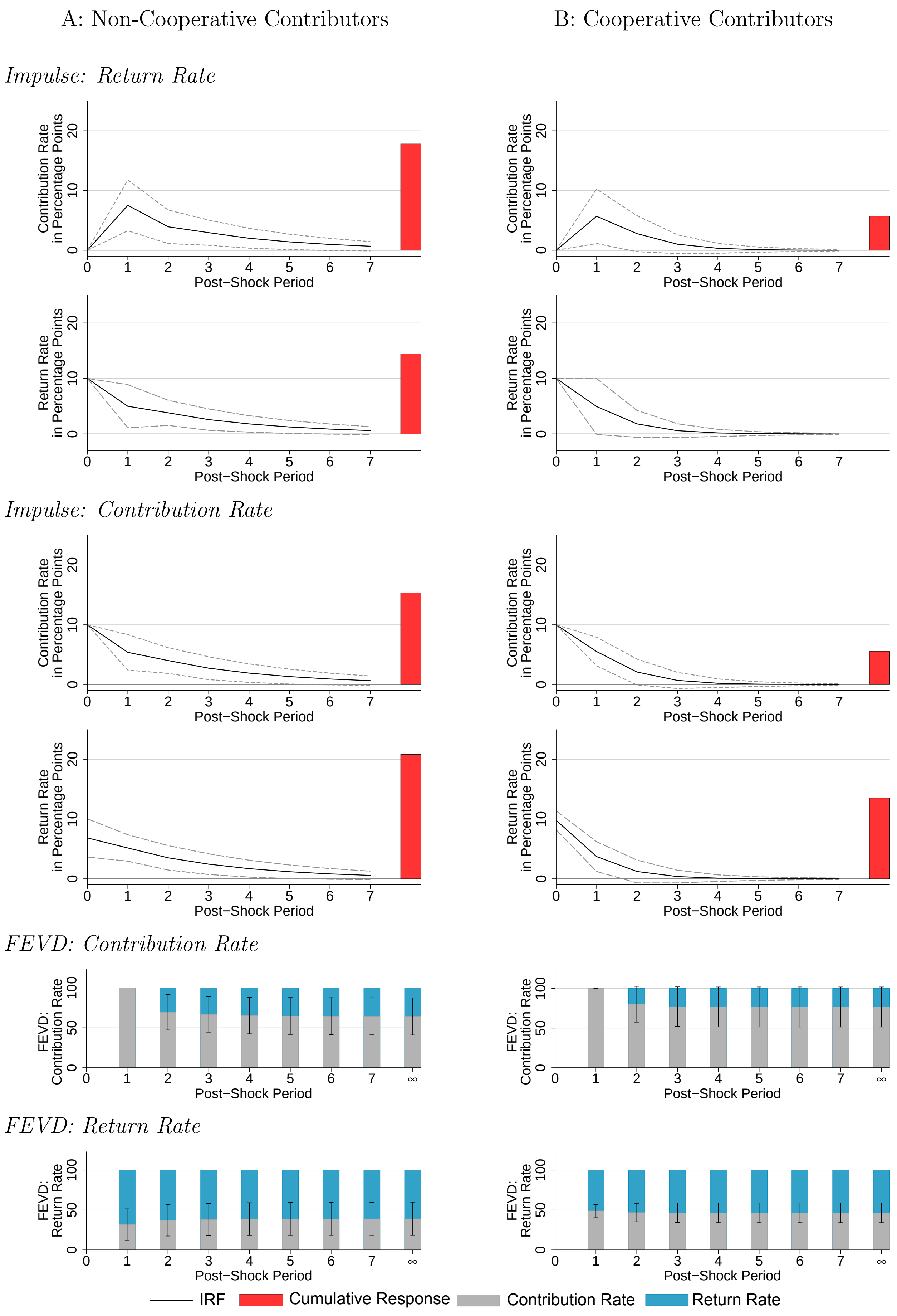
Impulse: Contribution Rate

FEVD: Contribution Rate

FEVD: Return Rate

Notes: The upper (lower) part of the figure shows IRFs (FEVDs). For detailed notes see Figure 6 in the paper.

Figure A2: Heterogenous Responses – Cooperative vs. Non-Cooperative Contributors



A: Non-Cooperative Contributors

B: Cooperative Contributors

Impulse: Return Rate

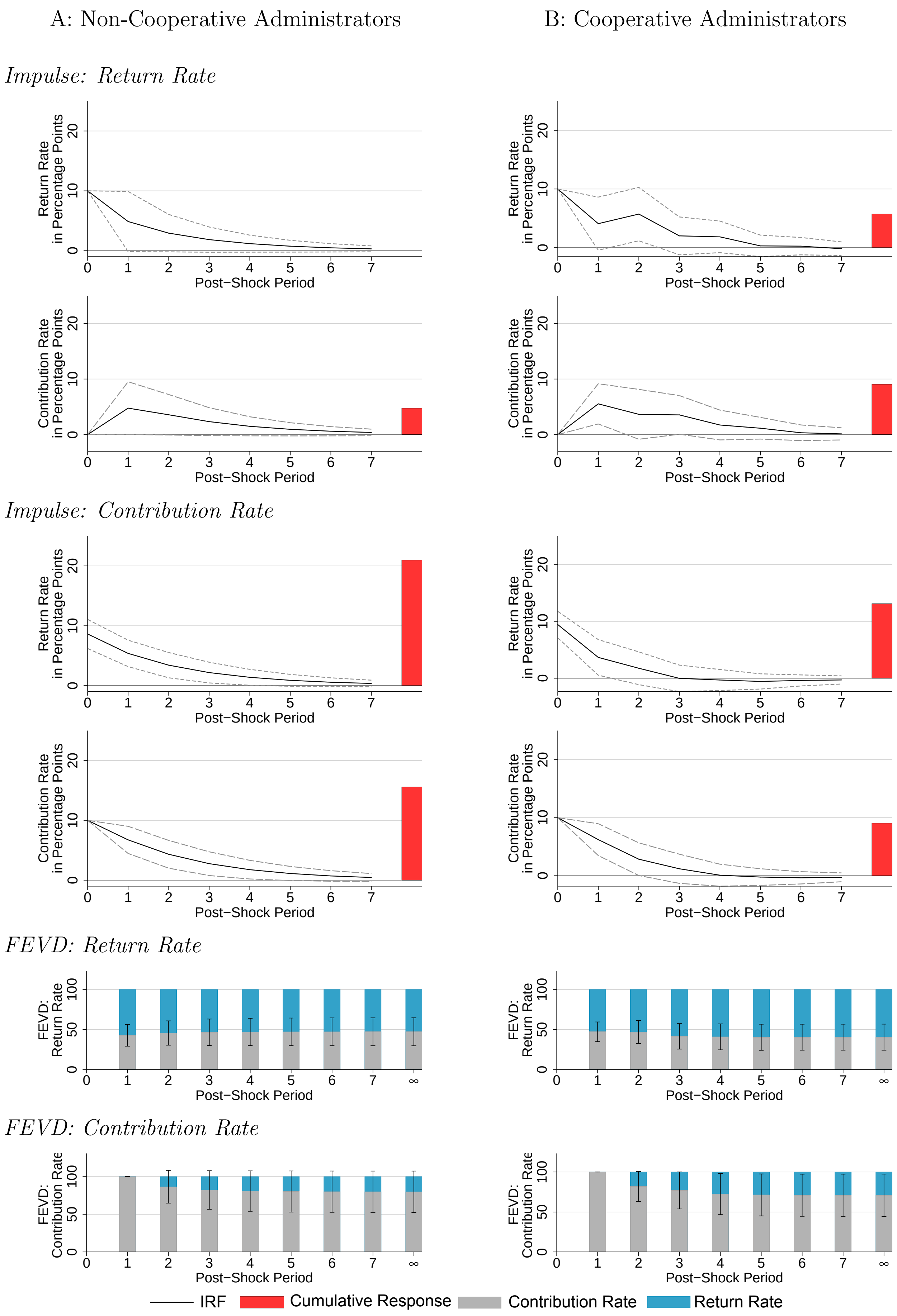
Impulse: Contribution Rate

FEVD: Contribution Rate

FEVD: Return Rate

Notes: The upper (lower) part of the figure shows IRFs (FEVDs). For detailed notes see Figure 7 in the paper.

Figure A3: Heterogenous Responses – Cooperative vs. Non-Cooperative Administrators



A: Non-Cooperative Administrators

B: Cooperative Administrators

Impulse: Return Rate

Impulse: Contribution Rate

FEVD: Return Rate

FEVD: Contribution Rate

Notes: The upper (lower) part of the figure shows IRFs (FEVDs). For detailed notes see Figure 8 in the paper.

Instructions PTG & PGG
(PGG Instructions exclude highlighted text components)

Welcome and thank you for participating in today's experiment. Please read the instructions carefully.

If you have any questions, please raise your hand. One of the experimenters will answer your questions. **You are not allowed to communicate with other participants of the experiment.** Violation of this rule will lead to exclusion from the experiment. Please turn off your cell phone.

This is an experiment in economic decision making. For showing up on time, you receive a one-time payment of EUR 2.5. For attending the second part of the experiment, you receive a one-time payment of EUR 6. During the experiment you will earn additional money. Your additional earnings depend on your behavior and the behavior of other participants. During the experiment, money is displayed in Experimental Currency Units (ECU). The exchange rate is **1 Euro = 40 ECU. Your entire earnings will be paid to you in cash at the end of the second part of the experiment.**

You will not learn about the identity of other participants. We will not communicate your earnings or your role in the experiment to other participants. The data will be analyzed anonymously.

Experiment

Duration

The experiment is divided into periods. In each period you face the same decision-making situation. The experiment consists of **30 periods**.

Roles

Every participant is assigned a role, either A or B. In the following we refer to participants as **A-participant** and **B-participant**. The roles are **randomly assigned** before the first period and will not change during the experiment. All participants are treated equally during the assignment. Before the first period, every participant is informed about her role.

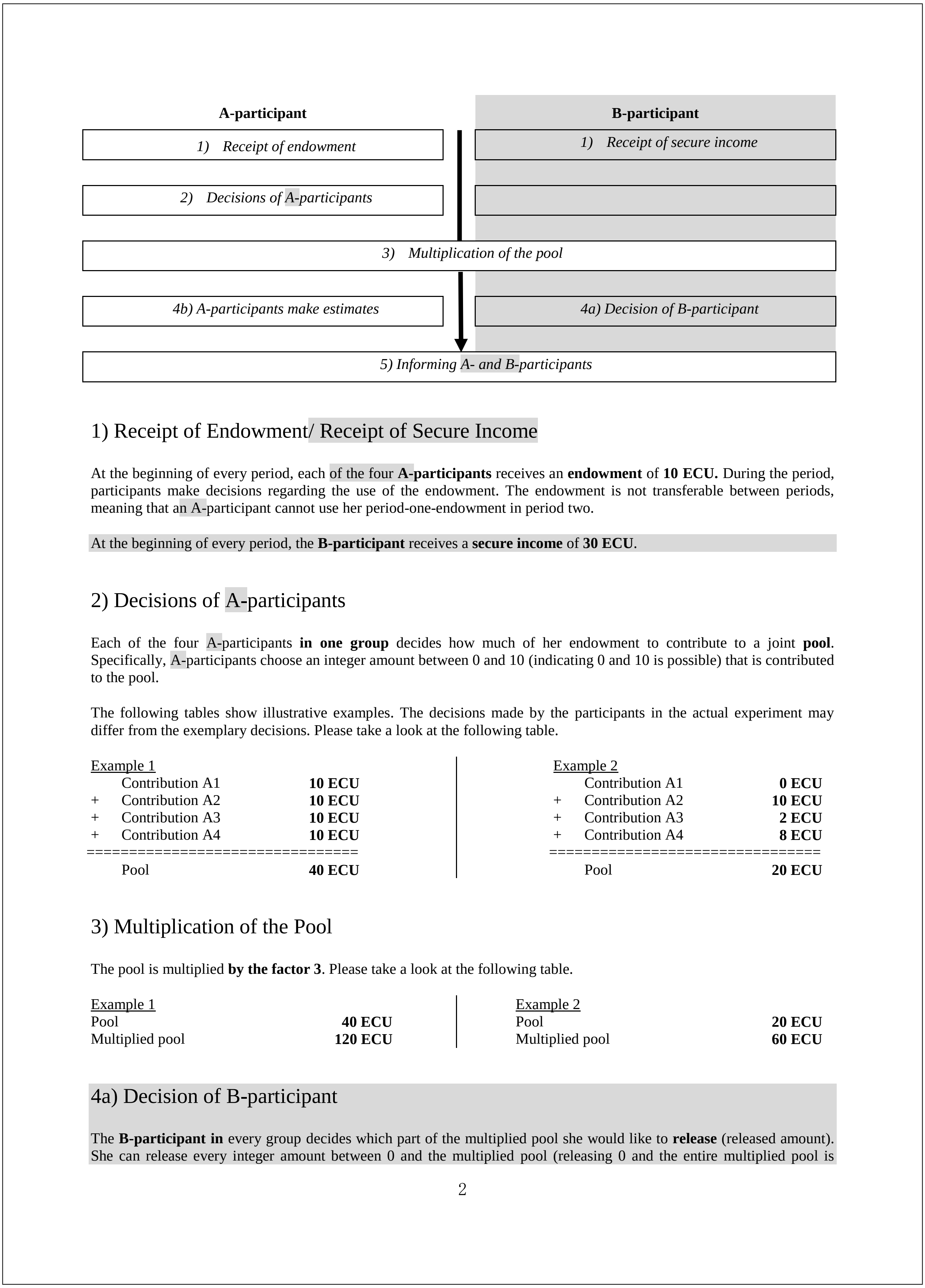
Groups

Prior to the first period, all participants are divided randomly into **independent groups** of five participants. Each group consists of **four A-participants** (in the following A1 to A4) and **one B-participant** (in the following B).

Groups remain the same throughout the experiment, meaning that you solely interact with members of your group. Decisions made by members of other groups will not affect your group.

Sequence

Every period follows the same sequence, illustrated in the following figure.



A-participant

B-participant

1) Receipt of endowment

1) Receipt of secure income

2) Decisions of A-participants

3) Multiplication of the pool

4b) A-participants make estimates

4a) Decision of B-participant

5) Informing A- and B-participants

1) Receipt of Endowment/ Receipt of Secure Income

At the beginning of every period, each of the four **A-participants** receives an **endowment** of **10 ECU.** During the period, participants make decisions regarding the use of the endowment. The endowment is not transferable between periods, meaning that an A-participant cannot use her period-one-endowment in period two.

At the beginning of every period, the **B-participant** receives a **secure income** of **30 ECU**.

2) Decisions of A-participants

Each of the four A-participants **in one group** decides how much of her endowment to contribute to a joint **pool**. Specifically, A-participants choose an integer amount between 0 and 10 (indicating 0 and 10 is possible) that is contributed to the pool.

The following tables show illustrative examples. The decisions made by the participants in the actual experiment may differ from the exemplary decisions. Please take a look at the following table.

<u>Example 1</u>

	Contribution A1	**10 ECU**
+	Contribution A2	**10 ECU**
+	Contribution A3	**10 ECU**
+	Contribution A4	**10 ECU**
================================		
	Pool	**40 ECU**

<u>Example 2</u>

	Contribution A1	**0 ECU**
+	Contribution A2	**10 ECU**
+	Contribution A3	**2 ECU**
+	Contribution A4	**8 ECU**
================================		
	Pool	**20 ECU**

3) Multiplication of the Pool

The pool is multiplied **by the factor 3**. Please take a look at the following table.

<u>Example 1</u>

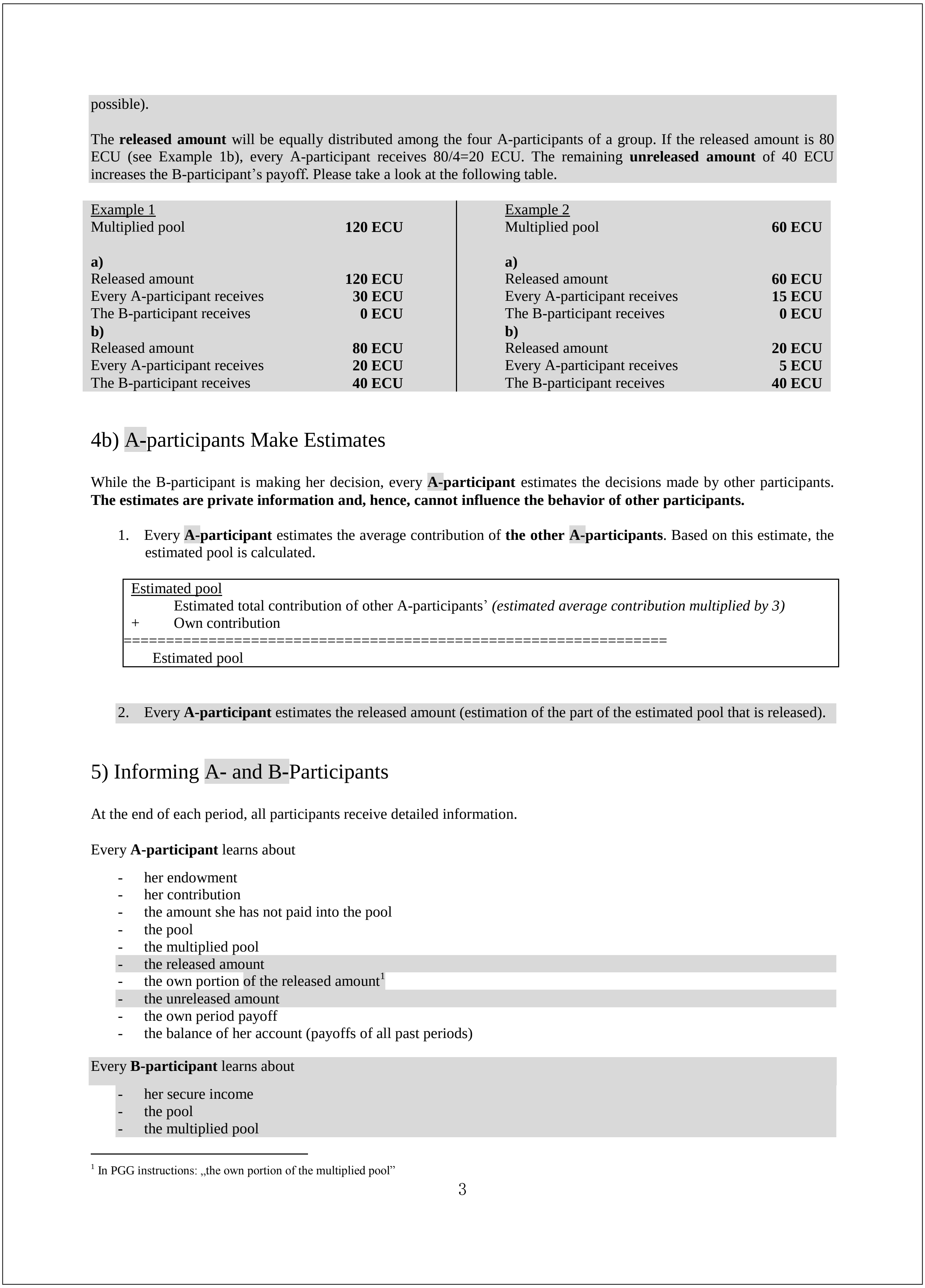
Pool	**40 ECU**
Multiplied pool	**120 ECU**

<u>Example 2</u>

Pool	**20 ECU**
Multiplied pool	**60 ECU**

4a) Decision of B-participant

The **B-participant in** every group decides which part of the multiplied pool she would like to **release** (released amount). She can release every integer amount between 0 and the multiplied pool (releasing 0 and the entire multiplied pool is



possible).

The **released amount** will be equally distributed among the four A-participants of a group. If the released amount is 80 ECU (see Example 1b), every A-participant receives 80/4=20 ECU. The remaining **unreleased amount** of 40 ECU increases the B-participant's payoff. Please take a look at the following table.

Example 1		Example 2	
Multiplied pool	**120 ECU**	Multiplied pool	**60 ECU**
a)		**a)**	
Released amount	**120 ECU**	Released amount	**60 ECU**
Every A-participant receives	**30 ECU**	Every A-participant receives	**15 ECU**
The B-participant receives	**0 ECU**	The B-participant receives	**0 ECU**
b)		**b)**	
Released amount	**80 ECU**	Released amount	**20 ECU**
Every A-participant receives	**20 ECU**	Every A-participant receives	**5 ECU**
The B-participant receives	**40 ECU**	The B-participant receives	**40 ECU**

4b) A-participants Make Estimates

While the B-participant is making her decision, every **A-participant** estimates the decisions made by other participants. **The estimates are private information and, hence, cannot influence the behavior of other participants.**

1. Every **A-participant** estimates the average contribution of **the other A-participants**. Based on this estimate, the estimated pool is calculated.

Estimated pool
Estimated total contribution of other A-participants' *(estimated average contribution multiplied by 3)*
+ Own contribution
==
Estimated pool

2. Every **A-participant** estimates the released amount (estimation of the part of the estimated pool that is released).

5) Informing A- and B-Participants

At the end of each period, all participants receive detailed information.

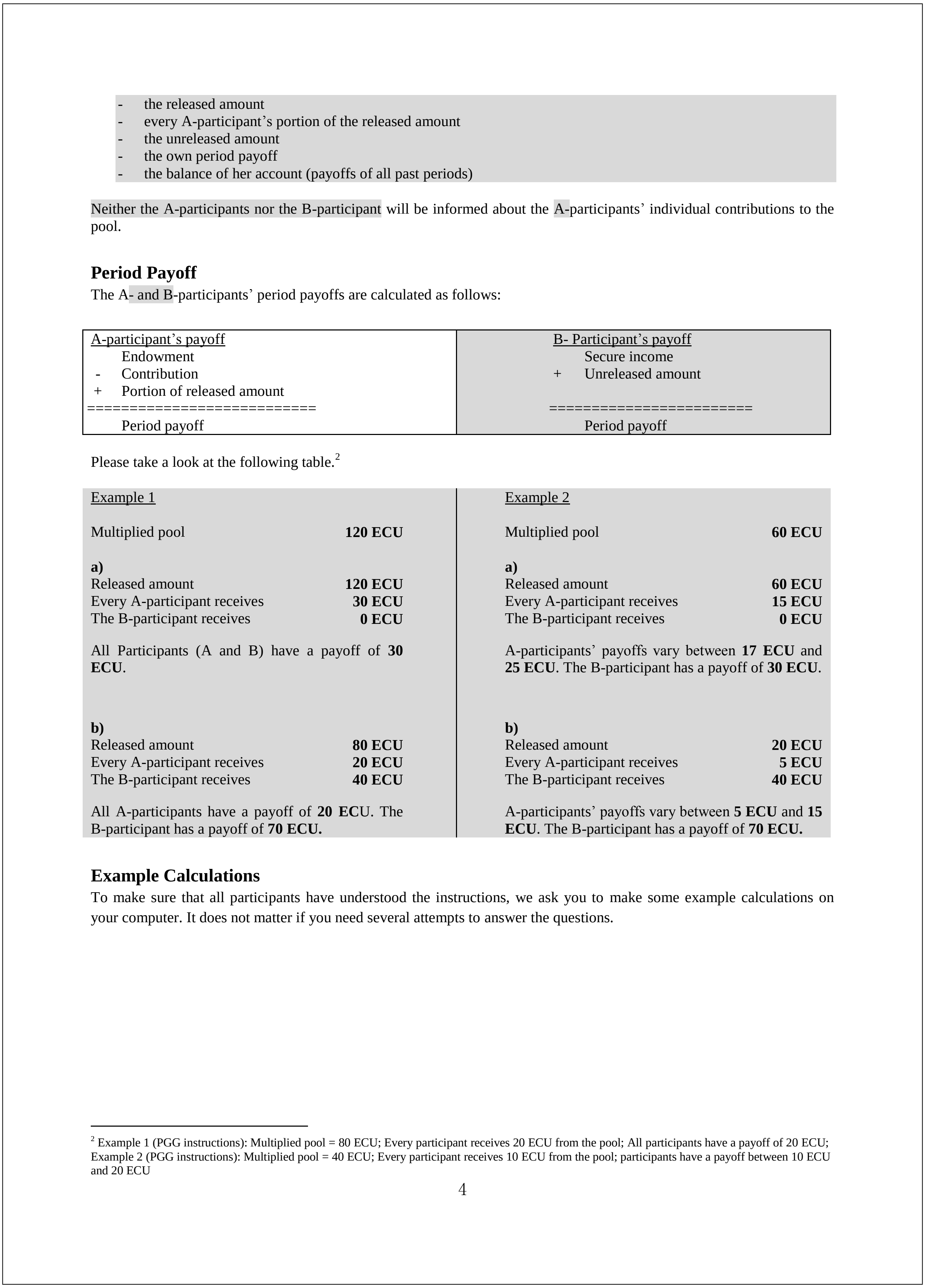
Every **A-participant** learns about

- her endowment
- her contribution
- the amount she has not paid into the pool
- the pool
- the multiplied pool
- the released amount
- the own portion of the released amount[1]
- the unreleased amount
- the own period payoff
- the balance of her account (payoffs of all past periods)

Every **B-participant** learns about

- her secure income
- the pool
- the multiplied pool

[1] In PGG instructions: „the own portion of the multiplied pool"



- the released amount
- every A-participant’s portion of the released amount
- the unreleased amount
- the own period payoff
- the balance of her account (payoffs of all past periods)

Neither the A-participants nor the B-participant will be informed about the A-participants’ individual contributions to the pool.

Period Payoff

The A- and B-participants’ period payoffs are calculated as follows:

A-participant’s payoff	B- Participant’s payoff
Endowment	Secure income
- Contribution	+ Unreleased amount
+ Portion of released amount	
==========================	========================
Period payoff	Period payoff

Please take a look at the following table.[2]

Example 1		Example 2	
Multiplied pool	**120 ECU**	Multiplied pool	**60 ECU**
a)		**a)**	
Released amount	**120 ECU**	Released amount	**60 ECU**
Every A-participant receives	**30 ECU**	Every A-participant receives	**15 ECU**
The B-participant receives	**0 ECU**	The B-participant receives	**0 ECU**
All Participants (A and B) have a payoff of **30 ECU**.		A-participants’ payoffs vary between **17 ECU** and **25 ECU**. The B-participant has a payoff of **30 ECU**.	
b)		**b)**	
Released amount	**80 ECU**	Released amount	**20 ECU**
Every A-participant receives	**20 ECU**	Every A-participant receives	**5 ECU**
The B-participant receives	**40 ECU**	The B-participant receives	**40 ECU**
All A-participants have a payoff of **20 EC**U. The B-participant has a payoff of **70 ECU.**		A-participants’ payoffs vary between **5 ECU** and **15 ECU**. The B-participant has a payoff of **70 ECU.**	

Example Calculations

To make sure that all participants have understood the instructions, we ask you to make some example calculations on your computer. It does not matter if you need several attempts to answer the questions.

[2] Example 1 (PGG instructions): Multiplied pool = 80 ECU; Every participant receives 20 ECU from the pool; All participants have a payoff of 20 ECU; Example 2 (PGG instructions): Multiplied pool = 40 ECU; Every participant receives 10 ECU from the pool; participants have a payoff between 10 ECU and 20 ECU

Theoretical Analysis of the Public Trust Game

In this Online Appendix, we provide detailed proofs for the theoretical results discussed in section 3 of our paper. In particular, we analyze infinitely repeated interaction and reciprocity concerns. We are aware that reciprocity and effects from repeated interaction might work together in our setup. To keep keep the analysis simple, we examine them separately.

Consider the PTG among five players $i = \{1, \ldots, 5\}$, where agents 1 to 4 are the contributors and agent 5 is the administrator. Contributors have similar endowments $w_i \equiv w$, $i = \{1, \ldots, 4\}$, while the administrator has an endowment $w_5 > w$. Contributions in period t are $(m_{1t}, \ldots, m_{4t})$ and $M_t = r \sum_1^4 m_{it}$ is the pool in period t. Furthermore, let $\gamma_t \in [0, 1]$ be the share of the pool kept by the administrator. Denote by x_{it} the agents' payoffs in period t. It holds that

$$\begin{aligned} x_{it} &= w - m_{it} + \frac{1}{4} r \sum_{j=1}^{4} m_{jt} - \frac{1}{4} r\gamma_t \sum_{j=1}^{4} m_{jt}, \quad i = \{1, \ldots, 4\}, \\ &= w - \left(1 - \frac{1}{4} r(1-\gamma_t)\right) m_{it} + \frac{1}{4} r(1-\gamma_t) \sum_{j \neq i} m_{jt}, \qquad (1) \\ x_{5t} &= w_5 + r\gamma_t \sum_{j=1}^{4} m_{jt}. \qquad (2) \end{aligned}$$

In any equilibrium of the one-shot PTG, contributions are zero if all agents are rational payoff maximizers and this is common knowledge among them. Consequently, any subgame perfect equilibrium of the finitely repeated game implies zero contributions in every period. The same is true if the administrator is absent and the contributors play a standard PGG with an efficiency factor of r The predictions change if the PTG is infinitely repeated (or the end is unknown) or if agents have reciprocity concerns.

Repeated Interaction

Let us consider repeated interactions and assume that participants share a common discount factor δ. Because the infinitely repeated PTG has a continuum of equilibria (including those equilibria with zero contributions)[1], we focus on conditions on δ under which full cooperation can be sustained in an equilibrium of the repeated game.

Let us first consider a standard Public Goods Game (PGG) without an administrator. The efficiency factor is r. It is well known that, if δ is sufficiently high, the following grim trigger strategies constitute an equilibrium of the infinitely repeated PGG:

$$m_{it} = \begin{cases} w & \text{if } \; m_{jt-1} = w \; \forall j = \{1, \ldots 4\} \\ 0 & \text{else.} \end{cases} \qquad (3)$$

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

This is summarized in the following lemma.

Lemma 1 (Infinitely Repeated PGG) *The infinitely repeated PGG has an equilibrium where all agents adopt the grim trigger strategy (3) iff* $\delta \geq \delta_{PGG} = \frac{4-r}{3r}$.

Proof. In the PGG there is no administrator (i.e. $\gamma_t = 0$). It follows from (1) that

$$x_{it}(m_{it}) = w - \left(1 - \frac{1}{4}r\right) m_{it} + \frac{1}{4}r \sum_{j \neq i} m_{jt}. \tag{4}$$

Now consider player i's decision to either choose the grim trigger strategies (3) or to deviate from it given that all other players $j \neq i$ follow these strategies. Contributing w in a given round (and consequently planning to do the same in all upcoming periods) yields a net present value of

$$\pi_i(w) \quad = \quad \sum_{t=0}^{\infty} \delta^t rw = \frac{rw}{1-\delta}.$$

Deviation to $m_{it} = 0$ in a given period implies future zero contributions by all agents and yields

$$\begin{aligned} \pi_i(0) \quad &= \quad rw + (1 - \frac{1}{4}r)w + \delta \sum_{t=0}^{\infty} \delta^t w, \\ &= \quad rw + (1 - \frac{1}{4}r)w + \frac{\delta}{1-\delta} w. \end{aligned}$$

Cooperation is sustainable if $\pi_i(w) \geq \pi_i(0)$, i.e.

$$\frac{rw}{1-\delta} \geq rw + (1 - \frac{1}{4}r)w + \frac{\delta}{1-\delta} w \quad \Leftrightarrow \quad \delta \geq \frac{4-r}{3r}.$$

■

In the PTG, the incentives of contributors to cooperate depend not only on the discount factor, but also on the level of rent extraction by the administrator. Extraction rates are naturally constrained by the potential impact on future profits: an administrator who chooses full rent extraction early in the game could trigger zero future contributions and, thereby, severely limit her further opportunities to generate payoffs. In our analysis, we focus on the question under which levels of rent extraction cooperation can be sustained in equilibrium and how the possibility of rent extraction affects the critical discount factor. For simplicity, we assume that the level of rent extraction is constant $\gamma_t = \hat{\gamma}$ and contributors expect the administrator to choose $\hat{\gamma}$ throughout all stages. Let us consider the following grim trigger strategies:

$$m_{it} = \begin{cases} w & \text{if } m_{jt-1} = w \; \forall j = \{1, \ldots 4\}, \text{ and } \gamma_{t-1} = \hat{\gamma} \\ 0 & \text{else,} \end{cases} \tag{5}$$

$$\gamma_t = \begin{cases} \hat{\gamma} & \text{if } m_{jt-1} = w \; \forall j = \{1, \ldots 4\}, \text{ and } \gamma_{t-1} = \hat{\gamma} \\ 1 & \text{else.} \end{cases} \tag{6}$$

The following proposition states the lowest possible discount factor that sustains full cooperation by the contributors and the associated level of rent extraction by the administrator.

Proposition 1 (Infinitely Repeated PTG) *The infinitely repeated PTG has an equilibrium where all agents adopt the grim trigger strategies (5) and (6) iff* $\delta \geq \delta_{PTG} = \sqrt{\frac{4}{3r} + \frac{1}{36}} - \frac{1}{6}$. *In this equilibrium it holds that* $\hat{\gamma} = \hat{\gamma}^* = \frac{7}{6} - \sqrt{\frac{4}{3r} + \frac{1}{36}}$.

Proof. Suppose that all players $j \neq i$ play the proposed grim trigger strategies (5) and (6). A contributor i's profit from cooperation in a given period t is

$$x_{it}(w) = w - \left(1 - \frac{1}{4}r(1-\hat{\gamma})\right) w + \frac{1}{4}r(1-\hat{\gamma})3w = r(1-\hat{\gamma})w,$$

and her period-profit from deviation is

$$x_{it}(0) = w + \frac{3}{4}r(1-\hat{\gamma})w.$$

For the administrator it holds that

$$\begin{aligned} x_{5t}(\hat{\gamma}) &= w_5 + 4r\hat{\gamma}w, \\ x_{5t}(1) &= w_5 + 4rw. \end{aligned}$$

The net present value of cooperation for a contributor i is

$$\pi_i(w) = \sum_{t=0}^{\infty} \delta^t r(1-\hat{\gamma})w = \frac{r(1-\hat{\gamma})w}{1-\delta}.$$

Deviation to $m_{it} = 0$ in a given period implies zero contributions in the future and yields

$$\pi_i(0) = w + \frac{3}{4}r(1-\hat{\gamma})w + \delta \sum_{t=0}^{\infty} \delta^t w = w + \frac{3}{4}r(1-\hat{\gamma})w + \frac{\delta}{1-\delta}w.$$

The administrator's net present value of choosing $\hat{\gamma}$ is

$$\pi_5(\hat{\gamma}) = \sum_{t=0}^{\infty} \delta^t \left(w_5 + 4r\hat{\gamma}w\right) = \frac{w_5 + 4r\hat{\gamma}w}{1-\delta}.$$

Deviation to $\gamma_t = 1$ in a given period implies zero contributions in the future and yields

$$\pi_5(1) = w_5 + 4rw + \delta \sum_{t=0}^{\infty} \delta^t w_5 = w_5 + 4rw + \frac{\delta}{1-\delta} w_5.$$

Cooperation is sustainable if contributors cooperate and the administrator refrains from full rent extraction. Contributors cooperate if $\pi_i(w) \geq \pi_i(0)$, i.e.

$$\frac{r(1-\hat{\gamma})w}{1-\delta} \geq w + \frac{3}{4}r(1-\hat{\gamma})w + \frac{\delta}{1-\delta}w \quad \Leftrightarrow \quad \delta \geq \frac{4-r(1-\hat{\gamma})}{3r(1-\hat{\gamma})}.$$

The administrator refrains from full rent extraction if $\pi_5(\hat{\gamma}) \geq \pi_5(1)$, i.e.

$$\frac{w_5 + 4r\hat{\gamma}w}{1-\delta} \geq w_5 + 4rw + \frac{\delta}{1-\delta}w_5 \quad \Leftrightarrow \quad \delta_5 \geq 1-\hat{\gamma}.$$

Let us define the critical discount factor of contributors and the administrator as $\delta_i(\hat{\gamma}) = \frac{4-r(1-\hat{\gamma})}{3r(1-\hat{\gamma})}$ and $\delta_5(\hat{\gamma}) = 1-\hat{\gamma}$. Noting that $\frac{\partial\delta_i}{\partial\hat{\gamma}} > 0$ and $\frac{\partial\delta_5}{\partial\hat{\gamma}} < 0$, we can identify the level of $\hat{\gamma}$, associated with the lowest possible discount factor that sustains cooperation by all parties, by solving

$$\frac{4-r(1-\hat{\gamma}^*)}{3r(1-\hat{\gamma}^*)} = 1-\hat{\gamma}^*.$$

We obtain $\hat{\gamma}^* = \frac{7}{6} - \sqrt{\frac{4}{3r} + \frac{1}{36}}$ and $\hat{\delta} = \sqrt{\frac{4}{3r} + \frac{1}{36}} - \frac{1}{6}$. ■

The analysis points to an important tradeoff in the repeated PTG: the level of anticipated rent extraction affects the incentives to cooperate and, thus, future rent extraction possibilities. Consider the case of our experiment ($r = 3$). Whereas a critical discount factor of $\delta \geq \frac{1}{9} \approx 0.11$ sustains cooperation in the PGG, the critical discount factor in the PTG is higher: $\delta_{PTG} = \frac{\sqrt{17}-1}{6} \approx 0.52$. The associated level of rent extraction is $\hat{\gamma} = \frac{7-\sqrt{17}}{6} \approx 0.48$. Rent extraction affects the efficiency factor and, hence, diminishes the scope for cooperation.

Comparing the infinitely repeated versions of the PTG and the PGG, we find that the critical discount factor in the PTG is identical to the critical discount factor in the PGG with an exogenously given efficiency factor $\hat{r} = (1-\hat{\gamma})r$. The analysis of the repeated game suggests similar levels of cooperation in the PTG and the reference PGG that we analyze in our experimental setup.

Reciprocity Concerns

To shed light on how concerns for reciprocity might affect play in the PTG, we apply Dufwenberg and Kirchsteiger (2004) *Theory of Sequential Reciprocity* to the (one shot) stage game. Dufwenberg and Kirchsteiger assume that individuals derive utility from material payoffs and from reciprocity. The utility is

$$U_i(x_1, \ldots, x_5) = x_i + Y_i \sum_{j \neq i} (\kappa_{ij}\lambda_{iji}), \tag{7}$$

where x_i is the agent's own material payoff, Y_i is her sensitivity for reciprocity, κ_{ij} is i's kindness to agent j, and λ_{iji} is i's belief about j's kindness to her. Both terms build on i's beliefs about

j's behavior, assuming that j' behavior coincides with the belief in equilibrium. κ_{ij} is the payoff that i gives to j minus the average of the minimum and maximum payoff she could give to j. λ_{iji} denotes i's belief about her payoff from j minus the average of the minimum and maximum payoff that j could give to i. We can establish the following proposition:

Proposition 2 (Sequential Reciprocity Equilibrium) *Suppose agents are sensitive to reciprocity as in Dufwenberg and Kirchsteiger (2004).*

(i) Iff $Y_5 \geq \frac{44}{rw(4\sqrt{15}-3)}$ *and* $Y_i \geq \frac{16(1-\frac{1}{4}(1-\gamma)r)}{3r^2w[\frac{1}{2}(1-\gamma)^2+3\gamma(1-2\gamma)]}$ *for all* $i = \{1, \ldots, 4\}$ *a Sequential Reciprocity Equilibrium exists where* $\gamma = \frac{1}{4} + \frac{1}{Y_5 rw}$ *and* $m_i = w$ *for all* $i = \{1, \ldots, 4\}$.

(ii) In a reciprocity equilibrium with full contributions the extraction rate γ *is at least* $\frac{1}{4}$ *and at most* $\frac{1}{11}(2+\sqrt{15}) \approx 0.53$.

Proof. For our analysis we need κ_{i5}, κ_{ij} κ_{5i}, λ_{i5i}, λ_{iji}, and λ_{5i5}. To establish under which conditions a Sequential Reciprocity Equilibrium with full cooperation exists, we study one contributor i's utility and the administrator's utility, assuming that all other contributors choose $m_j = w$. For contributor i's utility from reciprocity we define $j = \{1, \ldots, 4\}$ and $j \neq i$. For the administrator's utility from reciprocity, j denotes the group of contributors. For contributor i we get

$$\begin{aligned}
\kappa_{i5} &= \gamma r m_i - \frac{1}{2}[\gamma r w] \\
&= \gamma r(m_i - \frac{1}{2}w),
\end{aligned}$$

$$\begin{aligned}
\lambda_{i5i} &= \frac{1}{4}(1-\gamma)r(m_i+3w) - \frac{1}{2}\left[\frac{1}{4}r(m_i+3w)\right] \\
&= \frac{1}{4}(m_i+3w)r(\frac{1}{2}-\gamma),
\end{aligned}$$

$$\begin{aligned}
\kappa_{ij} &= \frac{1}{4}(1-\gamma)r(m_i+3w) - \frac{1}{2}\left[\frac{3}{4}(1-\gamma)rw + (1-\gamma)rw\right] \\
&= \frac{1}{4}(1-\gamma)r(m_i - \frac{1}{2}w),
\end{aligned}$$

$$\begin{aligned}
\lambda_{iji} &= \frac{1}{4}(1-\gamma)r(m_i+3w) - \frac{1}{2}\left[\frac{1}{4}(1-\gamma)r(m_i+2w) + \frac{1}{4}(1-\gamma)r(m_i+3w)\right] \\
&= \frac{1}{8}(1-\gamma)rw.
\end{aligned}$$

For the administrator we get

$$\begin{aligned}
\kappa_{5j} &= (1-\gamma)rw - \frac{1}{2}[rw] \\
&= (\frac{1}{2}-\gamma)rw,
\end{aligned}$$

$$\begin{aligned}
\lambda_{5j5} &= 4\gamma rw - \frac{1}{2}[3\gamma rw + 4\gamma rw] \\
&= \frac{1}{2}\gamma rw.
\end{aligned}$$

The administrator's utility is then

$$\begin{aligned}
U_5(\gamma) &= w_5 + 4\gamma rw + Y_5\left[4\left(\left(\frac{1}{2}-\gamma\right)rw\right)\left(\frac{1}{2}\gamma rw\right)\right] \\
&= w_5 + 4\gamma rw + Y_5\left[\gamma r^2 w^2(1-2\gamma)\right].
\end{aligned}$$

Reciprocity concerns cannot induce the administrator to abstain from rent extraction. Recall from the experimental design that the administrator could choose any level of rent extraction $\gamma \in [0,1]$. Because for $\gamma = 0$ no other player can affect the administrator's payoff, her belief about the kindness of player j towards her (λ_{5j5}) must equal to zero if she chooses $\gamma = 0$. In this case, the model implies that the administrator gains no utility from being kind or unkind to the contributors. Differentiation of $U_5(\gamma)$ with respect to γ yields

$$\begin{aligned}
\frac{\partial U_5}{\partial \gamma} &= 4rw + Y_5 r^2 w^2(1-4\gamma) \geq 0 \qquad (8) \\
\Leftrightarrow \quad \gamma &\leq \tfrac{1}{4} + \tfrac{1}{Y_5 rw} \quad \text{or} \quad Y_5 \geq \tfrac{4}{rw(4\gamma-1)}.
\end{aligned}$$

Thus the administrator extracts at least one fourth of the pool (if Y_5 tends to infinity) and extracts more than half of the pool if she has almost no reciprocity concerns, i.e. $Y_5 < \frac{2}{15}$.

Contributor i's utility and the first order condition are given by

$$\begin{aligned}
U_i(m_i, w, \gamma) &= w - m_i + \frac{1}{4}(1-\gamma)r(m_i + 3w) \qquad (9) \\
&\quad + Y_i\left[3\left(\frac{1}{4}(1-\gamma)r(m_i - \frac{1}{2}w)\right)\left(\frac{1}{8}(1-\gamma)rw\right)\right. \\
&\quad \left. + \left(\gamma r(m_i - \frac{1}{2}w)\right)\left(\frac{1}{4}(m_i+3w)r(\frac{1}{2}-\gamma)\right)\right] \\
&= w + m_i\left(\frac{1}{4}(1-\gamma)r - 1\right) + \frac{3}{4}(1-\gamma)rw \\
&\quad + Y_i\left[\frac{3}{32}(1-\gamma)^2 r^2 w(m_i - \frac{1}{2}w) + \frac{1}{4}\gamma r^2(m_i - \frac{1}{2}w)(m_i+3w)(\frac{1}{2}-\gamma)\right],
\end{aligned}$$

$$\frac{\partial U_i}{\partial m_i} \quad = \quad \frac{1}{4}(1-\gamma)r - 1 + Y_i\left[\frac{3}{32}(1-\gamma)^2 r^2 w + \frac{1}{4}\gamma r^2(\frac{1}{2}-\gamma)(2m_i + \frac{5}{2}w)\right] \geq 0.$$

The critical value of the contributors' sensitivity to reciprocity depends on the level of contributions. In any equilibrium where all contributors choose $m_i = w$, the FOC simplifies to

$$\begin{aligned} \frac{\partial U_i}{\partial m_i} &= \frac{1}{4}(1-\gamma)r - 1 + Y_i\left[\frac{3}{32}(1-\gamma)^2 r^2 w + \frac{1}{4}\gamma r^2(\frac{1}{2}-\gamma)(2w + \frac{5}{2}w)\right] \geq 0 \\ \Leftrightarrow Y_i &\geq \frac{16(1-\frac{1}{4}(1-\gamma)r)}{3r^2 w[\frac{1}{2}(1-\gamma)^2 + 3\gamma(1-2\gamma)]}. \end{aligned}$$

Note that a Sequential Reciprocity Equilibrium where all contributions equal the endowment can only be established if the extraction rate γ is not too high. If $\gamma \to \frac{1}{11}(2+\sqrt{15}) \approx 0.53$, the critical sensitivity for reciprocity (Y_i) approaches infinity. However, because of reciprocal behavior towards other contributors, there can be non-zero contributions despite unkind administrator behavior, i.e. $\gamma > \frac{1}{2}$.

We finally look into the administrator's minimal sensitivity for reciprocity that ensures an extraction of at most $\gamma = \frac{1}{11}(2+\sqrt{15})$, which is the highest possible extraction rate for which non-zero contributions in equilibrium are possible. Substitution of this value of γ into the second equation in (8) yields a minimal sensitivity for reciprocity of $Y_{5,min} = \frac{44}{rw(4\sqrt{15}-3)}$. ■

Proposition 3 (Administrator vs. No Administrator) *Suppose agents are sensitive to reciprocity as in Dufwenberg and Kirchsteiger (2004).*

(i) If in the PTG extraction behavior is kind (i.e. $0 < \gamma < \frac{1}{2}$), cooperation is easier to sustain in the PTG than in a reference PGG where agents face the same true efficiency factor but no administrator.

(ii) If in the PTG the extraction behavior is unkind (i.e. $\gamma > \frac{1}{2}$), cooperation is easier to sustain in a reference PGG where agents face the same true efficiency factor but no administrator.

Proof. Without an administrator, contributor i's utility is

$$\begin{aligned} U_i(m_i, w, \gamma) = w + m_i\left(\frac{1}{4}(1-\gamma)r - 1\right) + \frac{3}{4}(1-\gamma)rw \\ +Y_i\left[\frac{3}{32}(1-\gamma)^2 r^2 w(m_i - \frac{1}{2}w)\right], \end{aligned}$$

which is the utility in (9) without the reciprocity utility from interaction with the administrator. The FOC is

$$\frac{\partial U_i}{\partial m_i} = \frac{1}{4}(1-\gamma)r - 1 + Y_i\left[\frac{3}{32}(1-\gamma)^2 r^2 w\right] \geq 0 \qquad \Leftrightarrow \qquad Y_i \geq \frac{32(1-\frac{1}{4}(1-\gamma)r)}{3(1-\gamma)^2 r^2 w}.$$

To see under which conditions cooperation is easier to sustain in the PTG than in the PGG (holding the true efficiency factor constant), we compare the critical values of γ_i for both games

$$\begin{aligned}\frac{32(1-\frac{1}{4}(1-\gamma)r)}{3(1-\gamma)^2r^2w} &\leq \frac{16(1-\frac{1}{4}(1-\gamma)r)}{3r^2w[\frac{1}{2}(1-\gamma)^2+3\gamma(1-2\gamma)]}\\ \frac{1}{\frac{1}{2}(1-\gamma)^2} &\leq \frac{1}{\frac{1}{2}(1-\gamma)^2+3\gamma(1-2\gamma)}\\ 3\gamma(1-2\gamma) &\leq 0.\end{aligned}$$

■

If cooperation is sustained depends on the administrator's kindness. Whenever her action is kind (i.e. $0 < \gamma < \frac{1}{2}$), it is easier to sustain cooperation in the game with an administrator. Whenever her action is unkind, it is easier to sustain cooperation in the absence of an administrator.[2] The reason is that the administrator's kindness adds to the motivational effect of other contributors' kindness.

References

Dufwenberg, M. and Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, **47** (2), 268–298.

Friedman, J. W. (1971). A non-cooperative equilibrium for supergames. *Review of Economic Studies*, **38** (113), 1–12.

[2] Note that in case of $\frac{1}{2}(1-\gamma)^2r < 3\gamma(1-2\gamma)$ no Sequential Reciprocity Equilibrium exists because the extraction rate is too high. In this case, contributors expect excessive extraction by the administrator and therefore would not contribute if there is an administrator.