

Gallier, Carlo; Sturm, Bodo

**Working Paper**

## The ratchet effect in social dilemmas

ZEW Discussion Papers, No. 20-015

**Provided in Cooperation with:**

ZEW - Leibniz Centre for European Economic Research

*Suggested Citation:* Gallier, Carlo; Sturm, Bodo (2020) : The ratchet effect in social dilemmas, ZEW Discussion Papers, No. 20-015, ZEW - Leibniz-Zentrum für Europäische Wirtschaftsforschung, Mannheim

This Version is available at:

<https://hdl.handle.net/10419/215674>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

DISCUSSION

// NO.20-015 | 04/2020

# DISCUSSION PAPER

// CARLO GALLIER AND BODO STURM

## The Ratchet Effect in Social Dilemmas

# The Ratchet Effect in Social Dilemmas

Carlo Gallier<sup>1</sup> and Bodo Sturm<sup>1,2</sup>

<sup>1</sup> ZEW – Leibniz Centre for European Economic Research (ZEW)

<sup>2</sup> Leipzig University of Applied Sciences (HTWK Leipzig)

**This version:** April, 2020

**Abstract:** In this paper, we investigate whether dynamic incentive schemes lead to a ratchet effect in a social dilemma. We test whether subjects strategically restrict their contribution levels at the beginning of a cumulative public goods game in order to avoid high obligations in the future and how this affects efficiency. The incentive schemes prescribe that individual contributions have to be at least as high as, or strictly higher than, contributions in the previous period. We observe a substantial and statistically significant ratchet effect. Participants reduce their public good contribution levels at the beginning of the game, anticipating that higher contributions imply higher minimum contribution levels in the future, which increases the risk of being exploited by free-riders. While the dynamic incentive schemes lead to increasing contribution levels over the course of the game, this increase is not strong enough to compensate the efficiency losses at the beginning.

**Keywords:** Public goods; dynamic incentives; minimum contribution rules; ratchet effect

**JEL:** C72, C92, H41

**Acknowledgement:** We are grateful for valuable comments provided by audiences at the Helmholtz Centre for Environmental Research, Leipzig University, and University of Mannheim. Financial support by the German Federal Ministry of Education and Research (FKZ 01LA1825A/C) is gratefully acknowledged. We thank the MaXLab team, in particular Justine Richter and Lucas Tautz, for their support in conducting the experiment. This experiment has been registered and reviewed by the institutional review board of the German Association for Experimental Economics e.V. (<https://gfew.de/ethik>).

**Correspondence:** Carlo Gallier, ZEW Mannheim, L 7, 1, 68161 Mannheim, Germany, [carlo.gallier@zew.de](mailto:carlo.gallier@zew.de)

## 1 Introduction

Minimum contribution levels (MCLs) are a widely used mechanism to overcome free-riding incentives in social dilemma situations, where actions by rational and self-interested individuals do not align with common group interests. For example, governments impose minimum levels regarding the energy efficiency in household appliances and the sorting of waste or levy taxes as a minimum monetary contribution to environmental public goods.<sup>1</sup> A variety of economic experiments (e.g. Andreoni 1993; Eckel et al. 2005; Gronberg et al. 2012) shows that MCLs typically do not completely crowd out voluntary contributions, thus effectively increasing public good provision levels.

A common feature of the aforementioned studies is that the implemented minimum contributions are static and agents' current decisions do not have dynamic implications for their future MCLs. In many cooperation problems, though, minimum contributions are subject to dynamic incentive schemes where current decisions act as a benchmark for future contributions. The Paris Agreement under the United Nations Framework Convention on Climate Change (UNFCCC) is a prominent example. After years of unfruitful negotiations to reach a globally binding "top-down" climate agreement, in 2015 Parties of the UNFCCC agreed to set out a "bottom-up" global action plan with the objective to avoid dangerous climate change and limit global warming to well below 2°C above pre-industrial levels (Green et al. 2014). From a current perspective, however, countries' nationally determined contributions are insufficient to reach the agreed long-term climate objective (Rogelj et al. 2016). To ensure that this gap will be closed and to achieve the two-degree target, the agreement contains a dynamic component in form of an incremental contribution scheme that prescribes that nationally determined contributions have to increase over time. The agreement requires a global stocktake of nationally determined contributions every five years combined with a gradual ratcheting-up of ambitions. When countries set their climate targets at the start of each five-year period, they already know that they will have to increase their ambitions in the next period. High hopes are placed on this process of persistently increasing mitigation efforts, called "*ratcheting*", that should help to increase contributions to the global public good of climate change

---

<sup>1</sup> The interaction between private and public climate change mitigation efforts might serve as an important example. Voluntary contributions to climate protection (e.g. Löscher et al. 2013; Diederich and Goeschl 2014) are made given public regulations, like carbon prices, which could be interpreted as minimum climate change mitigation effort.

mitigation (e.g. Young 2016; Holz et al. 2018).<sup>2</sup> However, ratcheting might backfire since agents could anticipate that their current contributions to the public good increase their minimum contribution level in the future. The effect of such a dynamic incentive scheme that enacts public good contribution levels to increase over time, thus strikes us as an important but understudied question in public economics, and in environmental economics in particular.

It is exactly this effect of dynamic incentive schemes in a social dilemma situation that we address in this paper. This relates our paper to the economic literature on dynamic incentive schemes in general and the effect of incremental contribution schemes in cooperation problems in particular. In general, the ratchet principle (or simply “ratcheting”) implies the use of current performance to determine future economic targets.<sup>3</sup> This principle, which is an almost universal feature of economic planning, creates a dynamic incentive problem (e.g. Weitzman 1980), as higher rewards from better current performance must be weighed against the future assignment of more ambitious targets. Rational agents potentially restrict current efforts in order to avoid increased obligations in the future. So far, there is a substantial experimental literature on the effect of ratcheting in principal-agent problems in labor market contexts (e.g. Chaudhuri 1998; Charness et al. 2011). A key finding, the so called *ratchet effect*, refers to a situation where agents compensated based on productivity strategically “masquerading” as low-output workers relative to their true capacity, as the agents rationally anticipate that current high levels of output will be met with increased future requirements by the principal, who does not have complete information about agents’ skills (e.g. Charness et al. 2011; Cardella and Depew 2018). In the context of piece-rate compensation, Charness et al. (2011), for instance, find strong evidence for a ratchet effect. In an environment without any competition among workers, they observe that participants systematically restrict their current output levels in order to avoid reduced rewards in the future. Amano and Ohashi (2018) use field data from the Japanese television market in order to investigate the effect of regulation that holds that televisions sold during a future target year must be at least as energy efficient as the most efficient product sold in the base year. The result is that firms hold back on the energy

---

<sup>2</sup> Ratcheting is an important element of the Paris Agreement and the agreed “pledge-and-review” process: “As nationally determined contributions to the global response to climate change, all Parties are to undertake and communicate efforts [...] the efforts of all Parties will present progression over time [...]” and “Each Party’s successive nationally determined contribution will represent a progression beyond the Party’s then current nationally determined contribution and reflect its highest possible ambition, [...]” (UNFCCC 2015, Article 3 and 4.3).

<sup>3</sup> The term “ratchet principle” goes back to Berliner’s treatment on managerial behavior in a centralized planning system. See Berliner (1957), pp. 78-80.

efficiency of their products in order to be able to continue selling less efficient products for the foreseeable future.<sup>4</sup>

We investigate whether a ratchet principle, that prescribes that cooperation has to increase over time, creates an incentive problem in a repeated social dilemma by inducing incentives to strategically restrict cooperation at the beginning in order to avoid high obligations in the future. We implement dynamic MCLs into a cumulative public goods game (e.g. Duffy et al. 2007) where agents have the option to repeatedly contribute to a public good in discrete time with cumulating benefits. The MCLs we analyze prescribe that agents have to contribute as much (*weak* ratcheting) or even strictly more (*strong* ratcheting) than what they have contributed in the period before. Both ratcheting treatments are compared to a voluntary contribution mechanism without a dynamic contribution scheme.

This relates our paper to the literature on incremental contribution schemes in social dilemmas. Schelling (1960) suggests small, sequential, and contingent commitments in order to limit the extent to which agents expose themselves to the risk of being free ridden, establish trust, and, finally, foster cooperation. Support for Schelling's hypothesis comes from the literature on static versus dynamic cooperation problems (e.g. Marx and Matthews 2000; Duffy et al. 2007) as well as cooperation problems in discrete and continuous time (e.g. Friedman and Oprea 2012; Oprea et al. 2014). Having multiple contribution decisions and the opportunity to (rapidly) adjust contribution levels depending on others' contributions facilitates cooperation as individuals only need to sacrifice small contributions to test how cooperative others are.

To the best of our knowledge, Dorsey (1992) is the first paper that investigates the effect of a ratchet principle that prescribes that participants cannot adjust their commitments to contribute to a public good downwards. In each period of his experiment participants have the opportunity to pledge their public good contribution levels and continuously update their pledges in real-time. The actual contribution level is given by the pledge at the end of each period. He finds that public good contribution levels are higher when participants can only increase their pledges compared to when pledges can be in- or decreased. Also Kurzban et al. (2001) modify how pledges in a public goods game in continuous time can be constantly revised, i.e., whether pledges can be adjusted up- and downwards, or only upwards. In addition, they vary whether participants have access to full information about others' pledges, or only receive information about the lowest or highest pledge

---

<sup>4</sup> Macartney (2016) detects ratcheting also in the context of education accountability schemes in a study using field data.

per group. In line with Dorsey (1992), they find that a ratchet principle, that prescribes that participants who raised their pledges to a certain level are unable to reverse their decision, facilitates cooperation, especially when participants are only informed about the smallest pledge per group during the course of a period.<sup>5</sup> Ye et al. (2020) use a discrete coordination game to test whether gradually increasing thresholds can facilitate coordination within groups. If all participants contribute the required threshold, they do not only receive their contribution back but also gain an extra return; if not, participants finish the period with their remaining endowment. As treatment variable, they vary the level of thresholds across periods. They find that groups who start at low but gradually increasing thresholds are especially successful. One possible explanation could be that high thresholds at the beginning of the game increase the extent to which participants expose themselves to the risk of coordination failure, since participants have to sacrifice high contributions in order to test how others response.

Within the literature on ratcheting in social dilemmas, our experiment differs from those previous studies, in particularly Dorsey (1992), Kurzban et al. (2001), and Ye et al. (2020). In addition to our cooperation game, which differs from Ye et al. (2020), our experiment includes dynamic contribution schemes over discrete periods with cumulating benefits. Dorsey (1992) as well as Kurzban et al. (2001) rely on public goods games in continuous time where commitments can be constantly revised in each period of the game and focus at participants' final contributions at the end of each period. Furthermore, in both studies, decisions in one period do not depend on the previous period. All this precludes them from investigating whether a ratchet principle, that prescribes that contributions have to increase over time, induces a ratchet effect in that sense that participants strategically restrict cooperation at the beginning in order to avoid increased obligations in the future. In order to test for a ratchet effect in a social dilemma, our public goods game is played in discrete time with cumulating benefits and contributions that depend on the period before. In addition, we use different ratchet principles that prescribe that contributions cannot be adjusted downwards or have to increase to investigate how the stringency of the

---

<sup>5</sup> Similar results are obtained in other cooperation problems played in continuous time, such as threshold public goods games (Diev and Hichri 2008) and public goods games with different aggregation technologies (Güth et al. 2002). Fellner et al. (2003) made an early attempt to relate the ratchet effect to the private provision of public goods in discrete time. In their experiment, the cost of the public good provision depend on previous contribution levels in that sense that it is cheaper to provide the public good below than above the previous level. They observe that public good provision with such a dynamic production technology is lower than compared to a standard linear production technology.

mechanism affects participants' incentives to be reluctant to contribute at the beginning of the game.

We complement the existing literature by testing if dynamic MCLs lead to such a ratchet effect in a cumulated public goods game in discrete time and how this affects efficiency. While a continuous public goods game does adequately reflect many types of social dilemmas, such as fundraising activities where donations are made asynchronously in continuous time, we focus on more institutionalized social dilemma situations such as international climate policy where discrete political decisions that cannot be constantly revised are prevalent. The global stocktake within the UNFCCC, for instance, stipulates a simultaneous and periodical review of the collective climate change mitigation efforts every five years to assess the collective progress towards achieving the purpose of the Paris Agreement.

In the theory part of the paper, we show that the cumulated public good provision level is not expected to be affected by ratcheting (neither by weak nor strong ratcheting). Transferred to the corresponding contribution levels per period, this implies that the equilibrium cumulated contribution level is divided equally across all periods and is therefore equal to the Nash equilibrium in our baseline treatment, as well as in our *weak* ratcheting treatment. In *strong* ratcheting, however, the optimal cumulated provision level is predicted to be divided differently over the course of the game. While contributions in *strong* ratcheting are predicted to be below the Nash equilibrium at the beginning of the game, they have to increase – by design – from period-to-period and exceed the contribution in the Nash equilibrium at the end of the game. Consequently, the aggregated contribution level is not reached at lowest cost in *strong* ratcheting.

Contrary to our theoretical predictions, we find a distinct ratchet effect. Ratcheting clearly affects participants' cumulated public good provision levels. Both *weak* and *strong* ratcheting substantially reduce public good contributions at the beginning of the game compared to the baseline. While contributions increase over the course of the game in both ratcheting treatments, this increase is not strong enough to compensate for the initial loss in efficiency. Consequently, *weak* and *strong* ratcheting reduce cumulated public good provision levels and have a negative effect on efficiency. Thus, our results challenge the idea that ratcheting is a promising mechanism in order to solve or at least mitigate the social dilemma situation which global climate policy is confronted with. While already from a theoretical perspective one should be skeptical whether ratcheting increases contributions in a public goods game, our empirical results suggest that ratcheting on average even



reduces contributions to the public good. Our results thus complement the existing critical evaluation of the Paris Agreement from the economic perspective (e.g. Cooper et al. 2017).

Furthermore, the analysis at the individual level suggests that participants' *fear of the risk of being exploited* is especially strong at the beginning of the ratcheting treatments, which helps to explain the low contribution levels at the beginning of the game. In *weak* and *strong* ratcheting, positive public good contribution levels determine a path dependency that cannot simply be adjusted downwards in the next decision. Therefore, the risk of being exploited by free-riders is more pronounced with ratcheting and, consequently, participants are especially reluctant at the beginning of the game and contribute significantly less in the ratcheting treatments than in the baseline.

Section 2 of this paper presents the experimental design, the theoretical predictions, and the procedure, we discuss our results in Section 3, and we provide our conclusion in Section 4.

## 2 Experimental design, hypotheses and procedure

### 2.1 Game and treatments

Our experimental design extends the standard linear public goods game in order to provide a careful test of a ratchet effect in a social dilemma. A necessary condition for detecting a ratchet effect is that the Nash equilibrium is interior. In addition, based on dynamic voluntary contribution games (e.g. Duffy et al. 2007), subjects' payoffs depend on the cumulated benefits from the public goods game. These features of our experiment present a break with the majority of public goods games conducted in the past (e.g. Ledyard 1995; Zelmer 2003; Chaudhuri 2010).

We define groups of  $n$  identical members,  $i \in \{1, \dots, n\}$ . In all treatments, the game consists of  $T$  periods. In every period  $t \in \{1, \dots, T\}$ , participants receive a monetary endowment of  $w$  and decide which amount of this endowment is contributed to the public good,  $g_{i,t}$ , while the rest,  $w - g_{i,t}$ , goes to the private account. At the end of the game, subject  $i$  receives the benefits from her cumulate contributions to her private account ( $\alpha \sum_{t=1}^T (w - g_{i,t}) - \beta \sum_{t=1}^T (w - g_{i,t})^2$ ) as well as the cumulated benefits from the public good ( $\gamma \sum_{t=1}^T G_t$ ) which depends on the cumulated contributions made by all group members over all periods of the game. Formally, subject  $i$ 's payoff at the end of period  $T$  is given by:

$$\Pi_i = \sum_{t=1}^T \pi_{i,t} = \alpha \sum_{t=1}^T (w - g_{i,t}) - \beta \sum_{t=1}^T (w - g_{i,t})^2 + \gamma \sum_{t=1}^T G_t - \sum_{t=1}^T Tax$$

where  $\pi_{i,t} = \alpha(w - g_{i,t}) - \beta(w - g_{i,t})^2 + \gamma G_t - Tax$  is  $i$ 's payoff in period  $t$ ,  $G_t = \sum_{i=1}^n g_{i,t}$  is the group contribution in period  $t$ , and  $\alpha > \gamma > 0$ ,  $\beta > 0$ ,  $Tax > 0$  are constants. The tax is introduced in order to calibrate payoffs. Since the tax is lump sum, it has no effect on the incentives to contribute. The payoff function is characterized by linearity in the benefits of the public good, and by convexity in the costs of the public good (e.g. Keser 1996; Dannenberg et al. 2014).

Besides our baseline treatment (*base*), we implement two experimental treatments with different dynamic incentive schemes. In *base*, participants decide in each period  $t$  anonymously, and simultaneously about individual contributions to the public good, i.e.,  $0 \leq g_{i,t} \leq w$ . In our weak ratcheting treatment (*weakR*), participants decide exactly as in *base*, but each contribution has to be at least as high as the contribution in the previous period, i.e.,  $g_{i,t-1} \leq g_{i,t} \leq w$ . In the first period, this minimum contribution is equal to zero. In the strong ratcheting treatment (*strongR*), each contribution has to be strictly higher than in the previous period, i.e.,  $g_{i,t-1} < g_{i,t} \leq w$ . As in *weakR*, the minimum contribution is equal to zero in the first period.

## 2.2 Theoretical considerations and hypotheses

We derive theoretical predictions based on the standard model, which assumes that agents are rational and purely selfish. The corresponding contribution paths are illustrated in Figure 1. The unique Nash equilibrium in dominant strategies is given by subjects' payoff maximizing public good contribution levels in period  $t$  of  $g_{i,t}^* = w + \frac{\gamma - \alpha}{2\beta}$ . Consequently, over all periods the payoff maximizing contribution level of subject  $i$  cumulates to  $G_i^* = \sum_{t=1}^T g_{i,t}^* = T g_{i,t}^*$ .<sup>6</sup> The welfare maximizing public good contributions per subject over all periods sum up to  $G_i^\circ = \sum_{t=1}^T g_{i,t}^\circ$  with individual contribution in period  $t$  in the social optimum given by  $g_{i,t}^\circ = w + \frac{n\gamma - \alpha}{2\beta}$ . For  $n > 1$ , there is a social dilemma where the individually rational contribution is below the collectively rational contribution. Each subject's payoff is maximized by the cumulated contribution level equal to  $G_i^* = T(w + \frac{\gamma - \alpha}{2\beta})$  in all three treatments. This conjecture is summarized in our first hypothesis (H1) that states that cumulated contribution levels per subject over all periods are equal to the Nash equilibrium in *base*, *weakR*, and *strongR*.

$$\mathbf{H1.} \quad G_i^{base} = G_i^{weakR} = G_i^{strongR} = G_i^*$$

<sup>6</sup> See Appendix 1 for a derivation of the contribution level in the Nash equilibrium and in the social optimum.

In consequence, the unique Nash equilibrium in cumulated public good contributions per subject is below the welfare maximizing contribution level of  $G_i^\circ = T(w + \frac{n\gamma - \alpha}{2\beta})$  for  $n > 1$ .

Transferred to the individual contribution levels per period, this first hypothesis implies that subjects in *base* and *weakR* divide the optimal cumulative contribution level of  $G_i^*$  equally across all periods. This leads to our second hypothesis (H2) which states that individual contribution levels in *base* and *weakR* are equal to the Nash equilibrium in any period  $t$ .

$$\mathbf{H2.} \quad g_{i,t}^{weakR} = g_{i,t}^{base} = g_{i,t}^*$$

In *strongR*, in contrast, it is by design not possible to contribute  $g_{i,t}^*$ , or any other constant contribution level, in all  $T$  periods of the game. Nevertheless, it is payoff maximizing for subject  $i$  to contribute the cumulated contribution level of  $G_i^*$  in sum over all  $T$  periods. Consequently, the overall equilibrium contribution level is divided differently over the periods of the game. Since individual public good contributions have to increase, subjects contribute less than the Nash solution at the beginning of the game,  $g_{i,t}^{strongR^*} < g_{i,t}^*$  for  $t < T_{[0.5]}$  where  $T_{[0.5]}$  is the median of all  $T$  periods. They contribute the Nash solution in the middle,  $g_{i,t}^{strongR^*} = g_{i,t}^*$  for  $t = T_{[0.5]}$ , and contribute more than the Nash solution at the end,  $g_{i,t}^{strongR^*} > g_{i,t}^*$  for  $t > T_{[0.5]}$ .<sup>7</sup> This conjecture leads to the third hypothesis (H3), which refers to the contribution pattern across all periods. While in period 1 contributions in *strongR* are below those in *base* and *weakR*, in period  $T$  contributions in *strongR* exceed contributions in *base* and *weakR*.

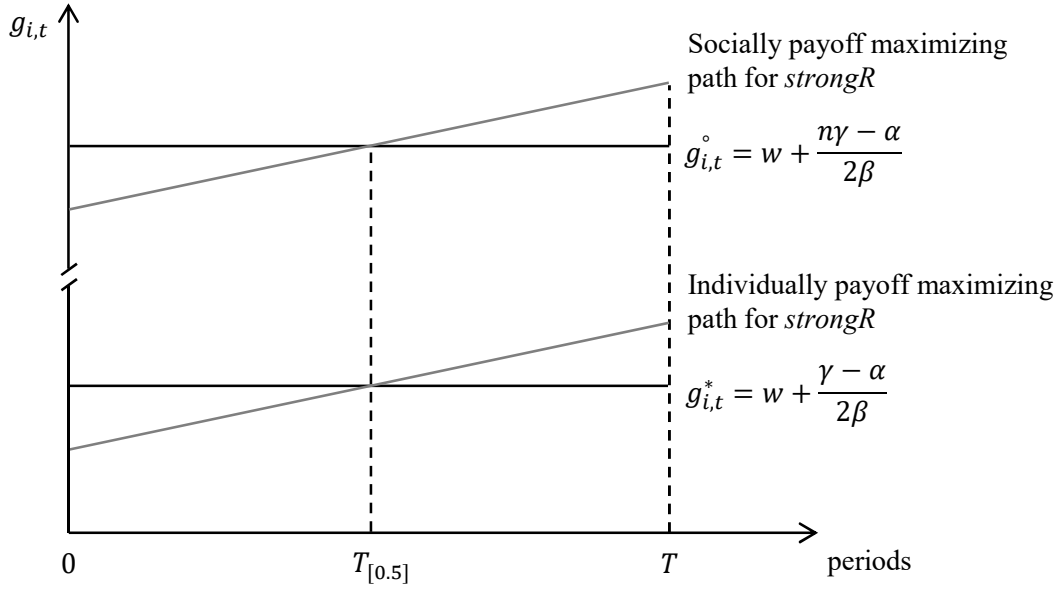
$$\mathbf{H3.} \quad \begin{aligned} \text{(a)} \quad & g_{i,t}^{strongR} < g_{i,t}^* \text{ if } t < T_{[0.5]} \\ \text{(b)} \quad & g_{i,t}^{strongR} = g_{i,t}^* \text{ if } t = T_{[0.5]} \\ \text{(c)} \quad & g_{i,t}^{strongR} > g_{i,t}^* \text{ if } t > T_{[0.5]} \end{aligned}$$

To sum up, strong ratcheting will divide the aggregated contribution level  $G_i^*$  differently from the Nash solution over the periods; therefore, the aggregated contribution level is not reached at lowest cost. Note also that the socially optimal ratcheting path in *strongR*, which can be derived analogously, deviates from the social optimum without ratcheting (see Figure 1).

---

<sup>7</sup> A detailed derivation of our third hypothesis is provided in Appendix 1.

**Figure 1: Theoretical predictions for contribution levels**



### 2.3 Parameterization

The parameters are chosen to meet the following specifications of the Nash equilibrium, the Pareto efficient allocation, and the equilibrium contribution path. Both the Nash equilibrium and social optimum are in the interior of subjects' strategy space and easy to identify and quantify. The differences between subjects' contribution levels in the Nash equilibrium and the social optimum are substantial. Finally, choosing the Nash equilibrium is associated with considerable losses in efficiency. The parameters are set as follows: group size  $n = 4$ , periods  $T = 5$ , endowment per period  $w = 100$  of the experimental currency (ECU),  $\alpha = 4.4$ ,  $\beta = 0.02$ ,  $\gamma = 1$ , and  $Tax = 100$ . The unique Nash equilibrium is where each subject contributes  $G_i^* = 75$  over all five periods and  $g_{i,t}^* = 15$  per period. The symmetric Pareto efficient outcome is where each subject contributes  $G_i^{\circ} = 450$  over all periods and  $g_{i,t}^{\circ} = 90$  per period. In *strongR*, the equilibrium contribution path starts at  $g_{i,1} = 13$  in period 1 and increases to  $g_{i,5} = 17$  in period 5.<sup>8</sup> Consequently, the individual payoffs per subject in the social optimum are given by  $\Pi_i^{\circ} = 1,510$  and are substantially higher than those in the Nash equilibrium  $\Pi_i^* = 947.5$ . The efficiency gain in the social optimum

<sup>8</sup> To simplify the experimental procedure, public good contribution levels are restricted to integer numbers. Furthermore, if participants reach the maximum contribution level of  $g_{i,t} = 100$  in *strongR*, the minimum contribution level in the next period is set to 100.

compared to the Nash solution is about 60 percent. There is practically no difference between the equilibrium payoff for *strongR* and the payoff in the Nash solution (for *base* and *weakR*).<sup>9</sup>

#### 2.4 Procedure

The experiment was conducted at the MaxLab laboratory of the University of Magdeburg, Germany. We used the software z-Tree developed by Fischbacher (2007) for programming the experiment, and participants were recruited via hroot (Bock et al. 2012). In terms of statistical inference, we based our sampling on experimental pilot sessions.<sup>10</sup> Based on this pilot, we decided to conduct 20 sessions in total in May and June 2019 with 340 participants overall. This results in 85 independent observations distributed relatively evenly across treatments, i.e. 30 independent observation in *base*, 27 in *strongR*, and 28 in *weakR*. This allows us to detect treatment effects comparable to those of the experimental pilot at a conventional level of statistical significance (5 percent) at a very high level of statistical power of more than 99 percent.

At the beginning of each session, we randomly created groups of four players which remained fixed (“partner matching”). Participants received a set of experimental instructions which included written descriptions, numerical examples, and control questions. Furthermore, participants were able to make use of a payoff table to verify the numerical examples, answer control questions, and simulate different contribution decisions.<sup>11</sup> During the game, information on group members’ individual contributions to the public good, payoffs, and corresponding average values within the group were transmitted via screen.

Important to note is that a session consists of five consecutive but independent games, called “phases”, each with five periods of the public goods game. Before the first phase started, participants played two non-payoff relevant practice periods. At the end of each session, cumulated earnings of one randomly chosen phase were selected to determine participants’ earnings. Sessions lasted around an hour and earnings ranged from 3.10 euros to 15.50 euros, with an average of 9.50 euros.

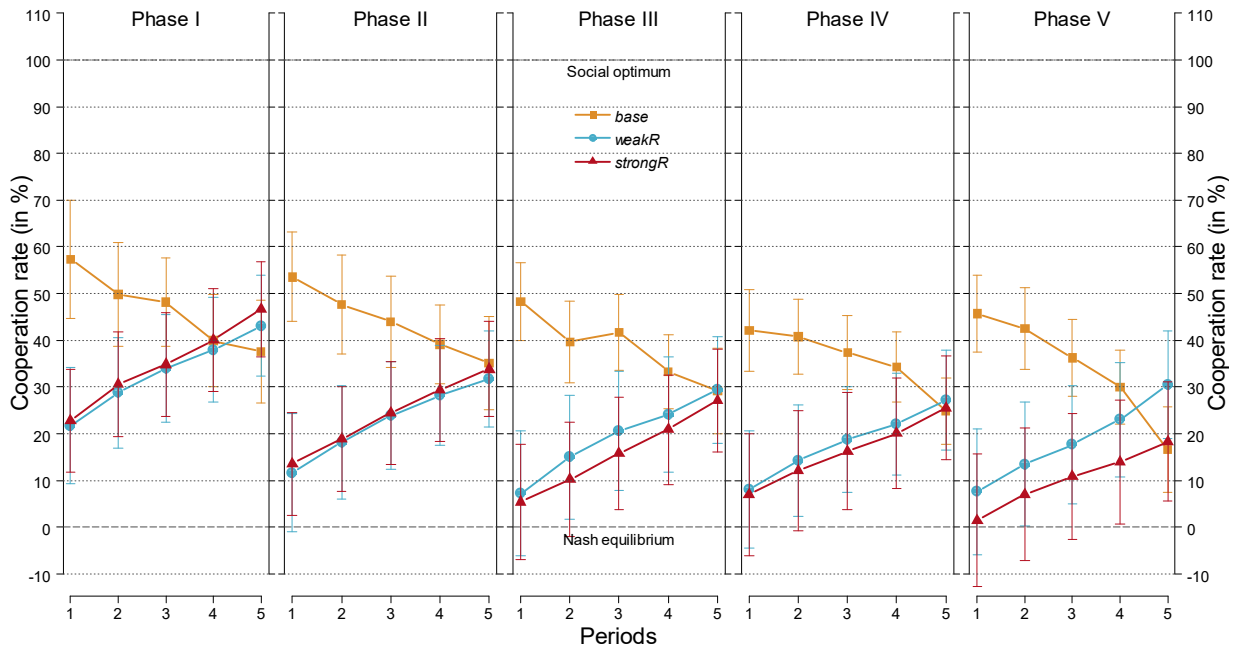
---

<sup>9</sup> Due to the convexity of the costs in the payoff function the cumulated equilibrium payoff in *strongR* (947.0) is slightly below the cumulated payoff in the Nash solution for *base* and *weakR* (947.5).

<sup>10</sup> The first three sessions were run as an experimental pilot in May 2019. All other sessions were executed with exactly the same procedure in June 2019. We used the data generated in the pilot experiment for a power analysis (see Appendix 2).

<sup>11</sup> We provide an example of instructions and screenshots in Appendix 4 and 5. The payoff table was a matrix which indicated the individual payoff in a period given average contributions by the other group members in columns and own contributions in rows (each in 5 units steps from 0 to 100).

**Figure 2: Cooperation rates per group by phase, period, and treatment**



Note: Average cooperation rates by phase, period, and treatment. Cooperation rates are given by  $\left(\frac{g_{i,t} - g_{i,t}^*}{g_{i,t} - g_{i,t}^*}\right) \times 100\%$  in order to ease interpretation and assure that contributions at the Nash equilibrium and the social optimum correspond to 0 and 100 percent, respectively. Confidence intervals are calculated at the 95 percent level.

### 3 Results

#### 3.1 Group level analysis

At first, we analyze participants' public good contribution levels at the group level. Average contribution levels by phase, period, and treatments are presented in Figure 2 and summarized in Table 1. In our *base* treatment, both the level of average contributions as well as its pattern are comparable to other voluntary contribution mechanisms (e.g. Ledyard 1995; Zelmer 2003; Chaudhuri 2011).

On average, the cooperation rate is 39.81 percent.<sup>12</sup> Average cooperation rates over all phases decrease from 49.42 percent in period 1 to 28.66 percent in period 5. Cooperation rates in our ratcheting treatments show a different trend. Our first result shows that cooperation rates in *weakR* and *strongR* are significantly below those in *base* at the beginning of each phase.

<sup>12</sup> In order to ensure comparability to results of linear public goods games, cooperation rates in our non-linear public goods game with interior solutions are calculated such that contributions at the Nash equilibrium are set to 0 percent and contributions at the social optimum are set to 100 percent (see note to Figure 2).

**Table 1: Cooperation rates – Group level analysis**

	Overall	Per phase				
		Phase I	Phase II	Phase III	Phase IV	Phase V
<b>Panel A. baseline (<i>base</i>)</b>						
All Periods	39.81 (2.094)	46.58 (1.627)	43.90 (1.650)	38.44 (1.651)	35.91 (1.685)	34.22 (1.656)
First Period	49.42 (4.302)	57.37 (3.528)	53.57 (3.635)	48.33 (3.812)	42.08 (3.635)	45.73 (3.650)
Last Period	28.66 (4.518)	37.57 (3.705)	35.09 (3.612)	29.13 (3.544)	24.90 (3.749)	16.62 (3.539)
<b>Panel B. weak ratcheting (<i>weakR</i>)</b>						
All Periods	22.34 (1.602)	33.13 (1.565)	22.73 (1.401)	19.26 (1.248)	18.08 (1.338)	18.47 (1.249)
First Period	11.25 (3.039)	21.74 (3.347)	11.17 (2.856)	7.19 (2.443)	8.05 (2.731)	7.59 (2.542)
Last Period	32.27 (3.691)	43.10 (3.502)	31.70 (3.373)	29.38 (3.078)	27.17 (3.183)	30.51 (3.261)
<b>Panel C. strong ratcheting (<i>strongR</i>)</b>						
All Periods	20.30 (1.656)	34.96 (1.557)	24.03 (1.453)	15.93 (1.220)	16.20 (1.241)	10.36 (1.143)
First Period	10.08 (3.365)	22.77 (3.383)	13.58 (3.098)	5.46 (2.427)	7.04 (2.565)	1.54 (2.340)
Last Period	30.29 (3.700)	46.58 (3.575)	33.82 (3.346)	27.18 (3.029)	25.52 (3.026)	18.32 (2.661)

Note: Overall cooperation rates as well as cooperation rates per phase in *base* (Panel A), *weak* ratcheting (Panel B), and *strong* ratcheting (Panel C). Standard errors in parentheses.

**Result 1.** *At the beginning of each phase, cooperation rates in weakR and strongR are significantly lower than those in base. The difference in cooperation rates between weakR and strongR is statistically indistinguishable from zero.*

Support for Result 1 is presented in Table 1, which summarizes the average cooperation rates (see Table 1 - column 1) as well as the cooperation rates separately for each phase (see Table 1 - column 2 - 6). In *strongR*, right at the beginning of each of the five phases, cooperation rates are on average significantly below those in *base* (10.08 vs. 49.42 percent, p-value < 0.000, Wilcoxon rank sum test). The same holds true for the cooperation rates in *weakR*. Contribution rates in period 1 are on average significantly lower in *weakR* compared to *base* (11.25 vs. 49.42 percent, p-value < 0.000, Wilcoxon rank sum test). There are no significant differences between our two ratcheting treatments right at the beginning of the phases (10.08 vs. 11.25 percent, p-value = 0.3588, Wilcoxon rank sum test). Result 1 shows that contribution levels differ significantly across treatments right at the beginning of each phase. Our next result reveals that contribution levels also show clearly different trends across treatments.

**Table 2: Ratchet effects – Group level analysis**

	Dependent variable: Cooperation rate		
	(1)	(2)	(3)
<i>weakR</i>	-17.474*** (5.352)	-17.474*** (2.567)	-38.169*** (5.496)
<i>strongR</i>	-19.513*** (5.302)	-19.513*** (2.543)	-39.337*** (5.444)
<i>period 2</i>		2.140 (3.320)	-5.316 (5.349)
<i>period 3</i>		4.366 (3.320)	-7.896 (5.349)
<i>period 4</i>		4.940 (3.320)	-14.067*** (5.349)
<i>period 5</i>		6.042* (3.320)	-20.753*** (5.349)
<i>weakR x period 2</i>			12.056 (7.772)
<i>strongR x period 2</i>			11.006 (7.699)
<i>weakR x period 3</i>			19.641** (7.772)
<i>strongR x period 3</i>			18.284** (7.699)
<i>weakR x period 4</i>			29.899*** (7.772)
<i>strongR x period 4</i>			28.869*** (7.699)
<i>weakR x period 5</i>			41.879*** (7.772)
<i>strongR x period 5</i>			40.960*** (7.699)
Constant	39.809*** (3.684)	36.312*** (2.744)	49.416*** (3.782)
Observations	85	425	425
R <sup>2</sup>	0.167	0.154	0.239
Adjusted R <sup>2</sup>	0.147	0.141	0.213

Note: OLS regressions. Standard errors in parentheses. The dependent variable is the cooperation rate at the group level. The explanatory variables in column 1 are indicator variables for the treatments. Column 2 contains indicator variables for treatments and periods. Column 3 also contains the interaction of indicator variables for treatments with indicators for periods. *weakR*: indicator for weak ratcheting treatment, *strongR*: indicator for strong ratcheting treatment, *period 1 - 5*: indicator for period 1 to 5. \*  $p < 0.10$ , \*\*  $p < 0.05$ , and \*\*\*  $p < 0.01$ .

**Result 2.** *In base, cooperation rates show a significant downward trend over time. Cooperation rates in weakR and strongR, in contrast, increase significantly over time.*



Figure 2 and Table 1 show the trends in cooperation rates across treatments. In *base*, average cooperation rates decrease by around 21 percentage points from 49.42 percent in period 1 to 28.66 percent in period 5. Cooperation rates in the two ratcheting treatments show a different trend. In *weakR*, cooperation rates increase steadily by 21 percentage points from 11.25 percent in period 1 to 32.27 percent in period 5. Furthermore, cooperation rates in *strongR* increase by 20 percentage points from 10.08 percent in period 1 to 30.29 percent in period 5.

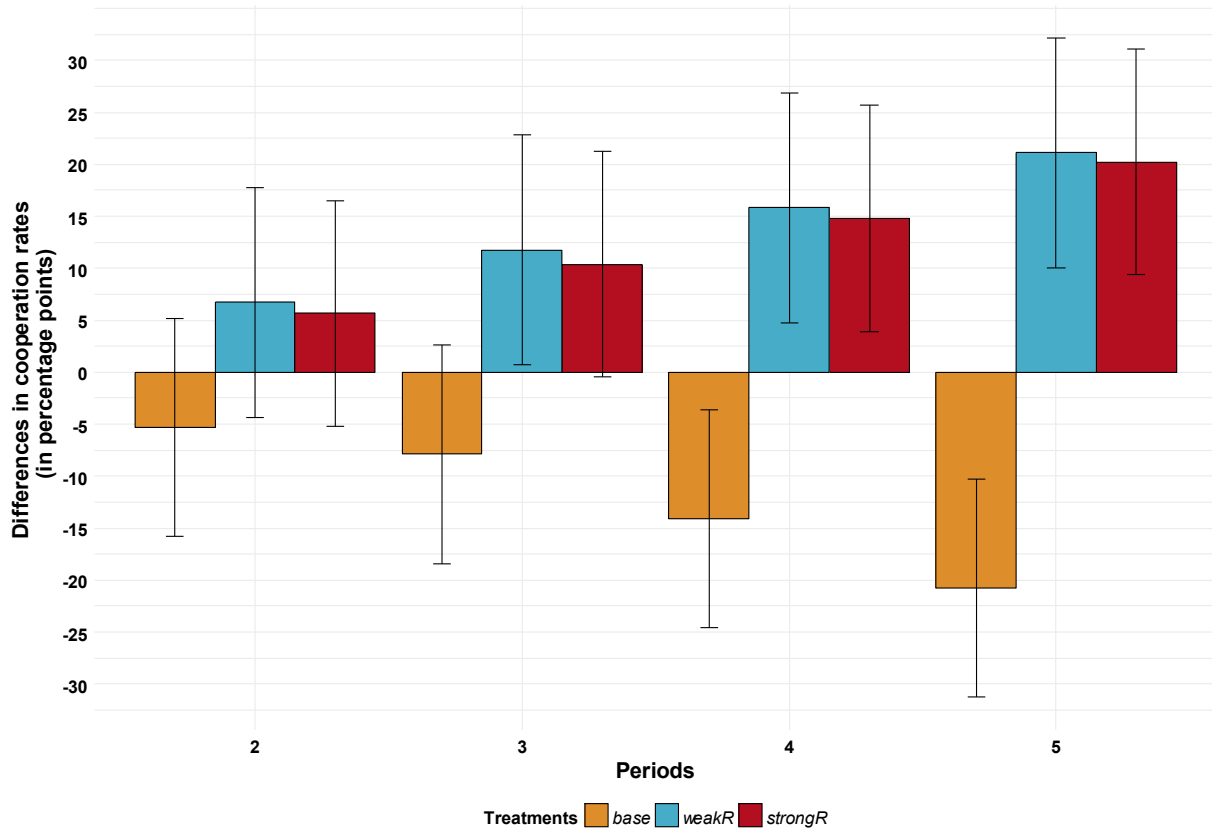
Result 2 is supported by a series of linear regression models. Table 2 reports the results of OLS models of the groups' cooperation rates. The first model in Table 2 includes indicator variables for the two ratcheting treatments but does not allow for different effects by periods. The second model controls for general trends and the third model accounts for different trends across treatments by including interactions between periods and treatments. To ease interpretation, Figure 3 shows the estimated interaction effects, i.e. the deviation from first period contribution levels by treatment of all phases. While cooperation rates in *base* significantly decrease over time, those in *weakR* and *strongR* show a significantly positive trend. While the increasing trend in *strongR* is externally enforced, the increase in *weakR* is clearly above the theoretical prediction. Consequently, we have to reject our hypothesis H2, which stated that participants' cooperation rates in *base* and *weakR* are identical to the Nash equilibrium over the course of the experiment. Furthermore, the average level of contributions in *strongR* is clearly above the theoretical prediction. Thus, we also have to reject our hypothesis H3, which stated that contribution patterns in *strongR* are centered around the Nash equilibrium.

All in all, trends in cooperation rates in both ratcheting treatments run in the opposite direction than our baseline treatment. They are, however, not strong enough to outweigh the initial differences. This leads to our third result.

**Result 3.** *On average, cooperation rates in weakR and strongR are significantly below the cooperation rates in base. The difference in cooperation rates between weakR and strongR is statistically indistinguishable from zero.*

The average cooperation rates in *weakR* (22.34 vs. 39.81 percent, p-value = 0.0032, Wilcoxon rank sum test) and *strongR* (20.30 vs. 39.81 percent, p-value = 0.0012, Wilcoxon rank sum test) treatments are significantly lower than in *base* (see also Table 2, Model 1). Furthermore, we do not

**Figure 3: Trends in cooperation rates per group by period and treatment**



Note: Trends by treatment are given by the difference in cooperation rates (in percentage points) in period 2 – 5 compared to cooperation rates in period 1. Trends are calculated separately by treatment and periods. Confidence intervals are calculated at the 95 percent level.

find significant differences between the cooperation rates in *weakR* and *strongR* (22.34 vs. 20.30 percent, p-value = 0.5642, Wilcoxon rank sum test). Cumulated public good contribution levels in *weakR* and *strongR* are, therefore, significantly below those in our baseline treatment. In *base*, participants’ public good contributions cumulate to on average 897 per phase. In *weakR* and *strongR*, in contrast, contributions per phase cumulate to 635 and 604, respectively. Cumulated provision levels in *weakR* (p-value < 0.000) and *strongR* (p-value < 0.000) are significantly below those in *base* and, consequently, we have to reject our hypotheses H1, which stated that cumulated public good contributions are equal across all treatments.

Given the different trends in contribution levels, our results at the group level show that the comparable low cooperation rates at the beginning of each phase are decisive for the low cumulated public good contribution levels in both ratcheting treatments. Furthermore, cooperation rates under ratcheting decrease from phase to phase (see Figure 2).

### 3.2 Individual level analysis

Our analysis at the group level in Section 3.1 shows that cooperation rates in both ratcheting treatments (*weakR* and *strongR*) are significantly lower at the beginning of the game than in the baseline treatment. Although cooperation rates in *weakR* and *strongR* increase steadily over time, this positive trend is not strong enough to catch up the backlog to *base*. On average, cooperation rates in *weakR* and *strongR* are significantly below those in *base*. The group level analysis, however, leaves open what drives participants in *weakR* and *strongR* to cooperate so little at the beginning of the game. In this section, we investigate participants' cooperation behavior at the individual level in order to analyze why participants in both ratcheting treatments are especially cautious with their public good contributions at the beginning of the game.

Participants' willingness to take social risks plays a central role in their willingness to cooperate. There is always an element of *fear of exploitation* (De Cremer 1999) that causes participants to be cautious, leading to non-cooperative behavior (e.g. Yamagishi and Sato 1986, Parks and Hulbert 1995). In the realm of the private provision of public goods, the fear of exploitation can have both monetary as well as non-monetary components. The monetary component comprises participants' concern that they will receive no payoff from their engagement. A motive for non-cooperative behavior could be the attempt to minimize the pure monetary risk of being exploited by not contributing to the public good, since a player's own payoffs are lowest if she cooperates while others do not (e.g. Dawes et al. 1986). Beyond this purely outcome-based motive, the fear of exploitation could also incur a psychological element as well as non-monetary costs. This includes disutility from the experience of betrayal defined as non-reciprocated trust (e.g. Aimone et al. 2015, Bohnet et al. 2008).

The fear of being exploited by free-riders could explain non-cooperative behavior in virtually all public goods games (e.g. Cubitt et al. 2017). However, ratcheting increases the social risk of being exploited at the beginning of the game, which could help to explain why participants in *weakR* and *strongR* are especially cautious and contribute initially less than in our baseline treatment. With ratcheting, positive contributions determine a path dependency, meaning participants cannot simply adjust their contributions downward in the next period. They are bound from below to what they have contributed in *weakR* and *strongR*. They cannot directly mirror uncooperative behavior of other group members in the corresponding phase of the experiment; the risk of being exploited is especially high with ratcheting. Consequently, participants' fear of the risk of being exploited

**Table 3: Exploitation effect – Individual level analysis**

	Dependent variable: Cumulated cooperation rates per phase					
	(1)	(2)	(3)	(4)	(5)	(6)
<i>gap</i>	-0.268*** (0.028)		-0.239*** (0.029)		-0.119*** (0.046)	
<i>exploitation</i>		-6.365*** (1.400)		-5.405*** (1.391)		-0.673 (2.109)
<i>weakR</i>			-4.637*** (1.510)	-6.471*** (1.518)	-8.158*** (1.872)	-3.188 (2.071)
<i>strongR</i>			-7.238*** (1.501)	-9.086*** (1.508)	-10.314*** (1.856)	-5.006** (2.065)
<i>gap x weakR</i>					-0.176*** (0.056)	
<i>exploitation x weakR</i>						-6.642** (2.944)
<i>gap x strongR</i>					-0.154*** (0.055)	
<i>exploitation x strongR</i>						-8.339*** (2.916)
<i>sum con Phase prev</i>	0.812*** (0.023)	0.721*** (0.022)	0.773*** (0.025)	0.683*** (0.022)	0.773*** (0.025)	0.686*** (0.022)
Constant	-4.100*** (1.279)	6.941*** (0.895)	1.481 (1.765)	12.660*** (1.308)	3.881** (1.899)	10.163*** (1.549)
Observations	1,360	1,360	1,360	1,360	1,360	1,360
R <sup>2</sup>	0.515	0.490	0.523	0.504	0.527	0.507
Adjusted R <sup>2</sup>	0.514	0.489	0.521	0.503	0.525	0.505

Note: OLS regressions. Standard errors in parentheses. The dependent variable is cumulated cooperation rate per phase at the individual level. *gap* captures the difference between participants' cumulated cooperation rates and the average cumulated cooperation rates by their group members in the previous phase. *exploitation* indicates that participants contributed more (1) or less (0) than their group members in the previous phase. *sum con Phase prev* captures participants' cumulated cooperation rates in the previous phase. Furthermore, we include indicator variables for the different ratcheting treatments as well as their interactions with *gap* and *exploitation*. \*  $p < 0.10$ , \*\*  $p < 0.05$ , and \*\*\*  $p < 0.01$ .

could help to explain why participants are especially reluctant at the beginning of the each phase in *weakR* and *strongR* as compared to *base*.

In order to investigate whether participants' fear of the risk of being exploited drives our results, we analyze participants' public good contributions at the individual level. The results in Section 3.1 suggest that a non-negligible amount of participants anticipate right at the beginning of the game that cooperative behavior can easily be exploited, and that this has especially severe consequences under ratcheting. They are especially cautious and restrict their cooperation levels already at the beginning of the first phase. In addition, our experimental design allows us to test

**Table 4: Exploitation effect by phase and treatment**

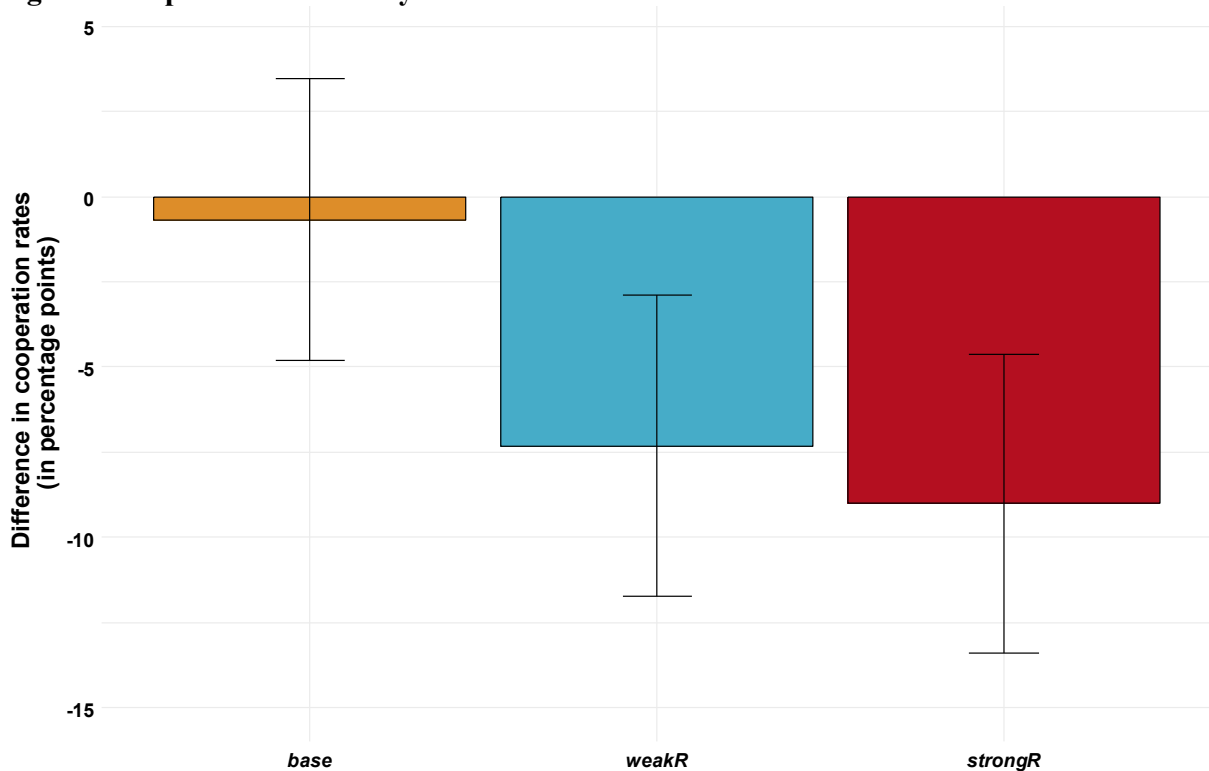
	Overall	Per phase			
		Phase II	Phase III	Phase IV	Phase V
<b>Panel A. baseline (<i>base</i>)</b>					
	-0.67 (0.021)	-2.85 (4.634)	-0.98 (4.165)	-0.20 (3.980)	-0.46 (4.052)
<b>Panel B. weak ratcheting (<i>weakR</i>)</b>					
	-7.32*** (2.256)	-15.90*** (5.173)	-9.28** (4.343)	-6.65 (4.216)	0.38 (4.270)
<b>Panel C. strong ratcheting (<i>strongR</i>)</b>					
	-9.01*** (2.230)	-14.00*** (5.170)	-12.03*** (4.343)	-4.80 (4.418)	-6.50 (4.207)

Note: Overall exploitation effects as well as exploitation effects per phase in *base* (Panel A), *weak* ratcheting (Panel B), and *strong* ratcheting (Panel C). Standard errors in parentheses. Estimates for the overall exploitation effect are based on the regression in Table 3 – model (6). Corresponding estimates for the exploitation effect per phase are based on the regressions in Table A3.1 (model (1) – (4)) in the appendix 3. \*  $p < 0.10$ , \*\*  $p < 0.05$ , and \*\*\*  $p < 0.01$ .

how participants who have been exploited in one phase of the experiment adjust their public good contribution levels in the next phase of the game. Based on a series of linear regression models (see Table 3), we summarize the main findings in Table 4 and illustrate them in Figure 4. In the analysis we use participants' cumulated cooperation rates in one phase of the experiment as a dependent variable. Participants' cumulated cooperation rates in the previous phase (*sum con Phase prev*) and the difference between participants' cumulated cooperation rates and the average of the cumulated cooperation rates by other group members in previous phase (*gap*) are our explanatory variables of major interest. Furthermore, the indicator variable *exploitation* indicates that a participant has been exploited, and cooperated more than the other group members in the previous phase.

We find that participants who have been exploited contribute less in the subsequent phase of the experiment. Across all treatments, participants who contribute more to the public good than their group members in one phase of the experiment contribute significantly less to the public good than those who have not have been exploited ( $p$ -value  $< 0.001$ , Table 2 – Column 1). On average, their cooperation rates decrease by 6.37 percentage points. Most importantly, this effect is especially strong with ratcheting. Based on the model in Table 3 – Column 6, we separately estimate the exploitation effect by treatments. Results are summarized in Table 4 and illustrated in Figure 4. Separating this effect by treatments shows that the effect of exploitation is especially strong in the weak and strong ratcheting treatment. In *base*, the effect is slightly negative but does not reach significance at conventional statistical levels (-0.67 percentage points,  $p$ -value = 0.7498, Table 4 – Column 1). In *weakR*, exploited participants contribute in the subsequent phase significantly less than non-exploited participants. With 7.32 percentage points, this effect is quite substantial and

**Figure 4: Exploitation effect by treatment**



Note: Exploitation effect given by the difference (in percentage points) in cumulated cooperation rates per phase between participants who cooperated more or less than their group members in the previous phase. The exploitation effect is estimated across all phases, and separately for the different treatments. Confidence intervals are calculated at the 95 percent level.

statistically significant ( $p$ -value = 0.002, Table 4 – Column 1). In *strongR*, the exploitation effect is even bigger and amounts to 9.01 percentage points ( $p$ -value < 0.000, Table 4 – Column 1).

By examining the exploitation effect on a phase-to-phase basis (see Table 4 – columns 2-5), we find that this effect loses its strength over the course of the experiment. While it is especially strong at the beginning of the experiment, the exploitation effect does not reach statistical significance at the end of the experiment. Participants who have been exploited in phase I reduce their contribution in phase II by 15.90 percentage points ( $p$ -value = 0.0023) in *weakR* and 14.00 percentage points ( $p$ -value = 0.0071) in *strongR*. The effect of being exploited in phase II also significantly affects participants' cooperation rates in phase III (*weakR*: -9.28 percentage points,  $p$ -value = 0.0371; *strongR*: -12.03 percentage points,  $p$ -value = 0.0059). From phase III onwards, the exploitation effect does not reach statistical significance at conventional levels, neither in the weak nor in the strong ratcheting treatment.

Our individual level analysis suggests that ratcheting enhances participants' fear of being exploited at the beginning of the game, thereby explaining the low cooperation rates. One could argue that

this effect is reversed once the first period has been completed. From period 2 onwards, cooperation rates are bounded from below and, consequently, the risk of being exploited is reduced. More precisely, the risk of being exploited does not increase over the course of the game if participants do not increase their own cooperation rates. This determines an upper bound for the level of exploitation, reduces the risk of being exploited in the subsequent period, and helps to explain the positive trend in cooperation rates in our weak ratcheting treatment.

## **4 Conclusion**

In this paper, we investigate the effect of dynamic minimum contribution levels in a repeated social dilemma. We distinguish between strong ratcheting, where individual contributions to a cumulative public good have to increase over time, and weak ratcheting, which only prevents individual contributions from decreasing over time. Our core result is a statistically and economically significant ratchet effect. In both ratcheting treatments, participants strategically restrict their public good contribution levels at the beginning of the game. While cooperation rates under ratcheting increase over the course of the game, this increase is not sufficient to compensate the distinct drop of cooperation rates at the beginning of the game.

Our work suggests that ratcheting loses its efficiency when contribution decisions cannot be constantly revised. Extending the experiments by Dorsey (1992) and Kurzban et al. (2001), who find that ratcheting fosters cooperation in public goods games in continuous time, we show that cumulated cooperation rates under ratcheting are almost halve as compared to the baseline in our discrete game where contributions cannot be revised any number of time. This finding is not merely of technical or application oriented interest because many institutional mechanisms to foster cooperation in social dilemmas rely on discrete political decisions, such as international environmental agreements, that involve dynamic incentive schemes and cannot be revised any number of time.

The observed treatment effect is not consistent with standard theory which has – as it generally does in social dilemma games – only very limited explanatory power. We identify participants' fear of the risk of being exploited as a potential explanation of our results. Ratcheting deprives participants of the opportunity to directly reciprocate non-cooperate behavior. Cooperative behavior can easily be exploited by free-riders without the opportunity to directly reduce individual contributions. In this line, subjects either anticipate that cooperative behavior at the beginning of

the game entails the elevated risk of being exploited by free-riders, or subjects who experienced such exploitation reduce their contributions in the subsequent phases of the experiment.

Our results might have some interesting policy implications, particularly for recent climate policy under the Paris Agreement. The observed negative effect of ratcheting on efficiency clearly challenges the belief that ratcheting is able to increase the contributions to the global public good climate change mitigation.<sup>13</sup> We have shown, at least, that there is neither theoretical nor empirical support for the effectiveness of ratcheting.

While the internal validity of our experiment is high (due to the fact that we are able to identify a statistically robust cause-and-effect relationship), the external validity, in particular for global climate policy, is more difficult to gauge. Our design is highly stylized and captures only the main features of the game which is played by the countries' representatives. There are several potentially relevant features where our design simplifies the game played outside the laboratory. For example, it is well documented that subjects behave differently when rules to increase public goods contributions are endogenously implemented instead of being exogenously imposed. Furthermore, the negotiation process for the real world climate policy is much richer than our experiment's, where subjects simply announce simultaneously and anonymously their contributions, and this process might interact with ratcheting. Also, our assumption of identical players might be critical since heterogeneity is an important aspect of real world climate policy, and laboratory evidence (Kesternich et al. 2014, 2018) shows that the performance of MCLs depends on the type of heterogeneity. Furthermore, our design assumes perfect compliance to the rules of the game. In the field, i.e. in global climate policy, compliance is, however, not enforceable. Countries do have the choice to leave an agreement if the conditions seem to be too unfavorable for them. The U.S. retreat from the Paris Agreement is a prominent example.<sup>14</sup> Our experimental results show that even under perfect compliance there is no empirical evidence that ratcheting is able to mitigate or even solve the social dilemma of reducing global warming. On the contrary, ratcheting leads to lower contributions to the cumulative public good. Our results thus complement the existing economic criticism of the Paris Agreement, which states in particular that the agreement is not addressing the free-rider problem (e.g. Cooper et al. 2017). Simply to formulate a collective goal and let parties

---

13 According to the UNFCCC, the "pledge-and-review" implemented in the Paris Agreement process will "enable an upward spiral of ambition over time" (Ad Hoc Working Group on the Durban Platform, 2014, p. 18).

14 On June 1, 2017, U.S. President Donald Trump announced that the U.S. would cease all participation in the Paris Agreement. In accordance with Article 28 of the Paris Agreement, the earliest possible effective withdrawal date by the U.S. cannot be before November 4, 2020, four years after the Agreement came into effect in the U.S.



decide about individual contributions, which are hopefully increasing over time, is very likely not sufficient. Assuming general compliance with the agreement, the effect could not only cause countries to restrict their national contributions at the beginning, it could also cause countries to delay the process of determining new and more ambitious national determined contributions (in order to avoid the risk of being exploited). Furthermore, since the announced individual commitments are in aggregate well below the necessary overall CO<sub>2</sub> emissions reductions level for achieving the two-degree-goal (Rogelj et al. 2016), the hope that more ambitious emissions reduction targets can be realized by ratcheting, seems to be “wishful thinking” (Cooper et al. 2017, p. 3), given our results. What is needed is a reciprocal commitment of countries where countries price carbon in case other countries do the same (MacKay et al. 2015).

The need for a reciprocal commitment leads us to the question of how the weak performance of ratcheting might be improved. A plausible supplement could be the implementation of a “reciprocal commitment device” into a game with ratcheting. We leave this issue for future research.

## 5 References

- Aimone, J. Ball, S., and King-Casas, B. (2015): The betrayal aversion elicitation task: An individual level betrayal aversion measure, *PLoS One*, 10(9), e0137491.
- Andreoni, J. (1993): An experimental test of the public-goods crowding-out hypothesis, *American Economic Review* 83, 1317-1327.
- Amano, T. and Ohashi, H. (2018): Ratcheting, Competition, and the Diffusion of Technological Change: The Case of Televisions under an Energy Efficiency Program, Harvard Business School Working Paper No. 19-021.
- Berliner, J.S. (1957): *Factory and manager in the USSR*. Cambridge, Harvard University Press.
- Bock, O., Nicklisch, A., and Baetge, I. (2012): hroot: Hamburg registration and organization online tool, WiSo-HH Working Paper Series No. 1.
- Bohnet, I., Greig, G., Herrmann, B., and Zeckhauser, R. (2008): Betrayal aversion: Evidence from Brazil, China, Oman, Switzerland, and the United States. *American Economic Review* 98, 294-310.
- Bohnet, I. and Zeckhauser, R. (2004): Trust, risk and betrayal. *Journal of Economic Behavior & Organization* 55, 467-484.
- Cardella, E. and Depew, B. (2018): Output restriction and the ratchet effect: Evidence from a real-effort work task, *Games and Economic Behavior* 107, 182-202.
- Charness, G., Kuhn, P., and Villeval, M.C. (2011): Competition and the Ratchet Effect, *Journal of Labor Economics* 29(3), 513-547.
- Chaudhuri, A. (1998): The ratchet principle in a principal agent game with unknown costs: an experimental analysis, *Journal of Economic Behavior & Organization* 37, 291-304.
- Chaudhuri, A. (2011): Sustaining cooperation in laboratory public good games experiments: a selective survey of the literature. *Experimental Economics* 14, 47-83.
- Cooper, R.N., Cramton, P., Edenhofer, O., Gollier, C., Laurent, E., MacKay, D.J.C., Nordhaus, W., Ockenfels, A., Stiglitz, J., Stoft, S., Tirole, J., and Weitzman, M. (2017): Why Paris Did not Solve the Climate Dilemma, in: *Global Carbon Pricing. The Path to Climate Cooperation*, edited by Cramton, P., MacKay, D.J.C., Ockenfels, A., and Stoft, S., 2017, MIT Press.
- Cubitt, R., Gächter, S., and Querica, S. (2017): Conditional cooperation and betrayal aversion, *Journal of Economic Behavior & Organization* 141, 110-121.
- Dannenberg, A., Lange A., and Sturm, B. (2014): Participation and Commitment in Voluntary Coalitions to Provide Public Goods, *Economica* 81(322), 257-275.
- Dawes, R., Orbell, J., Simmons, R., and Van De Kragt, A. (1986): Organizing groups for collective action. *American Political Science Review* 80, 1171-1185.
- De Cremer, D. (1999): Trust and fear of exploitation in public goods dilemma. *Current Psychology* 18, 153-163.
- Diederich, J., and Goeschl, T. (2014): Willingness to pay for voluntary climate action and its determinants: field experimental evidence, *Environmental and Resource Economics* 57(3), 405-429.
- Diev, P. and Hichri, W. (2008): Dynamic voluntary contributions to a discrete public good: Experimental evidence. *Economics Bulletin*, 3(23), 1-11.
- Dorsey, R. (1992): The voluntary contribution mechanism with real-time revisions. *Public Choice* 73, 261-282.
- Duffy, J., Ochs, J., and Vesterlund, L. (2007): Giving little by little: Dynamic voluntary contribution games, *Journal of Public Economics* 91(9), 1708-1730.
- Eckel, C., Grossman, P., and Johnston, R. (2005): An experimental test of the crowding out hypothesis, *Journal of Public Economics* 89, 1543-1560.
- Fellner, G., Margreiter, M., and Eraso, N.O. (2003): When the past is present – The ratchet effect in the local commons, discussion paper.

- Fischbacher, U. (2007): Z-Tree: Zurich Toolbox for Ready-made Economic Experiments, *Experimental Economics* 10(2), 171-178.
- Friedman, D. and Oprea, R. (2012): A continuous dilemma, *American Economic Review* 102, 337-363.
- Green, J.F., Sterner, T., and Wagner, G. (2014): A balance of bottom-up and top-down in linking climate policies, *Nature Climate Change* 4 (12), 1064-1067.
- Gronberg, T.J., Luccasen, R.A., Turocy, T.L., and Van Huyck, J.B. (2012): Are tax-financed contributions to a public good completely crowded-out? Experimental evidence, *Journal of Public Economics* 96(7-8), 596-603.
- Güth, W., Levati, M., and Stiehler, A. (2002): Privately contributing to public goods over time: An experimental study, SFB 373 Discussion Paper No. 18, Humboldt University of Berlin.
- Holz, C., Siegel, L.S., Johnston, E., Jones, A.P., and Sterman, J. (2018): Ratcheting ambition to limit warming to 1.5 °C—trade-offs between emission reductions and carbon dioxide removal, *Environmental Research Letters* 13, <https://doi.org/10.1088/1748-9326/aac0c1>.
- Kesternich, M., Lange, A., and Sturm, B. (2018): On the Performance of Rule-Based Contribution Schemes Under Endowment Heterogeneity, *Experimental Economics* 21, 180-214.
- Kesternich, M., Lange, A., and Sturm, B. (2014): The Impact of Burden Sharing Rules on the Voluntary Provision of Public Goods, *Journal of Economic Behavior & Organization* 105, 107-123.
- Keser, C. (1996): Voluntary contributions to a public good when partial contribution is a dominant strategy, *Economics Letters* 50(3), 359-366.
- Kurzban, R., McCabe K., Smith, V., and Wilson B. (2001): Incremental commitment and reciprocity in a real-time public goods game, *Personality and Social Psychology Bulletin* 27, 1662-1673.
- Ledyard, J.O. (1995): Public Goods: A Survey of Experimental Research, in J.H. Kagel and A.E. Roth, eds., *Handbook of experimental economics*. Princeton: Princeton University Press, 1995, 111-194.
- Löschel, A., Sturm, B. and Vogt, C. (2013): The demand for climate change mitigation – an empirical assessment for Germany, *Economics Letters* 118, 415-418.
- Macartney, H. (2016): The Dynamic Effects of Educational Accountability, *Journal of Labor Economics* 34(1), 1-28.
- MacKay, D.J.C., Cramton, P., Ockenfels, A., and Stoft, S. (2015): Price Carbon – I will If You Will, *Nature* 526, 315-316.
- Marx, L. and Matthews, S. (2000): Dynamic voluntary contribution to a public project, *Review of Economic Studies* 67, 327-358.
- Oprea, R., Charness, G., and Friedman, D. (2014): Continuous time and communication in a public-goods experiment, *Journal of Economic Behavior & Organization* 108, 212-223.
- Parks, C. and Hulbert, L. (1995): High and low trusters' responses to fear in a payoff matrix, *Journal of Conflict Resolution* 1995, 718-730.
- Rogelj, J., den Elzen, M., Höhne, N., Fransen, T., Fekete, H., Winkler, H., Schaeffer, R., Fu Sha, F., Riahi, K., and Meinshausen, M. (2016): Paris Agreement climate proposals need a boost to keep warming well below 2 °C, *Nature* 534, pages 631-639.
- Schelling, T. (1960): *The Strategy of Conflict*. Harvard University Press, Cambridge, MA.
- UNFCCC (2015): Paris Agreement, United Nations, [https://unfccc.int/files/meetings/paris\\_nov\\_2015/application/pdf/paris\\_agreement\\_english.pdf](https://unfccc.int/files/meetings/paris_nov_2015/application/pdf/paris_agreement_english.pdf).
- Weitzman, M.L. (1980): The 'Ratchet Principle' and Performance Incentives, *Bell Journal of Economics* 11(1), 302-308.
- Ye, M., Zheng, J., Nikolov, P., and Asher, S. (2020): One step at a time: Does gradualism build coordination?, *Management Science* 66, 113-129.
- Young, O.R. (2016): The Paris Agreement: Destined to Succeed or Doomed to Fail? *Politics and Governance* 4 (3), 124-132.

Zelmer, J. (2003): Linear Public Goods Experiments: A Meta-Analysis, *Experimental Economics* 6(3), 299-310.

## Appendix 1: Theoretical predictions

### *Treatments base and weakR*

The payoff function for subject  $i$ ,  $i = 1, \dots, n$ , over all  $T = 5$  periods is equal to

$$\Pi_i = \sum_{t=1}^{T=5} \pi_{i,t} = \alpha \sum_{t=1}^{T=5} (w - g_{i,t}) - \beta \sum_{t=1}^{T=5} (w - g_{i,t})^2 + \gamma \sum_{t=1}^{T=5} G_t - \sum_{t=1}^{T=5} Tax$$

where  $G_t = \sum_{i=1}^n g_{i,t}$  is the group contribution to the public good in period  $t$ . In the last period 5, subject  $i$  maximizes the following payoff function

$$\pi_{i,5} = \alpha(w - g_{i,5}) - \beta(w - g_{i,5})^2 + \gamma G_5 - Tax$$

We determine the FOC for payoff maximization

$$\frac{\delta \pi_{i,5}}{\delta g_{i,5}} = -\alpha - 2\beta(w - g_{i,5})(-1) + \gamma = 0 \Leftrightarrow g_{i,5} = w + \frac{\gamma - \alpha}{2\beta}$$

Since the decision situation in period 4 and all other periods 3 to 1 is exactly the same, we obtain as the individual Nash equilibrium contribution level

$$g_{i,t}^* = w + \frac{\gamma - \alpha}{2\beta}$$

In order to determine the individual contribution level in the social optimum we consider the group payoff function

$$\Pi = \sum_{t=1}^{T=5} \Pi_t = \alpha \sum_{t=1}^{T=5} (W - G_t) - \beta \frac{1}{n} \sum_{t=1}^{T=5} (W - G_t)^2 + \gamma n \sum_{t=1}^{T=5} G_t - n \sum_{t=1}^{T=5} Tax$$

where  $W$  is the group's endowment and  $G_t$  is the group contribution to the public good. In the last period 5, for a payoff maximum we obtain the following FOC and individual contribution level in the social optimum.

$$\frac{\delta \Pi_5}{\delta G_5} = -\alpha - 2\beta \frac{1}{n} (W - G_5)(-1) + \gamma n = 0 \Leftrightarrow g_{i,5}^\circ = w + \frac{n\gamma - \alpha}{2\beta}$$

For this, we are using  $W = nw$  and  $G_5 = ng_{i,5}$  since subjects are identical.

### *Treatment strongR*

We assume that subjects increase their contributions only by the minimum integer amount required (+1). This means that  $g_{i,1} = g_{i,2} - 1 = g_{i,3} - 2 = g_{i,4} - 3 = g_{i,5} - 4$ . For the group ( $n = 4$ ) contribution we obtain  $G_1 = G_2 - 4 = G_3 - 8 = G_4 - 12 = G_5 - 16$ .

We express the payoff function (see above) in contributions of period 1 only and determine the FOC for payoff maximization

$$\frac{\delta \pi_i}{\delta g_{i,1}} = -5\alpha - 2\beta(w - g_{i,1})(-1) - 2\beta(w - g_{i,1} - 1)(-1) - 2\beta(w - g_{i,1} - 2)(-1) - 2\beta(w - g_{i,1} - 3)(-1) - 2\beta(w - g_{i,1} - 4)(-1) + 5\gamma = 0$$

For the individual equilibrium contribution level in period 1 we obtain

$$-5\alpha + 10\beta(w - g_{i,1}) - 20\beta + 5\gamma = 0 \Leftrightarrow g_{i,1} = \frac{-\alpha + 2\beta w - 4\beta + \gamma}{2\beta}$$

$$\Leftrightarrow g_{i,1}^{strongR^*} = w + \frac{\gamma - \alpha}{2\beta} - 2$$

For the contribution level in periods 2-5 we obtain

$$g_{i,2}^{strongR^*} = w + \frac{\gamma - \alpha}{2\beta} - 1, g_{i,3}^{strongR^*} = w + \frac{\gamma - \alpha}{2\beta},$$

$$g_{i,4}^{strongR^*} = w + \frac{\gamma - \alpha}{2\beta} + 1, g_{i,5}^{strongR^*} = w + \frac{\gamma - \alpha}{2\beta} + 2$$

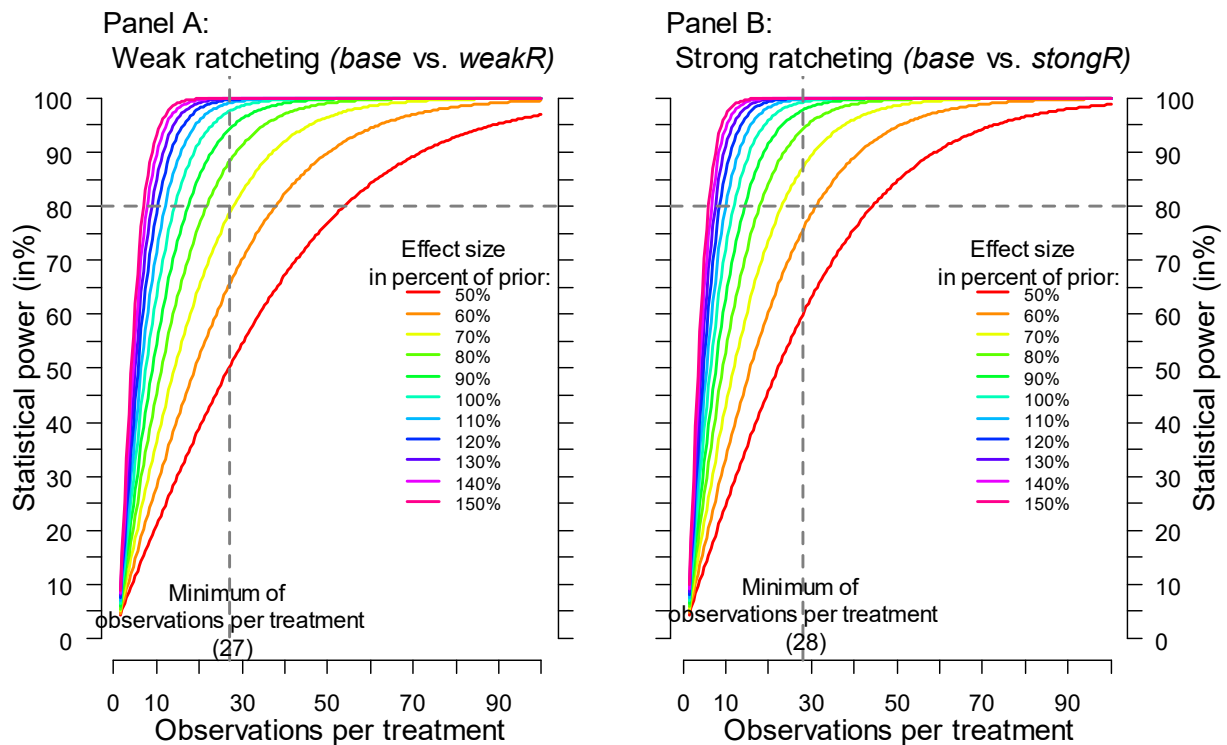
Analogously, we obtain the individual contribution level in the social optimum

$$g_{i,1}^{strongR^\circ} = w + \frac{n\gamma - \alpha}{2\beta} - 2, g_{i,2}^{strongR^\circ} = w + \frac{n\gamma - \alpha}{2\beta} - 1, g_{i,3}^{strongR^\circ} = w + \frac{n\gamma - \alpha}{2\beta},$$

$$g_{i,4}^{strongR^\circ} = w + \frac{n\gamma - \alpha}{2\beta} + 1, g_{i,5}^{strongR^\circ} = w + \frac{n\gamma - \alpha}{2\beta} + 2$$

## Appendix 2: Power calculations

Figure A2.1: Power curves – *weakR* and *strongR*



Note: Panel A (B) shows the power curves in *weakR* (*strongR*). For each treatment we simulate a variety of different treatment effects ranging from 50 percent to 150 percent of the prior.

We derive our priors for the effect sizes of the ratcheting treatments from a series of three pilot sessions conducted before the actual experimental sessions. 52 subjects participated in our pilot and we calculated average treatment effects to gain first priors. Based on these priors, we simulated a variety of treatment effects ranging from 50 percent to 150 percent of our initial estimates. The calculations are based on two-sided t-tests of means per treatment with independent samples conducted in R<sup>15</sup> using the “pwr”<sup>16</sup> package and illustrated for *weakR* and *strongR* separately in Panel A and B of Figure A2.1. As shown in Figure A2.1, we substantially oversample participants. With a minimum of 27 observations per treatment cell, we are able to detect a treatment effect comparable to our prior in *weakR* at a very high level of statistical power of more than 99 percent (see Figure A2.1 – Panel A). At conventional levels of statistical significance (5 percent) and power

<sup>15</sup> R Core Team (2018), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

<sup>16</sup> Champely, S. (2018): pwr: Basic Functions for Power Analysis. R package version 1.2-2. <https://CRAN.R-project.org/package=pwr>

(80 percent), 12 observations per treatment cell would have been sufficient to detect a treatment effect similar to the effect observed in our pilot sessions. Similarly, in *strongR* we have a minimum of 28 observations per treatment cell where around 12 observations would have been enough to detect a treatment effect similar to the effect observed in our pilot at conventional levels of statistical significance (5 percent) and power (80 percent) (see Figure A2.1 – Panel B). Based on the pilot, we oversample participants also in *strongR* and are able to detect an effect comparable to the effect in our pilot sessions at a very high level of statistical power of more than 99 percent.



## Appendix 3: Tables

**Table A3.1: Exploitation effect per Phase – Individual level analysis**

Dependent variable: Cumulated cooperation rates per Phase				
	(1)	(2)	(3)	(4)
	Phase II	Phase III	Phase IV	Phase V
<i>exploitation</i>	-2.849 (4.634)	-0.983 (4.165)	-0.196 (3.980)	-0.461 (4.052)
<i>weakR</i>	-5.447 (4.607)	-1.572 (3.974)	0.819 (3.984)	-4.979 (3.945)
<i>strongR</i>	-6.093 (4.614)	-4.111 (4.008)	0.524 (3.879)	-8.716** (3.972)
<i>exploitation x weakR</i>	-13.050** (6.470)	-8.294 (5.717)	-6.455 (5.580)	0.840 (5.690)
<i>exploitation x strongR</i>	-11.150* (6.430)	-11.051* (5.642)	-4.606 (5.560)	-6.044 (5.625)
<i>sum con Phase prev</i>	0.719*** (0.052)	0.664*** (0.044)	0.812*** (0.044)	0.628*** (0.042)
Constant	11.877*** (3.575)	9.785*** (2.999)	4.758 (2.957)	11.929*** (2.903)
Observations	340	340	340	340
Adjusted R <sup>2</sup>	0.457	0.500	0.568	0.495

Note: OLS regressions. Standard errors in parentheses. The dependent variable is cumulated cooperation rate in Phase II (model 1), Phase III (model 2), Phase IV (model 3), and Phase V (model 4) at the individual level. *exploitation* indicates that participants contributed more (1) or less (0) than their group members in the previous phase. *sum con Phase prev* captures participants' cumulated cooperation rates in the previous Phase. Furthermore, we include indicator variables for the different ratcheting treatments as well as their interactions with gap and exploitation. \*  $p < 0.10$ , \*\*  $p < 0.05$ , and \*\*\*  $p < 0.01$ .

## Appendix 4: Instructions

### Instructions (weakR)

Welcome to Magdeburg Experimental Laboratory MAXLAB!

Please read the instructions carefully and indicate by opening your door or raising your hand if you have any questions.

In the experiment in which you are participating now you can earn money depending on your decisions and the decisions of your fellow players. Your payoff from the experiment is calculated in LaborDollar (LD) at an exchange rate of **1:125** between € and LD, i.e., 125 LD equal 1 €.

You make your decision in the experiment **anonymously**. Only the experimenter learns your identity; your information is treated confidentially.

#### Experiment rules

In total, **four players** participate in the experiment, i.e., apart from you, there are three more players. The three participants you interact with remain the same in the entire experiment.

The experiment comprises **five phases**. In each of the five phases, you play the same **game**.

Every play consists of **five rounds**. Thus, in total, you play  $5 \times 5 = 25$  rounds. At the start of each round, you and your three fellow players each receive an **initial endowment of 100 LD**. Thus, in the course of five rounds, i.e., per phase, you receive 500 LD in total.

Your task, and the task of your fellow players, consists in deciding how many LD you want to contribute to a **joint project**. Your contribution can vary between 0 and 100 LD (only integers) in round 1. In all other rounds, your contribution needs to be **at least as high as in the previous round**. The maximum contribution is 100 LD.

**Your payoff for every round** is determined as follows:

$$= 4.4(100 - \text{your contribution}) - 0.02(100 - \text{your contribution})^2 \\ + (\text{your contribution} + \text{contributions of the other three players}) - 100$$

That is, for example, if all the other three players contribute on average 30 LD to the project in each round and you contribute 90 LD, **your payoff** will be:

$$= 4.4(100 - 90) - 0.02(100 - 90)^2 + (90 + 3 * 30) - 100 = 44 - 2 + 180 - 100 = 122$$

If, on the other hand, all other players contribute on average 30 LD and you contribute nothing at all, **your payoff** will be:

$$= 4.4(100 - 0) - 0.02(100 - 0)^2 + (3 * 30) - 100 = 440 - 200 + 90 - 100 = 230$$

On your table, you can find a “**payoff table**”, which helps you calculate your payoff. In this payoff table, you can select the **average** contribution of all the other players (column) and your own contribution (line) (both in intervals of 5) for every round. In the respective cell of the table, you can then find your payoff per round in LD.

At the end of the experiment, you receive the **payoff from one of these five phases** in € (in accordance with the exchange rate stated above). The phase for payoff is **randomly** chosen.

Therefore, you should act in **each** round as if it were relevant for payoff. At first, **two practice rounds** take place, which are **not relevant for payoff**.

### Control questions (weakR)

If you have read the instructions and do not have any questions, please answer the following control questions (Tip: use the payoff table).

1. Assuming your contribution to the project amounts to 20 LD while the average contribution of all other players is 50 LD: what is your payoff (in LD) in that round?  
My payoff is: \_\_\_\_\_
2. Assuming your contribution to the project amounts to 60 LD while the average contribution of all other players is 20 LD: what is your payoff (in LD) in that round?  
My payoff is: \_\_\_\_\_
3. Assuming that the average contribution of all other players is 15 LD: with which of the following contributions can you achieve the highest payoff in that round (please tick)?  
 5 LD             10 LD             15 LD             20 LD
4. Assuming that you want to maximize your payoff: does it make sense to make no contribution (i.e. zero) to the project?  
 yes             no
5. Assuming that all players choose to make the same contribution: which of the following contributions leads to the highest total payoff to all players in a round (please tick)?  
 0 LD     50 LD     70 LD     90 LD     100 LD
6. Which rule applies to your payoff?  
 One of the  $5 \times 5 = 25$  rounds is randomly chosen for payoff.  
 One of the five phases, which consists of five rounds, is randomly chosen for payoff.  
 All  $5 \times 5 = 25$  rounds are paid off.
7. Assuming that you have contributed 20 LD to the joint project in round 1 of a phase: how much do you need to contribute at least to the joint project in round 2?  
\_\_\_\_\_ LD
8. Assuming that you have contributed 20 LD to the joint project in round 1 of a phase: how much do you need to contribute at least to the joint project in round 5, i.e., in the last round of the phase?  
\_\_\_\_\_ LD

Once you have answered all questions, please give us a sign. We will then check your answers. The experiment starts after all subjects have passed the test.

Good luck in the experiment! The MaXLab team

## Appendix 5: Screenshots

Figure A5.1: Input screen for treatment weakR

Period 3 out of 5 Remaining time 13

### Contribution decision - Phase 1

Input Technisches Ausschneiden

Your task (as well as the task of your co-players) is to decide how much LabDollar you are willing to invest into the **joint project**.

**Note: You have to contribute at least as much as in the previous period.**

My contribution is

**Note:** In the previous period you contributed 15. Now you have to contribute at least 15. continue

Figure A5.2: Output screen for treatment weakR

Period 3 out of 5 Remaining time 22

### Contribution decision - Output - Phase 1

Technisches Ausschneiden

Period	contribution q1	contribution q2	contribution q3	contribution q4	av. contribution	payoff 1	payoff 2	payoff 3	payoff 4	av. payoff
1	15	15	15	15	15.00	189.50	189.50	189.50	189.50	189.50
2	15	15	15	15	15.00	189.50	189.50	189.50	189.50	189.50
3	15	15	15	15	15.00	189.50	189.50	189.50	189.50	189.50

---

You are **player 1**.

**Information about the current period:**

Your contribution q: 15  
Sum of all contributions Q: 60  
Average contribution: 15.00

Your payoff: 189.50  
Average payoff: 189.50

**Information about the current phase:**

Your payoff: 568.50

OK



Download ZEW Discussion Papers from our ftp server:

<http://ftp.zew.de/pub/zew-docs/dp/>

or see:

<https://www.ssrn.com/link/ZEW-Ctr-Euro-Econ-Research.html>

<https://ideas.repec.org/s/zbw/zewdip.html>



**IMPRINT**

**ZEW – Leibniz-Zentrum für Europäische  
Wirtschaftsforschung GmbH Mannheim**

ZEW – Leibniz Centre for European  
Economic Research

L 7,1 · 68161 Mannheim · Germany

Phone +49 621 1235-01

info@zew.de · zew.de

Discussion Papers are intended to make results of ZEW research promptly available to other economists in order to encourage discussion and suggestions for revisions. The authors are solely responsible for the contents which do not necessarily represent the opinion of the ZEW.