



Can we manage first
impressions in cooperation
problems?

An experiment

Christoph Engel
Sebastian Kube
Michael Kurschilgen





Can we manage first impressions in cooperation problems?

An experiment

Christoph Engel / Sebastian Kube / Michael Kurschilgen

March 2011

this version May 2014

Can we manage first impressions in cooperation problems?

An Experiment

Christoph Engel Sebastian Kube Michael Kurschilgen

Abstract

We study how cooperative behavior reacts to selective (favorable or unfavorable) pre-play information about the cooperativeness of other, unrelated groups within an experimental framework that is sufficiently rich for conflicting behavioral norms to emerge. We find that cooperation crucially depends on pre-play information, coinciding with a change in initial beliefs. Over time, behavior within both types of groups becomes increasingly homogeneous, indicating the formation of two rather different social norms, depending on whether pre-play information was favorable or unfavorable. In addition, we find unfavorable information to substantially reduce the effectiveness of peer punishment. For these differences to emerge it is immaterial whether each member or only one member of a four-person group receives the pre-play information.

JEL: C90, D63, H41

Keywords: Cooperation, Effectiveness of Sanctions, Information, Expectations, Norms, Experiment

1. Introduction

Cooperation problems are at the heart of many everyday situations. For example, when it comes to protecting the environment, defending one's country, generating new knowledge, joining a political party, extending the infrastructure, or exploiting the opposite market side, agents face a social dilemma. Jointly they are best off if everyone contributes her fair share. But individually, free-riding on others' efforts yields the highest payoff. A large number of theoretical and empirical papers have explored how people should be expected to and how they actually do behave in such situations. Many empirical contributions make use of laboratory experiments where subjects participate in a prisoners' dilemma or in a public-good game. Absent institutional interventions like punishment (Fehr and Gächter 2000a), contributions are heterogeneous in the beginning, but average cooperation quickly declines and most participants free-ride in the end (e.g. Andreoni 1988). These results have been stress-tested extensively, e.g., with respect to anonymity (Andreoni and Petrie 2004), culture (Herrmann et al. 2008), group size (Isaac et al. 1994), efficiency (Glöckner et al. 2010), or framing (Goerg and Walkowitz 2010). Interventions that mitigate the dilemma are not easily designed. Effective interventions tend to be heavy-handed, often altering the incentive structure of the game such that free-riding is not in an individual's self-interest any more (e.g. Falkinger et al. 2000, Fehr and Gächter 2002, Glöckner et al. 2010, Güerck et al. 2006, Ostrom et al. 1992). In this paper, we propose a very simple and light-handed mechanism and test its effectiveness by using laboratory experiments.

Our mechanism is based on an observation that has previously been made in public-goods experiments, namely that (initial) group composition and initial cooperation rates significantly affect the future development of cooperation rates in a particular group (Burlando and Guala 2005, Engel et al. 2014, Gächter and Thöni 2007). A likely reason for this behavioral pattern is provided by the influential concept of reciprocity and conditional cooperation (Dufwenberg and Kirchsteiger 2004, Falk and Fischbacher 2006, Fischbacher and Gächter 2010, Fischbacher et al. 2001). If persons are sensitive to the behavior of other group members, those groups who cooperate little in the beginning will become even less cooperative over time, while groups with substantial cooperation in the beginning are able to sustain cooperation over time. This suggests that outsiders might be able to moderate cooperation by manipulating experiences. Our mechanism is even less invasive in that it is confined to the first impressions subjects happen to make. More precisely, the idea is to moderate initial beliefs by providing participants with selective information about behavior in other, unrelated, groups. Arguably this information affects initial beliefs about others' behavior, which in turn changes how I myself behave initially. A virtuous (or vicious, depending on the kind of information that is provided) cycle starts: our manipulation affects cooperativeness in the beginning, which in turn affects cooperativeness later on.

Our intervention idea is in the spirit of James Q. Wilson's Broken Windows Theory (Wilson and Kelling 1982) which is based on the work of American psychologist Philip Zimbardo. In 1969, Zimbardo abandoned two identical cars in two different locations: the Bronx, NYC and Palo Alto, California. "The license plates of both cars were removed and the hoods opened to provide the necessary releaser signals" (Zimbardo 1969). In the Bronx, the abandoned car was stripped

and demolished after only 26 hours, the result of 23 separate incidents of vandalism. In contrast, the car in Palo Alto still sat unmolested after the course of an entire week. Zimbardo then decided to provide an example of vandalism to the affluent and seemingly non-violent neighborhood of Palo Alto. So he and two graduate students of his took a sledgehammer and started bashing the car. After they had taken the first blow, observers shouted encouragement and finally joined in the vandalism, until the car was completely wrecked (Zimbardo and Ebbesen 1969).

In order to clearly identify the short- and long-run effects of our intervention, we run laboratory experiments, which have the benefit of providing a sufficient degree of control over the environment. The environment that we use has been prominent lately to study social dilemmas and cooperation problems. It is a complex public-good game with decentralized sanctions (Fehr and Gächter 2000a) and counter-punishment opportunities (Nikiforakis 2008). We have chosen this rich paradigm, since it mimics several potentially important features available in natural environments. Moreover, it is sufficiently complex, so that conflicting behavioral norms might emerge. We expect belief manipulations to be particularly effective if the situation lends itself to competing expectations: participants might be optimistic since there is a sanction mechanism which could keep those tempted to freeride in check; or they might be pessimistic since those receiving a sanction have power to strike back which might deter punishment. We attempt to affect beliefs the following way: prior to making their first contribution decision in the game, we provide subjects with selected data from previous experiments. The information that we give is either favorable for cooperation (treatment FAV) or unfavorable (treatment UNFAV).

Initial cooperation rates between FAV and UNFAV differ substantially. The difference appears to be largely driven by initial beliefs, which in turn are highly susceptible to pre-play information. The difference does not narrow over time. Instead, the selective pre-play information presented to subjects appears to establish two distinct and robust norms of behavior, as evidenced by an increasing degree of behavioral homogeneity over time. Moreover, the difference in punishment patterns between the two treatments suggests that pre-play information not only affects subject's behavioral expectations ("how will others behave") but also their normative expectations ("how should one rightfully behave"). In FAV, a person who is sanctioned reacts immediately by substantially raising her contribution. This beneficial effect of sanctions almost disappears in UNFAV. Strikingly, a similar effect, both in terms of contributions and effectiveness of sanctions, is achieved when only one out of four group members receives either the positive (treatment 1FAV) or the negative (treatment 1UNFAV) information.

The next section embeds our study into the existing literature. Subsequently, we explain the design of the experiment. Section 4 presents and discusses the experimental results. Section 5 reports some additional treatments enhancing the robustness and precision of our results. Section 6 concludes the paper.

2. Related Literature

Our study closely relates to the experimental literature on reciprocity and conditional cooperation. This literature suggests that a substantial fraction of a typical subject pool consists of conditional cooperators, i.e. individuals who are willing to cooperate, provided they expect a sufficiently large fraction of the population to do the same (Fischbacher and Gächter 2010, Mengel 2007). For conditional cooperators, information about contribution patterns in other groups has a first- and a second-order effect. Given the uncertainty about the composition of the group of which they happen to be a member, information about other groups helps them form beliefs. Moreover, each participant knows that each other participant has received the same information. This makes it possible to also form a second-order belief, based on knowing that the remaining group members have received the same information.

Our research question is also related to the literature that explores interventions aiming at raising contributions to public goods. Usually, the proposed mechanisms alter the incentive structure such that free-riding becomes less attractive, or even a dominated strategy for self-centered money maximizers. For example, in Falkinger et al. (2000) the payoff structure is changed such that each individual gets a reward or has to pay a penalty depending on the deviation of its contribution from the mean contribution. In other studies, group composition is changed such that the payoff structure is changed implicitly. For example, in Gunnthorsdotir et al. (2007), subjects are re-matched every period according to their cooperativeness in the previous round; which is found to raise cooperation levels. Likewise, if groups have a chance to exclude free-riders, this improves cooperation in a dilemma setting (Cinyabuguma et al. 2005, Croson et al. 2008), as does a mechanism that allows members to self-select into groups (Page et al. 2005), in particular if free-riders are effectively excluded by a rule that sacrifices a portion of the group income to outsiders (Brekke et al. 2009). Our study differs from this literature in that we leave the incentive structure of the game completely unchanged. All we alter are the first impressions participants happen to make – which is an option that should naturally be available in any public good game, and in many real-life social dilemmas.

Parts of the legal literature, particularly those at the intersection of law and economics, have also been asking how socially desirable behavior and/or compliance with the law can be brought about. Again, a prominent approach is to change the incentive structure, e.g., by increasing the expected costs of breaking the law. For example, Braga et al. (1999) report evidence from a field experiment that randomly exposed 12 of 24 matched violent crime places in Jersey City to intense police scrutiny and intervention. In the places chosen, crime rates dropped substantially, while they did not in the unaffected places. In a similar vein, in a series of sociological field experiments, when there were signs of disorder, like graffiti, abandoned shopping carts, littering or bicycles locked where they were not supposed to be, this induced passers-by to also break these and other rules (Cialdini et al. 1990, Keizer et al. 2008, Ramos and Torgler 2010). In laboratory experiments, Galbiati and Vertova (2008) demonstrate that cooperation behavior increases when an explicit expectation is spelt out and enforced by (non-deterrent) sanctions (see also Kube and Traxler 2010).

Our study also connects to the literature on social norms and morality. Theoretically, that literature has been revolving around the notion of a dual self, selfish on the one hand and pro-social, moral on the other hand. Akerlof and Kranton (2000) capture this in a utility function where optimal choice reflects a trade-off between material (selfish) utility and identity utility. A person's identity utility is supposed to decrease as her actual behavior deviates from her own normative ideal. Bénabou and Tirole (2011) convey a similar idea in a game of self-signaling. People who care intrinsically about being a good person need to invest into their moral identity by behaving socially. Experimentally, Dal Bó and Dal Bó (2009) provide subjects in a public good game with definitions of moral behavior. They find that levels of cooperation increase but still decay over time. In a standard dictator game, Bicchieri and Xiao (2009) manipulate dictators' empirical expectations by telling them what the majority of participants did in previous studies as well as their normative expectations by telling them what the majority thought should be done. They find empirical expectations having a much stronger effect on behavior than normative expectations.

Finally, our findings also underline the power of information, as it has also been observed in other contexts. For example, previous work on voting behavior points out that information in the form of polls (Forsythe et al. 1993, Forsythe et al. 1996, Klor and Winter 2007) or of cheap-talk electoral campaigns (Corazzini et al. 2010) affect subsequent voting outcomes. Similarly, information gathered during pre-play communication¹ strongly affects subjects' decision in subsequent coordination games (Blume and Ortmann 2007), as does information from preceding asset-market outcomes (Kogan et al. 2010) or information about group members' previous decisions (Weber 2006). Interestingly, the literature on coordination games comes to the conclusion that pre-play information promotes Nash-equilibrium play (assuming self-centered money maximizers). By contrast, our experimental results show that the opposite might happen in a cooperation game like the one reported here, where the socially efficient outcome is usually not part of the subgame-perfect Nash equilibrium. This is noteworthy since, for conditional cooperators who expect to interact with other conditional cooperators, the original dilemma is transformed into a coordination game (about the degree of cooperation).

In other settings, however, it has been shown that information may well lead to behavior away from Nash-equilibrium. Whilst Fehr and Rockenbach (2003), do not find a change in subjects' behavior in a gift exchange game with punishment, Berg et al. (1995) show in a trust game that providing a social history increases cooperation. Bohnet and Zeckhauser (2004) report that informing responders about the average offers before they decide whether to accept or reject their specific offer increases offers and offer-specific rejection probabilities. In a binary dictator game Krupka and Weber (2009) find that showing subjects what others actually do produces more pro-social behavior. Interestingly, this is even the case when observed subjects are mostly selfish. They also find support for an informational effect: observing more people behaving pro-socially generally produces more pro-social behavior. Similar findings have been made in the field. Frey and Meier (2004) show that students are more likely to donate to a charity when they are in-

1 See also Crawford (1998) for a general survey of experiments on communication via cheap talk.

formed that previously a high fraction of other students from the same university has given. Likewise, if donors to a public radio were informed that another member had made a very high donation, they contributed significantly more themselves (Shang and Croson 2009). A similar effect of social information has been shown on the frequency by which users of an online platform for the rating of movies enter ratings themselves (Chen et al. 2010). Our results show that information about others playing the same game even influences behavior in a textbook public good, although this is a dilemma. Our experiment further shows that the effect is present even if participants know that the information is selective, not representative. Most importantly, our experiment is designed to not only show the effect of such information, but to also better understand the process by which it changes behavior.

3. Experimental Design, Procedure, and Behavioral Predictions

Experimental Paradigm: The experimental paradigm used in this study is a public-good game with punishment and counter-punishment opportunities as implemented by Nikiforakis (2008).² The basic game features $n=4$ players and consists of three stages: At the beginning of stage 1 (“contribution stage”),³ players are endowed with 20 tokens each. Players then decide simultaneously and independently how much of their endowment they want to contribute to a public account. We denote this decision with c_i . Each token that is contributed to the public account increases the payoff of each player in the group by 0.4 tokens (i.e., the MPCR is $\alpha=.4$). Each token unspent increases a player’s own payoff by one token. The preliminary payoff at the end of stage 1 is thus given by $\pi_i^1 = 20 - c_i + 0.4 \sum_{h=1}^4 c_h$.

Table 1: Punishment points p_{ij} per player j and associated costs $C(p_{ij})$ for punisher i

p_{ij}	0	1	2	3	4	5	6	7	8	9	10
$C(p_{ij})$	0	1	2	4	6	9	12	16	20	25	30

At the beginning of stage 2 (“punishment stage”), players are informed about every group member’s contribution to the public account. Every player i then has the opportunity to reduce the income of each other group member j by assigning costly punishment points p_{ij} . Each punishment point received reduces a player’s income from stage 1 by ten percent. At the same time,

2 We have chiefly chosen this design because it provides a rich environment in which different behavioral norms may emerge. Moreover, this design has the advantage to leave sufficient room for testing the effect of both favorable and unfavorable first impressions. Had we tested a mere voluntary contribution mechanism, previous experiments would have indicated that cooperation is very difficult to sustain in the first place. Conversely, had we chosen a public good game with one-step punishment, successful cooperation would have been very likely. By contrast, according to the existing literature, if we also add the counter-punishment option, expectations are in the middle between both extremes.

3 In the instructions and on the computer screens we only speak of „stage“, and do not use explanatory names; see instructions for detail.

each punishment point assigned reduces one's own payoff according to the cost function given in Table 1.⁴ The preliminary payoff at the end of stage 2 is thus given by:

$$\pi_i^2 = \pi_i^1 \frac{1}{10} \max \left\{ 0, 10 - \sum_{j \neq i} p_{ji} \right\} - \sum_{j \neq i} C(p_{ij}).$$

At the beginning of stage 3 ("counter-punishment stage"), players observe who punished them by how much in stage 2. They then have the opportunity to counter-punish the punishers by assigning them counter-punishment points cp_{ij} . The punishment technology is the same as in stage 2. Each counter-punishment point received reduces a player's preliminary income from stage 2 by ten percent and each counter-punishment point assigned reduces one's own payoff as given in Table 1, but the cost of counter-punishment also depends on this player's punishment decision. That is, if i assigns counter-punishment points to j , the specific costs of the counter-punishment points depend on the number of punishment points that i has assigned to j on stage two. For example, if i had already assigned four points to j on stage 2, assigning him a single counter-punishment point on stage three costs 3 tokens. Hence the costs of counter-punishment are given by $C(p_{ij} + cp_{ij}) - C(p_{ij})$. The final payoff at the end of stage 3 is thus given by:

$$\pi_i^3 = \pi_i^2 \frac{1}{10} \max \left\{ 0, 10 - \sum_{j \neq i} cp_{ji} \right\} - \sum_{j \neq i} [C(p_{ij} + cp_{ij}) - C(p_{ij})]$$

The only difference between the two treatments (FAV and UNFAV) is the information subjects receive before playing the game. On a one-page information sheet they are given selected information about unrelated groups who previously played this game – and players know that the information is actually taken from previous experiments.⁵ They are also told that the information is selective and that all players in their group receive the same information. The information we give is the development of mean contributions over time from four selected groups. Furthermore, we provide participants with selected data about the number of persons choosing to contribute zero in the first, resp. in the last period, and the average amount of counter-punishment that was meted out in these experiments. The way this data is presented and the selection of the four groups are such that they constitute a positive (i.e. cooperative) impression in treatment FAV, while they constitute a negative (i.e. uncooperative) impression in treatment UNFAV. Thus, the treatment manipulations allow us to study to what extent different contents of pre-play information affect behavior.

Procedure: The experiment was run at the Laboratory for Experimental Economics at the University of Bonn (Germany). All experiments were programmed using z-Tree (Fischbacher 2007). Subjects were invited using ORSEE (Greiner 2004). Subjects were not allowed to participate in more than one treatment (between-subject design). When subjects arrived in the lab, they were seated in separate cubicles. The experimental instructions were then handed out to them and read out aloud in order to create common knowledge and to ensure that everybody had read and un-

4 This non-linear punishment technology has been introduced in the paper by Fehr and Gächter (2000) that has started the literature, and has become standard.

5 The information was taken from the datasets of Nikiforakis (2008) and Engel et al. (2014). Both used the same experimental paradigm as the present study.

derstood the instructions. Additionally, subjects had to take a quiz and could pose comprehension questions in private before the game started. The instructions were written in neutral language, avoiding potentially loaded terms like punishment or public good (cp. Appendix 1).

After subjects had finished reading the instructions, they were provided with the additional information sheet (cp. Appendix 3 for FAV and Appendix 4 for UNFAV). Subjects were then randomly divided into groups of four and played the above-described game repeatedly for ten consecutive periods. A partner protocol was used, i.e., the group composition stayed constant over the entire 10 periods of anonymous interaction. In the end, subjects were privately paid their cumulated earnings and left. Participants received their accumulated earnings from the experiment (1 token = Euro 0.04) and an additional show-up fee of 5 Euro. On average, a session lasted 60 minutes and subjects earned Euro 14.46.

Behavioral Predictions: For rational, self-centered money maximizing players, the game at hand is a cooperation problem. Using backwards induction, it is straightforward that the unique subgame perfect equilibrium is to contribute nothing to the public good at the first stage, and not to punish nor counter-punish at subsequent stages. This leads to an equilibrium payoff of 20 tokens per period and player. By contrast, the socially efficient outcome would be achieved if everyone contributed their entire endowment, and nobody punished nor counter-punished, in which case every player would earn 32 tokens per period. But in that case, each player would have an individual incentive to free-ride on the others' contributions, which would yield her a payoff of 44 tokens, *ceteris paribus*. Since we announce the number of periods, through unraveling this is also the prediction for the repeated game. Providing subjects with additional information about other groups does not change the equilibrium prediction. Therefore, with rational, self-centered money maximizing players we should not expect to observe different behavior in FAV and UNFAV.

By contrast, things might change as soon as we allow for social preferences, in particular if we expect a substantial fraction of the population to act as conditional cooperators. If conditional cooperators are sufficiently optimistic about the cooperativeness of their interaction partners, cooperation gives them higher utility than defection. Facing a public good, conditionally cooperative subjects should therefore base their decision on what they believe other subjects to do. The beliefs are likely to be related to other group members' behavior in the previous period(s). Yet, upon the first encounter (in the first period), a conditional cooperator needs to form initial expectations of how others are likely to behave. This is where our treatment manipulation might make a difference. Pre-play information about other, unrelated groups might influence a participant in forming initial beliefs, which should then guide her first-period decision. If her (conditionally cooperative) group peers were also influenced by the selective pre-play information, then group contributions in the first period are likely to mirror the pre-play information. Her beliefs for the second period will thus be very similar again, and so will contributions be in the second period; and similarly in subsequent periods. In that case, we would expect to observe more cooperation in FAV than UNFAV throughout the game.

4. Experimental Results

A. Contribution Behavior

Does pre-play information change the way subjects behave in a social dilemma? Figure 1 suggests that the effect is substantial. Groups that were shown the favorable (FAV) information contributed on average 16.98 tokens to the public good over the entire duration of the game whereas those who had seen the unfavorable information managed only 10.71 tokens. The difference of contributions is highly significant (Mann-Whitney ranksum test, two-sided, $N=33$, $p=0.0003$). In fact, the drastically different level of cooperation between FAV groups and UNFAV groups starts already in the initial period of interaction where FAV groups contribute on average 14.79 tokens and UNFAV groups only 9.78 tokens. Also this difference is highly significant (Mann-Whitney ranksum test, two-sided, $N=132$, $p<0.0001$).

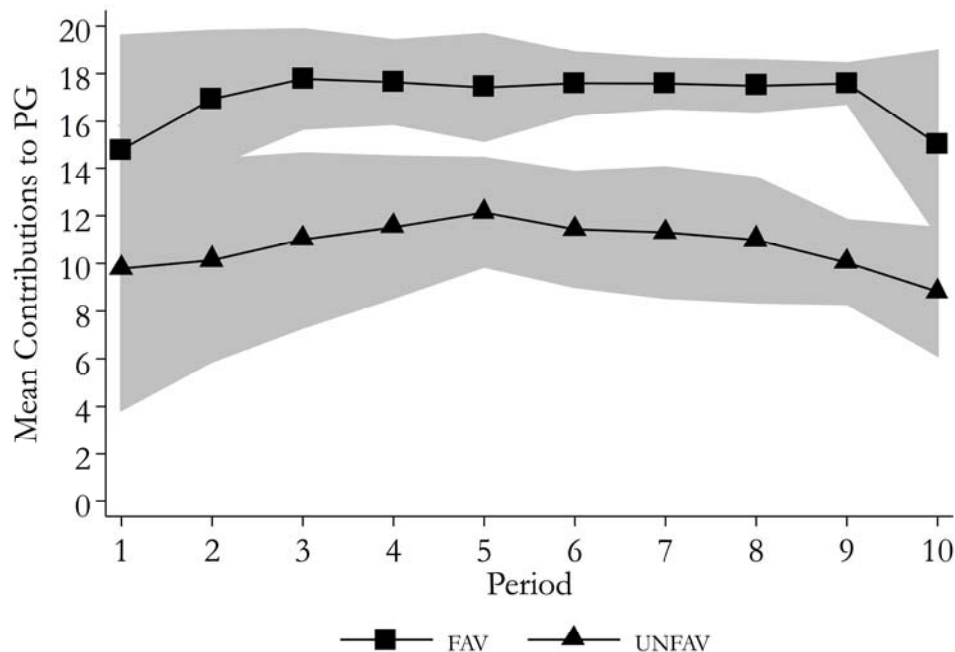


Figure 1: Public Good contributions in FAV and UNFAV

Note: FAV has 72 subjects (18 groups), UNFAV has 60 subjects (15 groups). The shaded areas around the treatment means show one standard deviation, as a measure of within-group heterogeneity. Standard deviations are calculated per group and then averaged over all groups of a treatment.

Strikingly, as Figure 1 illustrates, average behavior in the two treatments does not converge over time. On the contrary, mean contributions in the last period of interaction amount to 15.04 tokens in FAV and 8.82 tokens in UNFAV (Mann-Whitney ranksum test, two-sided, $N=33$, $p=0.0206$). Convergence of behavior over time despite the different pre-play information would have suggested that subjects have a strong home-grown contribution norm that is only temporarily disturbed by selective information. Instead, the clear lack of convergence suggests that the selective pre-play information presented to subjects manages to establish two distinct and robust norms of behavior.

One way to measure the robustness of a norm is by the degree of conformity of behavior (see Bernheim (1994)). As the shaded areas of Figure 1 show, in both treatments the behavioral norm seems to become stronger over time as contribution behavior becomes more and more homogeneous within groups. In FAV (UNFAV), the standard deviation from the group mean decreases significantly and almost monotonically from 4.85 (5.95) in period 1 to 0.84 (1.77) in period 9 (Wilcoxon signed-rank test, two-sided. FAV: $N=18$, $p=0.0002$. UNFAV: $N=15$, $p=0.0008$). In period 10, heterogeneity increases again in both treatments, as the lack of the shadow of the future entices many participants to behave opportunistically while others stick to the norm. This endgame effect is considerably more pronounced in FAV where the payoff from defecting from the group's cooperation norm looms larger.⁶

The first period of interaction appears to be crucial as it sets the tone for the subsequent development of cooperation. In a population with a substantial fraction of conditional cooperators (Fischbacher and Gächter 2010), behavior in the first period of interaction will critically depend on people's beliefs about others' behavior. A conditional cooperator will contribute high (low) amounts if he expects others to also contribute high (low) amounts.

To verify this conjecture, we ran additional experiments at the Bonn Econ Lab. In these experiments, subjects received exactly the same instructions and had to answer the same control questions as in the games above.⁷ However, instead of actually playing the game they learned that these experiments had been conducted before. Their task was to guess how much those previous participants contributed on average to the public good in the first period (rounded to the next integer). For a correct answer they received 4 Euro, for an incorrect answer 0 Euro. Altogether, 48 fresh subjects participated in these experiments.

Our results show that beliefs in treatments FAV-B and UNFAV-B correspond indeed very closely with observed behavior in FAV and UNFAV. Subjects expect others to contribute on average 12.63 tokens after being shown the favorable selective information, and 7.83 tokens after the unfavorable information. The difference is highly significant (Mann-Whitney ranksum test, two-sided, $N=48$, $p<0.0001$).

B. Punishment Behavior

Figure 2 illustrates that the overall use of punishment is rather similar across both treatments. In both treatments punishment mainly occurs in the early periods of interaction, coinciding with high heterogeneity of contribution behavior. As behavioral differences within a group decrease,

6 Ceteris paribus and considering stage 1 payoffs only, defecting (i.e. contributing zero) from a group norm of 17 (as in FAV), yields a net gain of about 10 tokens while defecting from a group norm of 9 (as in UNFAV) yields only additional payoffs of about 5 tokens. This difference could of course be neutralized in stage 2 if punishment in the light of a norm violation happened to be substantially more severe in FAV than in UNFAV. We will discuss punishment behavior extensively below.

7 In the belief treatment FAV-B (UNFAV-B) subjects received the same instructions as the subjects of treatment FAV (UNFAV).

so does the amount of punishment. The fact that there is substantial punishment in the last period of interaction contradicts the idea of punishment being purely strategic (i.e. meant to educate free-riders in order to enjoy higher cooperation in the future). However, it is fully reconcilable with the idea that punishment is triggered by a violation of the punisher’s normative expectations (Bicchieri 2006), most notably in FAV where after nine periods of successful cooperation, some people opt to defect in period ten (see Figure 1).

In FAV subjects punish in 143 out of 2160 (6.62%) instances and in UNFAV in 149 out of 1800 (8.38%). On average, subjects in FAV allot 0.41 punishment points per period, compared to 0.51 in UNFAV. Ranksum tests over means per group confirm that these differences are not statistically significant. In both treatments punishment is predominantly pro-social, i.e. it is directed towards the free-riders in the group. In fact, the descriptive difference of punishment volumes is solely driven by the amount of anti-social punishment⁸: 0.02 punishment points per period in FAV, and 0.12 in UNFAV. The incidence of anti-social punishment doubles from 15 out of 143 cases (10.49%) in FAV to 32 out of 149 (21.48%) in UNFAV. This difference is highly significant both with respect to incidence (Mann-Whitney ranksum test, two-sided, N=33, p=0.0442) and total volume (p=0.0293).

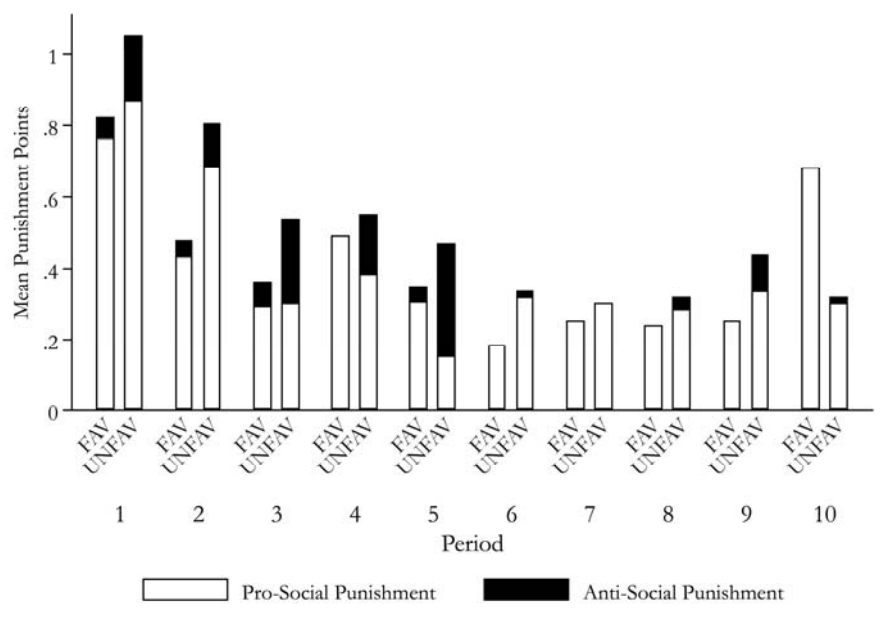


Figure 2: Punishment in FAV and UNFAV

Note: FAV has 72 subjects (18 groups), UNFAV has 60 subjects (15 groups). The graph illustrates the mean punishment points given by a subject in each period, split up into social punishment and anti-social punishment. We define as Pro-(Anti-)Social punishing a person who contributed less than (at least as much as) oneself.

According to Herrmann et al. (2008) anti-social punishment is an indicator of “weak norms of civic cooperation”. The authors seem to conceptualize norms as behavioral ideals, i.e. what is best for society. In that vein, the more (less) people cooperate, the stronger (weaker) one would

8 In line with (Herrmann, et al. 2008), we define as anti-social punishing a person who contributed at least as much as oneself.

judge the norm to be. Our data clearly support this conjecture as we observe high (low) anti-social punishment in the treatment with low (high) levels of cooperation. Interestingly, the conjecture also holds under the alternative conceptualization of norms as behavioral regularities. In UNFAV, where anti-social punishment is higher, cooperation is significantly less homogeneous⁹ than in FAV (Mann-Whitney ranksum test, two-sided, $N = 33$, $p = 0.0429$).

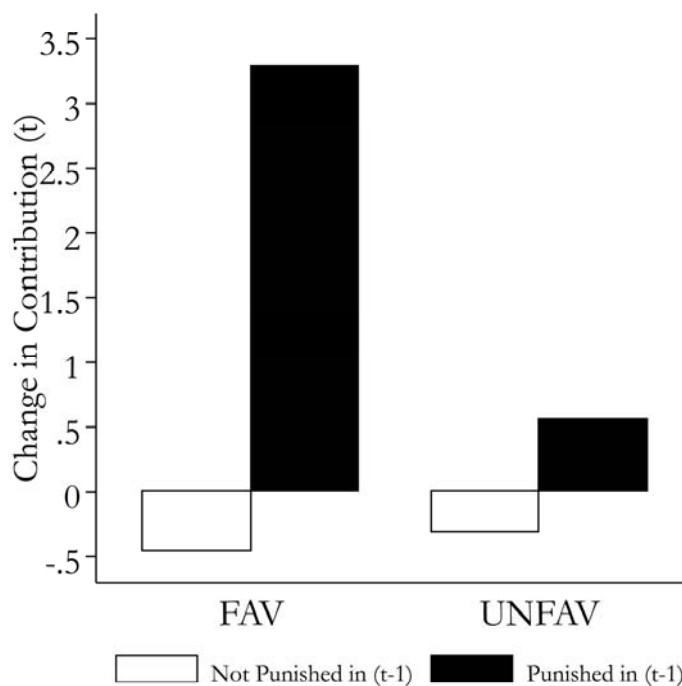


Figure 3: Reaction to being punished in FAV and UNFAV

Note: FAV has 72 subjects (18 groups), UNFAV has 60 subjects (15 groups). In FAV (UNFAV) there are 84 (127) instances of contributing after being punished in (t-1) and 564 (413) of contributing after not being punished in (t-1).

How do people respond to receiving punishment? In principle, there are two meaningful reactions: (1) strike back with counter-punishment, or (2) change one's contribution in the next period. By design, a subject could only use counter-punishment in stage 3 if she had been punished in stage 2 of that very same period. In FAV, subjects used counter-punishment in 36 out of 92 (39.13%) times they could. Descriptively, this figure increases slightly in UNFAV to 66 out of 137 (48.16%), which suggests that in UNFAV more people disagree with the punishment they received and take revenge on the punisher. Also the volume of counter-punishment per period and subject is lower in FAV (0.16 points) than in UNFAV (0.23 points). However, this difference is not statistically significant.

Fehr and Gächter (2000a) show that the beneficial effect of sanctions comes from people reacting to being punished with an immediate increase in contributions. As Figure 3 illustrates, this effect is present in FAV but not in UNFAV. The tendency in both treatments is to slightly decrease contributions in the subsequent period. However if they are punished, subjects in FAV react by increasing their contributions by 3.29 tokens. In contrast, participants who had seen the

⁹ Just as above, we compute the mean standard deviation for each of the 33 independent groups.

negative examples of behavior prior to playing the game (UNFAV) are considerably less sensitive to receiving a sanction and only increase their subsequent contribution by 0.55 tokens.

Table 2: Reaction to receiving punishment in FAV and UNFAV

DV: Change in Contribution (t)	(1)	(2)	(3)	(4)
UNFAV	0.148 (0.230)	0.148 (0.230)	0.148 (0.230)	0.487 (0.257)
Received punishment in (t-1)	3.743*** (0.432)	4.245*** (0.526)	4.438*** (0.608)	4.398*** (0.613)
Received punishment in (t-1) x UNFAV	-2.882*** (0.583)	-2.750*** (0.570)	-3.097*** (0.724)	-2.990*** (0.728)
Counter-punished in (t-1)		-1.277* (0.586)	-1.768* (0.821)	-1.651* (0.829)
Counter-punished in (t-1) x UNFAV			0.801 (1.166)	0.726 (1.150)
Others' Contribution in (t-1)				0.0572*** (0.0154)
Constant	-0.457** (0.156)	-0.457** (0.156)	-0.457** (0.157)	-1.443*** (0.355)
# observations	1,188	1,188	1,188	1,188
# subjects	132	132	132	132
# clusters	33	33	33	33
chi2	80.67	81.03	75.34	116.6
R squared (overall)	0.0604	0.0650	0.0655	0.0720

Robust standard errors, clustered by groups, in parentheses. *** p<0.001, ** p<0.01, * p<0.05. Random Effects Estimator.

Table 2 analyses subjects' reactions to receiving punishment in greater depth. In all four models, receiving punishment significantly increases contributions in the subsequent period but the effect is significantly smaller in UNFAV than in FAV. Notably, the effect is robust to controlling for the use of counter-punishment (columns 2-4) and the other group members' contribution level (column 4). This suggests that the negative examples in UNFAV not only affect people's behavioral expectations ("how will people behave") but also their normative expectations ("how should one rightfully behave"), at least for some participants. A person who contributes low because she thinks others will do so, too, seems more likely to increase her contribution after being punished than a person who contributes low because she believes low contributions are normatively appropriate.

5. Robustness

In the previous section, we have seen that pre-play information can have a substantial effect on people's willingness to cooperate. Our data from the belief treatments show that people expect others to initially contribute high (low) in FAV (UNFAV), which is exactly what happens in the contribution treatments. Moreover, over the course of the game subjects do not return to any

home-grown behavioral norm but rather two distinct norms emerge, one of high cooperation in FAV and one of low cooperation in UNFAV.

A. One informed player

On the basis of the results discussed so far, we cannot distinguish, however, whether selective information influences behavior directly through first-order expectations or indirectly through higher-order expectations. The former would mean that a player adapts her behavior because the pre-play information influences what she personally expects from the new environment. The latter would imply that a player adapts her behavior because she believes others will be influenced by the pre-play information (possibly because they themselves also believe others will be influenced).

To tackle this question, we run two additional treatments: 1FAV and 1UNFAV. 144 new subjects participated in these additional treatments. The 1FAV (1UNFAV) treatment is identical to the baseline with the sole exception that before starting the game only one out of the four members of every group saw an additional screen with the same selected information that all four members of a group saw in FAV (UNFAV). We call this group member the “informed player”. The informed player knew that her group peers were not informed. The three non-informed group members, on the other hand, did not know that there was an informed player. Put differently, in the FAV (UNFAV) treatments the selected examples of positive (negative) behavior in previous experiments were public information, whereas in 1FAV (1UNFAV) they were private information.

To avoid any interference with the behavior of the non-informed players, we first focus on the contribution decision in period 1. If higher-order expectations did not matter, the informed players in 1FAV (1UNFAV) would display the same behavior as players in FAV (UNFAV). As Figure 4 shows, this is exactly what happens. On average, the informed 1FAV players contribute in the first period 15.83 tokens, which is statistically indistinguishable from the 14.79 tokens of the FAV players (Mann-Whitney ranksum test, two-sided, $N = 90$, $p = 0.7128$). Similarly, the informed 1UNFAV players contribute in the first period 10.17 tokens, which is statistically indistinguishable from the 9.78 tokens of the UNFAV players (Mann-Whitney ranksum test, two-sided, $N = 78$, $p = 0.7738$). On the other hand, the difference between informed 1FAV and informed 1UNFAV players is highly significant (Mann-Whitney ranksum test, two-sided, $N = 36$, $p = 0.0111$), just as the difference between FAV and UNFAV (Mann-Whitney ranksum test, two-sided, $N = 132$, $p = 0.000$). As one would expect, there is no significant difference between the initial contributions of non-informed players in 1FAV and 1UNFAV (ranksum test, $N = 36$, $p = 0.1670$, two-sided).

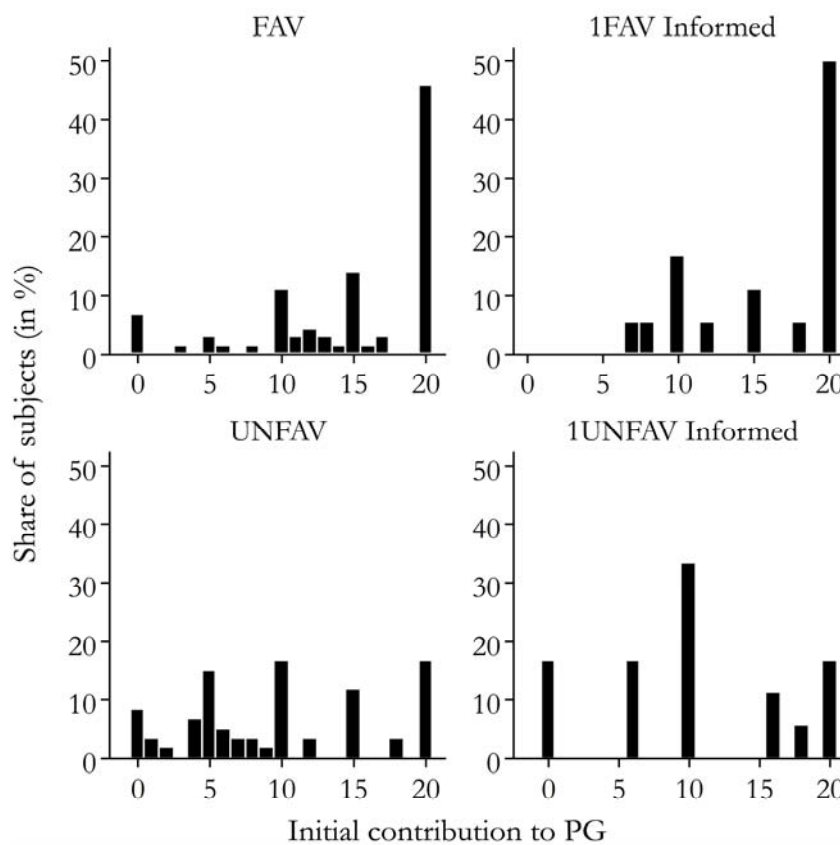


Figure 4: Initial PG contributions of the informed players in FAV, UNFAV, 1FAV, and 1UNFAV

Notes: FAV has 72 subjects, UNFAV has 60 subjects, 1FAV has 72 subjects, and 1UNFAV has 72 subjects. In FAV and UNFAV all players were informed whereas in 1FAV and 1UNFAV only one of the 4 members of each group was informed. Hence, there were 18 informed subjects in 1FAV and 18 in 1UNFAV.

Our results clearly suggest that selective information influences people’s behavior directly through altering their first-order expectations. The selective information seems to be used to update expectations, and to adjust behavior accordingly. To add robustness to this result, we ran again additional belief-treatments and asked new, unrelated subjects to guess how much the informed players in 1FAV and 1UNFAV would contribute in the first period. In total, we had 48 new subjects in treatment 1FAV-B and 1UNFAV-B.¹⁰ Beliefs in treatments 1FAV-B and 1UNFAV-B correspond again very closely to observed behavior. Subjects expect the informed players to contribute on average 14.17 tokens after being shown the favorable selective information, and 11.41 tokens after the unfavorable information. The difference is highly significant (Mann-Whitney ranksum test, two-sided, $N=48$, $p=0.0320$).

From the second period onwards, in every group of the 1FAV and 1UNFAV treatments three non-informed players interact with one informed player. Strikingly however, as shown in Figure 5, having one player selectively informed suffices to bring about a similar effect as informing all four players. Over all 10 periods of interaction, there is no significant difference between UNFAV and 1UNFAV (Mann-Whitney ranksum test, two-sided, $N=33$, $p=0.9712$), nor between FAV and 1FAV (Mann-Whitney ranksum test, two-sided, $N=36$, $p=0.4863$). In contrast, there is

¹⁰ The procedure was identical to treatments FAV-B and UNFAV-B. See above.

a highly significant difference between 1FAV and 1UNFAV (Mann-Whitney ranksum test, two-sided, $N=36$, $p=0.0026$).

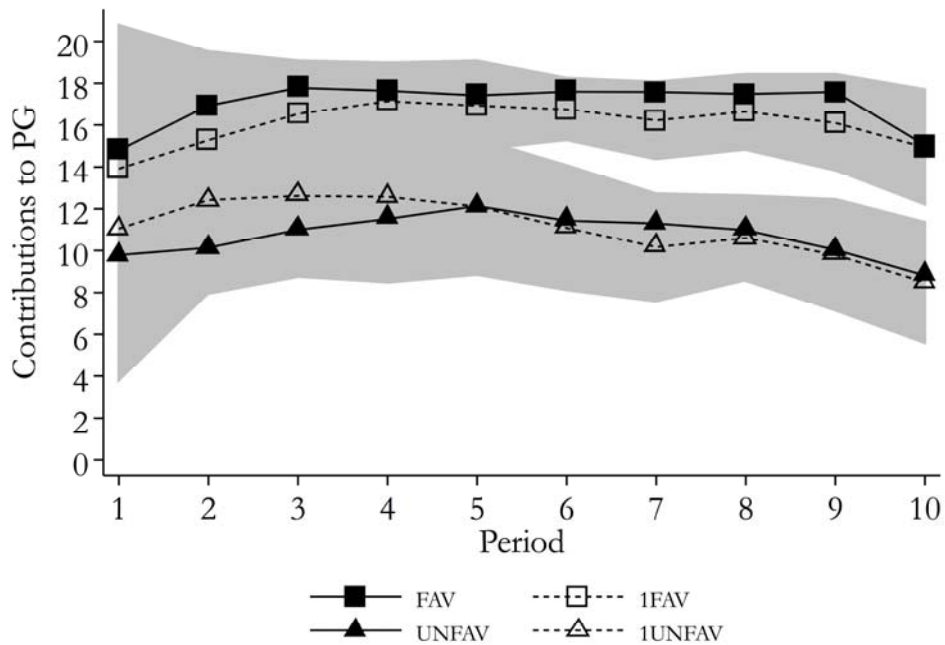


Figure 5: Public Good contributions in FAV, UNFAV, 1FAV, and 1UNFAV

Notes: FAV has 72 subjects (18 groups), UNFAV has 60 subjects (15 groups), 1FAV has 72 subjects (18 groups), and 1UNFAV has 72 subjects (18 groups). The shaded areas around the treatment means show the corresponding standard deviations of 1FAV and 1UNFAV, as a measure of within-group heterogeneity. Standard deviations are calculated per group and then averaged over all groups of a treatment.

Moreover, as the shaded areas in Figure 5 illustrate, also in 1FAV and 1UNFAV behavior within groups becomes increasingly homogeneous over time. In 1FAV (1UNFAV) the mean standard deviation decreases significantly from 6.87 (7.25) in period 1 to 2.30 (2.65) in period 9 (Wilcoxon signed-rank test, two-sided. 1FAV: $N=18$, $p=0.0012$. 1UNFAV: $N=18$, $p=0.0010$). Apparently, in this setting it is sufficient to selectively inform one out of four players to establish two distinct and robust behavioral norms.

The use of punishment in 1FAV and 1UNFAV is very similar with respect to the total volume of punishment (Mann-Whitney ranksum test, two-sided, $N=36$, $p=0.2228$), the use of anti-social punishment (Mann-Whitney ranksum test, two-sided, $N=36$, $p=0.2519$), and the use of counter-punishment. Just as with the comparison between FAV and UNFAV the most striking difference between 1FAV and 1UNFAV is the large gap in punishment effectiveness. As can be seen in Table 3, punishees in the 1FAV treatment increase their subsequent contribution by about 5 tokens. In 1UNFAV this effect is significantly smaller. This difference is robust to controlling for counter-punishment (columns 2-4) and the contribution level of one's group peers (column 4).

Table 3: Reaction to receiving punishment in 1FAV and 1UNFAV

DV: Change in Contribution (t)	(1)	(2)	(3)	(4)
1UNFAV	-0.412 (0.254)	-0.412 (0.254)	-0.412 (0.254)	0.159 (0.208)
Received punishment in (t-1)	5.358*** (0.411)	5.313*** (0.514)	5.869*** (0.627)	5.971*** (0.642)
Received punishment in (t-1) x 1UNFAV	-2.276** (0.851)	-2.279** (0.862)	-3.161** (1.224)	-3.209** (1.222)
Counter-punished in (t-1)		0.115 (0.939)	-1.315 (1.471)	-1.575 (1.456)
Counter-punished in (t-1) x 1UNFAV			2.219 (1.803)	2.239 (1.749)
Others' Contribution in (t-1)				0.117*** (0.0164)
Constant	-0.483** (0.151)	-0.483** (0.151)	-0.483** (0.151)	-2.378*** (0.274)
# observations	1,296	1,296	1,296	1,296
# subjects	144	144	144	144
# clusters	36	36	36	36
chi2	187.5	186.1	233.4	228.0
R squared (overall)	0.113	0.113	0.115	0.132

Robust standard errors, clustered by groups, in parentheses. *** p<0.001, ** p<0.01, * p<0.05. Random Effects Estimator.

B. No player informed

The previous sections have shown that it makes a considerable difference whether people enter a complex social dilemma with favorable or unfavorable pre-play expectations. The 1FAV and 1UNFAV treatments suggest that even if only a small fraction of participants is informed, totally different norms of cooperation may emerge. However, to better assess the magnitude of the effect, we need to compare the behavior of (favorably and unfavorably) informed subjects with that of uninformed ones. For that purpose, we ran another treatment with 68 new subjects: NoInfo. In this treatment, subjects play exactly the same game as in the previous four treatments but this time none of the players receives any pre-play information.

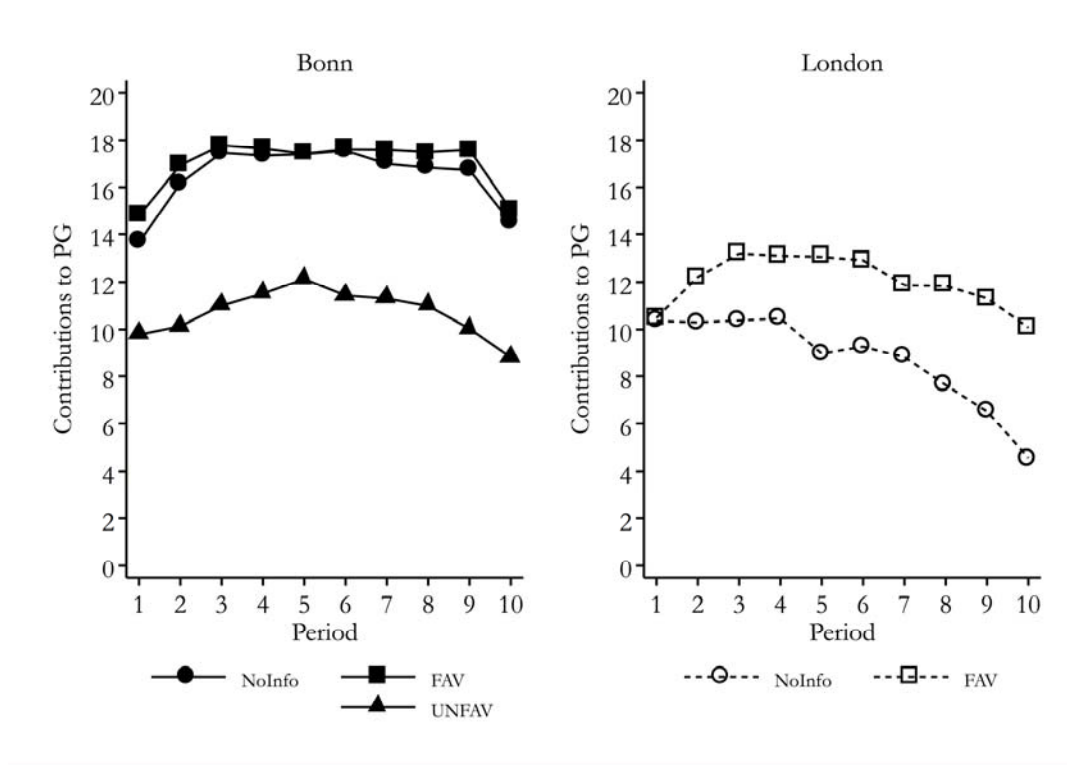


Figure 6: Public Good contributions in Bonn and London

Notes: Left panel: NoInfo Bonn has 68 subjects (17 groups), FAV Bonn has 72 subjects (18 groups), UNFAV Bonn has 60 subjects (15 groups), Right panel: NoInfo London (Nikiforakis, 2008) has 48 subjects (12 groups), FAV London has 64 subjects (16 groups).

As the left panel of Figure 6 shows, letting people play the game without pre-play information is virtually equivalent to giving them favorable information (Mann-Whitney ranksum test, two-sided, $N=35$, $p=0.2411$). This suggests that people’s home-grown expectations of one another and their resulting cooperativeness are rather high. But at the same time, the substantial drop of cooperation after seeing the unfavorable examples (Mann-Whitney ranksum test, two-sided, $N=32$, $p=0.0024$) suggests that those favorable home-grown expectations are very fragile. In fact, as we saw above it is sufficient to give just one player unfavorable information to provoke the drop of cooperativeness (Mann-Whitney ranksum test, two-sided, $N=35$, $p=0.0008$).

A plausible objection to the observed asymmetry between the effect of FAV and UNFAV with respect to NoInfo would be that FAV simply cannot improve cooperation upon NoInfo since NoInfo already displays very high levels of cooperation. The asymmetry would thus be simply due to a ceiling effect. To test this, we ran yet another additional experimental treatment. From Nikiforakis (2008) we know that a treatment identical to our NoInfo displayed much lower levels of cooperation at the Royal Holloway Lab in London, UK. Whereas our NoInfo treatment at Bonn had mean contributions of 16.46 tokens, subjects in London contributed on average merely 8.72. We thus conjecture that the FAV treatment should have a rather good shot at enhancing cooperation in London. Our additional FAV sessions in London comprise a total of 64 participants.

The effect of favorable information on contribution levels in London is shown in the right panel of Figure 6. Also in an environment with much lower home-grown cooperativeness the positive information is not able to increase initial contributions to the public good. In fact, mean first period cooperation in FAV London (10.48) and NoInfo London (10.35) are virtually identical (Mann-Whitney ranksum test, two-sided, $N=112$, $p=0.9881$). Over time, however, contributions in NoInfo London decrease even further whilst they remain rather constant in FAV London. Average contributions increase from 8.72 in NoInfo London to 12.00 in FAV London. This increase of almost 40% is remarkable, though it falls short of being statistically significant (Mann-Whitney ranksum test, two-sided, $p=.1567$). Still, the behavioral pattern seems to be strongly influenced by the favorable manipulation. While cooperation is sustained in FAV London, contribution rates quickly decline in NoInfo London. In FAV London, contributions rise sharply for the first three periods and stay more or less stable until period six before displaying a slight decay at the end. In contrast, in NoInfo London contributions never rise but rather decay from the fourth period on. This difference of time trends is highly significant (Mann-Whitney ranksum test over means of first differences, two-sided, $N = 28$, $p = .0243$).

6. Discussion

In this paper, we have explored how cooperative behavior reacts to selective pre-play information about other, unrelated groups. To this end, we have used an experimental framework that captures cooperation situations that are sufficiently rich, so that several potentially conflicting behavioral norms might emerge. We find that the aggregate level of cooperativeness dramatically depends on pre-play information. Additional treatments show that the initial change in behavior can be attributed to a change of initial beliefs. Over time, the difference between groups that had been shown positive examples of behavior and those that had seen negative examples does not narrow but rather widens. On the other hand, behavior within both types of groups becomes increasingly homogeneous over time, indicating the formation of rather different social norms: high cooperation in FAV, low cooperation in UNFAV.

Moreover, if punishment is the consequence of a perceived norm infraction of the punishee (as perceived by the punisher) and the punishee's reaction to being punished reflects her (dis)agreement with the punisher's normative judgment, our findings suggest two interesting things. On the one hand, the fact that punishment is substantially less effective in UNFAV than in FAV indicates that for some people pre-play information not only alters behavioral expectations ("how will people behave") but also normative expectations ("how should one rightfully behave"). Apparently, in UNFAV many defectors not only think that it is common to contribute low but also legitimate. On the other hand, the vast majority of punishment is pro-social, both in FAV and in UNFAV. Apparently thus, many players in UNFAV do not think that contributing low is legitimate, notwithstanding the negative examples of others' behavior seen at the beginning of the experiment.

This points at an interesting source of heterogeneity that calls for further exploration in the future. People seem to differ in the degree to which they condition their normative expectations on how other people behave. On one extreme of the spectrum, one's normative expectation would be derived from a social ideal and not change in the face of deviating behavior. On the other end of the spectrum, a person's notion of how one should rightfully behave would depend solely on what other people do. We might call the former pure moralism and the latter pure conformism. Most people will likely be somewhere in between. Further research should investigate the predictive power of a type classification based on the conditionality of people's normative expectations as well as the relationship between such a classification and the concept of conditional cooperation, which is purely based on behavioral expectations.

Our findings have a number of additional implications. First, they clearly point to the relevance of pre-game communication – a factor that has only lately started to receive significant attention in the literature. While the existing literature usually focuses on self-chosen cheap talk messages (for an overview, see Crawford 1998), we demonstrate that also exogenously selected, one-way information about other players can alter how players act in subsequent games. In particular the findings of our belief treatments might be of interest to this literature. They suggest a possible channel through which the observed effects of cheap talk are mediated, namely through the alteration of subjects' pre-game expectations.

Taken together, our results underline the power and importance of information and experience in shaping cooperative behavior. The bottom line is that observation matters. Interestingly, people do not only learn from what they experience themselves. They also seem to learn “vicariously”, by observing others, or by seeing the results (Bandura 1977). The effect is even present when participants are told that the information they are receiving is selective. Most strikingly, the substantial difference in behavior is even achieved when only one person out of four receives the pre-play information (1FAV and 1UNFAV). This connects our study with the growing literature on social learning (for a recent meta-study see Weizsäcker 2010). Our subjects' behavior critically depends on pre-game expectations, which we show to be easy to deteriorate in a complex setting – simply by providing subjects the opportunity of vicarious learning.

This suggests that, in appropriate circumstances, impression management might indeed be a feasible tool to avert, or at least to mitigate, the danger of social dilemmas. Of course this is a paternalistic intervention. But note that conditional cooperators need not even be deceived for the intervention to be successful. All that is required is that they expect the manipulation to matter for a sufficient fraction of the remaining members of their group, be that because they are deceived, or because they are skeptical themselves, but willing to give cooperation a try since the intervention gives all of them one and the same informational starting point.

For policy makers these findings represent both a chance and a peril. If they do not manage to repair broken windows quickly, both literally and metaphorically speaking, chances are that a vicious cycle starts. By contrast, if they can induce some to lead others by their socially beneficial example, this strategy may well work. In particular, policy makers would want to prevent

(perhaps wrong) pessimistic beliefs from spreading. In any case, we show that home-grown expectations must not be disregarded in order to attain socially desirable outcomes. Consequently in situations where the success of a law depends on the willingness of individual citizens to cooperate – for instance in areas like waste separation and sustainable water use – government might want to consider a PR campaign in order to create a general atmosphere of cooperativeness within the population.

Managing first impressions might certainly be less effective if the large majority of addressees know better. The intervention requires a sufficient degree of uncertainty. Yet in political reality, quite a few public goods are characterized by deep conceptual and factual uncertainty. Problems like climate change are heavily contested among scientists and not well understood by many. Addressees have to trust expertise. If in the eyes of addressees the underlying social problem is opaque, they are also likely to be uncertain how others will react to it. Most importantly, addressees face behavioral uncertainty whenever they newly enter a community. They do not know local mores, nor do they know how determined the group is when it comes to enforcing them.

The results in this paper should, of course, not be taken as arguments against the importance of elaborated, incentives-altering mechanisms in general. For instance, the effectiveness of our very simple mechanism seems to be asymmetric, since it has a hard time to increase cooperation in an otherwise uncooperative environment. Still, our findings suggest that minimal interventions *can* have a strong behavioral effect. Future research could try to explore the interaction between such simple and other, more complex and intrusive mechanisms.

References

- Akerlof, George A. and Rachel E. Kranton.** 2000. "Economics and Identity." *The Quarterly Journal of Economics*, 115(3), 715-53.
- Andreoni, James.** 1988. "Why Free Ride? Strategies and Learning in Public Goods Experiments." *Journal of Public Economics*, 37, 291-304.
- Andreoni, James and Ragan Petrie.** 2004. "Public Goods Experiments without Confidentiality. A Glimpse into Fund-Raising." *Journal of Public Economics*, 88, 1605-23.
- Bandura, Albert.** 1977. *Social Learning Theory*. Englewood Cliffs: Prentice Hall.
- Bénabou, Roland and Jean Tirole.** 2011. "Identity, Morals, and Taboos: Beliefs as Assets." *The Quarterly Journal of Economics*, 126(2), 805-55.
- Berg, Joyce, John Dickhaut and Kevin McCabe.** 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior*, 10, 122-42.
- Bernheim, B. Douglas.** 1994. "A Theory of Conformity." *Journal of Political Economy*, 102(5), 841-77.
- Bicchieri, Cristina.** 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge UK: Cambridge University Press.
- Bicchieri, Cristina and Erte Xiao.** 2009. "Do the Right Thing: But Only If Others Do So." *Journal of Behavioral Decision Making*, 22, 191-208.
- Blume, Andreas and Andreas Ortmann.** 2007. "The Effects of Costless Pre-Play Communication. Experimental Evidence from Games with Pareto-Ranked Equilibria." *Journal of Economic Theory*, 132, 274-90.
- Bohnet, Iris and Richard Zeckhauser.** 2004. "Social Comparisons in Ultimatum Bargaining." *Scandinavian Journal of Economics*, 106, 495-510.
- Braga, Anthony A., David L. Weisburd, Elin J. Waring, Lorraine Green Mazerolle, William Spelman and Francis Gajewski.** 1999. "Problem-Oriented Policing in Violent Crime Places. A Randomized Controlled Experiment." *Criminology*, 37, 541-80.
- Brekke, Kjell Arne, Karen Evelyn Hauge, Jo Thori Lind and Karine Nyborg.** 2009. "Playing with the Good Guys: A Public Good Game with Endogenous Group Formation,"
- Burlando, Roberto M. and Francesco Guala.** 2005. "Heterogeneous Agents in Public Goods Experiments." *Experimental Economics*, 8, 35-54.

- Chen, Yan, F. Maxwell Harper, Joseph Konstan and Sherry Xin Li.** 2010. "Social Comparisons and Contributions to Online Communities. A Field Experiment on Movielen." *American Economic Review*, 100(4), 1358-98.
- Cialdini, Robert B., Raymond R. Reno and Carl A. Kallgren.** 1990. "A Focus Theory of Normative Conduct. Recycling the Concept of Norms to Reduce Littering in Public Places." *Journal of Personality and Social Psychology*, 58, 1015-26.
- Cinyabuguma, Matthias, Talbot Page and Louis Putterman.** 2005. "Cooperation under the Threat of Expulsion in a Public Goods Experiment." *Journal of Public Economics*, 89, 1421-35.
- Corazzini, Lucca, Sebastian Kube, Michel André Maréchal and Antonio Nicolo.** 2010. "Elections and Deceptions. Theory and Experimental Evidence,"
- Crawford, Vincent.** 1998. "A Survey of Experiments on Communication Via Cheap Talk." *Journal of Economic Theory*, 78(2), 286-98.
- Croson, Rachel T.A., Enrique Fatas and Tibor Neugebauer.** 2008. "The Effect of Excludability on Team Production,"
- Dal Bó, Ernesto and Pedro Dal Bó.** 2009. "" Do the Right Thing." The Effects of Moral Suasion on Cooperation," National Bureau of Economic Research,
- Dufwenberg, Martin and Georg Kirchsteiger.** 2004. "A Theory of Sequential Reciprocity." *Games and Economic Behavior*, 47, 268-98.
- Engel, Christoph, Martin Beckenkamp, Andreas Glöckner, Bernd Irlenbusch, Heike Hennig-Schmidt, Sebastian Kube, Michael Kurschilgen, Alexander Morell, Andreas Nicklisch, Hans-Theo Normann, et al.** 2014. "First Impressions Are More Important Than Early Intervention: Qualifying Broken Windows Theory in the Lab." *International Review of Law and Economics*, 37, 126-36.
- Falk, Armin and Urs Fischbacher.** 2006. "A Theory of Reciprocity." *Games and Economic Behavior*, 54, 293-315.
- Falkinger, Josef, Ernst Fehr, Simon Gächter and Rudolf Winter-Ebmer.** 2000. "A Simple Mechanism for the Efficient Provision of Public Goods. Experimental Evidence." *American Economic Review*, 90, 247-64.
- Fehr, Ernst and Simon Gächter.** 2002. "Altruistic Punishment in Humans." *Nature*, 415, 137-40.
- _____. 2000a. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 90(4), 980-94.

- _____. 2000b. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 90, 980-94.
- Fehr, Ernst and Bettina Rockenbach.** 2003. "Detrimental Effects of Sanctions on Human Altruism." *Nature*, 422, 137-40.
- Fischbacher, Urs.** 2007. "Z-Tree. Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics*, 10, 171-78.
- Fischbacher, Urs and Simon Gächter.** 2010. "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments." *American Economic Review*, 100(1), 541-56.
- Fischbacher, Urs, Simon Gächter and Ernst Fehr.** 2001. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters*, 71, 397-404.
- Forsythe, Robert, Roger B. Myerson, Thomas Rietz and Robert J. Weber.** 1993. "An Experiment on Coordination in Multi-Candidate Elections. The Importance of Polls and Election Histories." *Social Choice & Welfare*, 10, 223-47.
- Forsythe, Robert, Thomas Rietz, Roger B. Myerson and Robert J. Weber.** 1996. "An Experimental Study of Voting Rules and Polls in Three-Candidate Elections." *International Journal of Game Theory*, 25, 355-83.
- Frey, Bruno and Stephan Meier.** 2004. "Social Comparisons and Pro-Social Behavior: Testing "Conditional Cooperation" in a Field Experiment." *American Economic Review*, 94, 1717-22.
- Gächter, Simon and Christian Thöni.** 2007. "Social Learning and Voluntary Cooperation among Like-Minded People." *Journal of the European Economic Association*, 3, 303-14.
- Galbiati, Roberto and Pietro Vertova.** 2008. "Law and Behaviours in Social Dilemmas. Testing the Effect of Obligations on Commitment." *Games and Economic Behavior*, 64, 146-70.
- Glöckner, Andreas, Bernd Irlenbusch, Andreas Nicklisch and Hans-Theo Normann.** 2010. "Leading with(out) Sacrifice? A Public-Goods Experiment with a Privileged Player." *Economic Inquiry*, ***, ***.
- Goerg, Sebastian J. and Gary Walkowitz.** 2010. "On the Prevalence of Framing Effects across Subject-Pools in a Two-Person Cooperation Game." *Journal of Economic Psychology*, ***, ***.
- Greiner, Ben.** 2004. "An Online Recruiting System for Economic Experiments," K. Kremer and V. Macho, *Forschung Und Wissenschaftliches Rechnen 2003*. Göttingen: 79-93.

- Gunnthorsdotir, Anna, Daniel Houser and Kevin McCabe.** 2007. "Disposition, History and Contributions in Public Goods Experiments." *Journal of Economic Behavior & Organization*, 62, 304-15.
- Güerker, Özgür, Bernd Irlenbusch and Bettina Rockenbach.** 2006. "The Competitive Advantage of Sanctioning Institutions." *Science*, 312, 108-11.
- Herrmann, Benedikt, Christian Thöni and Simon Gächter.** 2008. "Antisocial Punishment across Societies." *Science*, 319, 1362-67.
- Isaac, R. Mark, James M. Walker and Arlington W. Williams.** 1994. "Group Size and the Voluntary Provision of Public Goods. Experimental Evidence Utilizing Large Groups." *Journal of Public Economics*, 54, 1-36.
- Keizer, Kees, Siegwart Lindenberg and Linda Steg.** 2008. "The Spreading of Disorder," *Science*. 1681-85.
- Klor, Esteban and Eyal Winter.** 2007. "The Welfare Effects of Public Opinion Polls." *International Journal of Game Theory*, 35, 379-94.
- Kogan, Shimon, Anthony M. Kwasnica and Roberto A. Weber.** 2010. "Coordination in the Presence of Asset Markets." *American Economic Review*, ***, ***.
- Krupka, Erin L. and Roberto A. Weber.** 2009. "The Focusing and Informational Effects of Norms on Pro-Social Behavior." *Journal of Economic Psychology*, 30, 307-20.
- Kube, Sebastian and Christian Traxler.** 2010. "The Interaction of Legal and Social Norm Enforcement." *Journal of Public Economic Theory*, ***, ***.
- Mengel, Friedrike.** 2007. "The Evolution of Function-Value Traits for Conditional Cooperation." *Journal of Theoretical Biology*, 245, 564-75.
- Nikiforakis, Nikos.** 2008. "Punishment and Counter-Punishment in Public Good Games: Can We Really Govern Ourselves?" *Journal of Public Economics*, 92, 91-112.
- Ostrom, Elinor, James M. Walker and Roy Gardner.** 1992. "Covenants with and without Sword. Self-Governance Is Possible." *American Political Science Review*, 40, 309-17.
- Page, Talbot, Louis Putterman and Bulent Unel.** 2005. "Voluntary Association in Public Goods Experiments. Reciprocity, Mimicry and Efficiency." *Economic Journal*, 115, 1032-53.
- Ramos, Joao and Benno Torgler.** 2010. "Are Academics Messy? Testing the Broken Windows Theory with a Field Experiment in the Work Environment,"

- Shang, Jen and Rachel T.A. Croson.** 2009. "A Field Experiment in Charitable Contribution. The Impact of Social Information on the Voluntary Provision of Public Goods." *Economic Journal*, 119, 1422-39.
- Weber, Roberto A.** 2006. "Managing Growth to Achieve Efficient Coordination in Large Groups." *American Economic Review*, 96, 114-26.
- Weizsäcker, Georg.** 2010. "Do We Follow Others When We Should? A Simple Test of Rational Expectations." *American Economic Review*, ***.
- Wilson, James Q. and George L. Kelling.** 1982. "Police and Neighborhood Safety. Broken Windows." *Atlantic Monthly*, 127, 29-38.
- Zimbardo, Philip.** 1969. "The Human Choice. Individuation, Reason, and Order Versus Deindividuation, Impulse, and Chaos." *Nebraska Symposium on Motivation*, 17, 237-307.
- Zimbardo, Philip G. and Ebbe B. Ebbesen.** 1969. *Influencing Attitudes and Changing Behavior. A Basic Introduction to Relevant Methodology, Theory, and Applications.* Reading, Mass.,: Addison-Wesley Pub. Co.

Appendix 1: Experimental Instructions

You are now taking part in an economic experiment. If you read the following instructions carefully, you can, depending on your decisions, earn a considerable amount of money. It is therefore important that you take your time to understand the instructions.

The instructions which we have distributed to you are for your private information. **Please do not communicate with the other participants during the experiment.** Should you have any questions please ask us.

During the experiment we shall not speak of Pounds, but of Experimental Currency Units (ECU). Your entire earnings will be calculated in ECUs. At the end of the experiment the total amount of ECUs you have earned will be converted to Pounds at the rate of **1 ECU = 4 p** and will be immediately paid to you in cash.

At the beginning of the experiment the participants will be randomly divided into groups of four. You will therefore be in a group with 3 other participants. **The composition of each group will remain the same throughout the experiment.** The experiment lasts 10 periods and each period is divided into 3 stages.

The 1st stage:

At the beginning of each of the 10 periods each participant will receive 20 ECUs. In the following, we shall refer to this amount as the “endowment”. In the 1st stage, your task is to decide how to use your endowment. You have to decide how many of the 20 ECUs you want to contribute to a project (from 0 to 20) and how many of them to keep for yourself. The consequences of your decision are explained in detail below.

Once all the players have decided their contribution to the project you will be informed about, the group’s total contribution, your income from the project and your payoff in this period. Your payoff in each period is calculated using the following simple formula. Again, if you have any difficulties do not hesitate to ask us.

Income from the 1st stage	=	Endowment of ECUs	–	Your contribution to the Project	+	0.4*Total contribution to the Project
---	---	-------------------	---	----------------------------------	---	---------------------------------------

This formula shows that your 1st stage income consists of two parts:

- 1) The ECUs which you have kept for yourself (endowment – contribution)
- 2) The income from the project, which equals to the 40% of the group’s total contribution.

The income of each group member from the project is calculated in the same way. This means that each group member receives the same income from the project. Suppose the sum of the contributions of all group members are 60 ECUs. In this case, each member of the group receives an income from the project of: $0.4*60=24$ ECUs. If the total contribution to the project is 9 points, then each member of the group receives an income of: $0.4*9=3.6$ ECUs from the project.

You always have the option of keeping the ECUs for yourself or contributing them to the project. Each ECU that you keep raises your end of period income by 1 ECU. Supposing you contributed this point to the project instead, then the total contribution to the project would rise by 1 ECUs. Your income from the project would thus rise by $0.4*1=0.4$ ECUs. However, the income of the other group members would also rise by 0.4 ECUs each, so that the total income of the group from the project would be 1.6 points. Your contribution to the project therefore also raises the income of the other group members. On the other hand you also earn an income for each point contributed by the other members to the project. In particular, for each point contributed by any member you earn 0.4 ECUs.

In addition to the 20 ECUs per period, each participant receives a one-off lump sum payment of 25 ECUs at the beginning of this part. This one-off payment can be used to pay for eventual losses during the experiment. **However, you can always evade losses with certainty through your own decisions.** Note that this lump sum payment will not be used to calculate the income from the period. It will only be added to your total income from all the periods at the very end.

The 2nd stage:

At the 2nd stage you will be informed how much each group member contributed individually to the project at the 1st stage. At this stage you can **reduce or leave equal** the income of **each** member of your group by **distributing points**. The other group members can also reduce your income if they wish to.

If you choose 0 points for a particular group member, you do not change his or her income. However if you give a member 1 point you reduce his or her income by 10 percent. If you give a member 2 points you reduce his or her income by 20 percent, etc. The amount of points you distribute to each member determines, therefore, how much you reduce their income from the 1st stage. If one player receives in total 4 points his income will be reduced by 40% and if he receives **10 or more** his income from the 1st stage will be reduced by 100%.

If you distribute points you have costs in ECUs, which depend on the amount of points you distribute. You can distribute between 0 and 10 points to each group member. The more points you give **to any** group member, the higher your costs. Your total costs are equal to the **sum of the costs of distributing points to each of the other three group members**. The following table illustrates the relation between distributed points to **each** group member and

Points	0	1	2	3	4	5	6	7	8	9	10
Cost of points per person	0	1	2	4	6	9	12	16	20	25	30

the cost of doing so in ECUs.

Example: Supposing you give 2 points to player 1 this costs you 2 ECUs; if you also give 8 points to player 3 this costs you a further 20 ECUs; and if you give 0 points to the last group member this has no additional cost for you. In this case, your total costs of distributing points would be 22 ECUs (2+20+0) and not 30 ECUs.

The following equation summarizes the previous information. Your total income from the two stages is calculated as follows:

$$\text{Income at the end of the 2nd stage} = \text{Income from the 1st stage} * [(10 - \text{received points})/10] - \text{Costs of distributed points}$$

Please note that your income in ECUs at the end of the 2nd and the 3rd stage can be negative, if the costs of your points distributed exceeds your (possibly reduced) income from the 1st stage. **You can however evade such losses with certainty through your own decisions.** Should your income become zero or negative at the end of the 2nd stage you will not be able to continue to the 3rd stage. If your income becomes zero or negative at the end of the 3rd stage you can simply use your 25 ECUs that we gave you in the beginning in order to pay this off.

The 3rd stage:

In the 3rd and final stage, after being informed of the points that the other group members assigned to you, you will be given one last opportunity of assigning points back to the other participants, thus reducing their income. We shall call these points “counter-points”. **You will only be able to assign counter-points to participants who assigned points to you during the 2nd stage**

The costs of assigning points, as well as the income reduction caused by each point remain the same as before. The following table shows you how to calculate the costs for assigning counter-points.

Costs for assigning counter-points to <u>one specific group member</u> in the 3 rd stage											
Points that you already assigned to <u>one specific group member</u> in the 2 nd stage	Counter-points you assign to that <u>same group member</u> in the 3 rd stage										
	0	1	2	3	4	5	6	7	8	9	10
0	0	1	2	4	6	9	12	16	20	25	30
1	0	1	3	5	8	11	15	19	24	29	
2	0	2	4	7	10	14	18	23	28		
3	0	2	5	8	12	16	21	26			
4	0	3	6	10	14	19	24				
5	0	3	7	11	16	21					
6	0	4	8	13	18						
7	0	4	9	14							
8	0	5	10								
9	0	5									
10	0										

Example: if you distribute 2 points in the 2nd stage to player 1 you have a cost of 2 ECUs. If in the 3rd stage you decide to distribute 3 counter-points to player 1, a further 7 ECUs are added to your cost.

Your income after the 3rd stage (= period income) is therefore calculated as follows:

Income at the end of the 3rd stage	=	Income from the 2 nd stage	*	[(10 – received counter-points)/10]	–	Cost of distributed counter-points
--	---	---------------------------------------	---	-------------------------------------	---	------------------------------------

If you have any further questions please raise your hand and one of the supervisor.

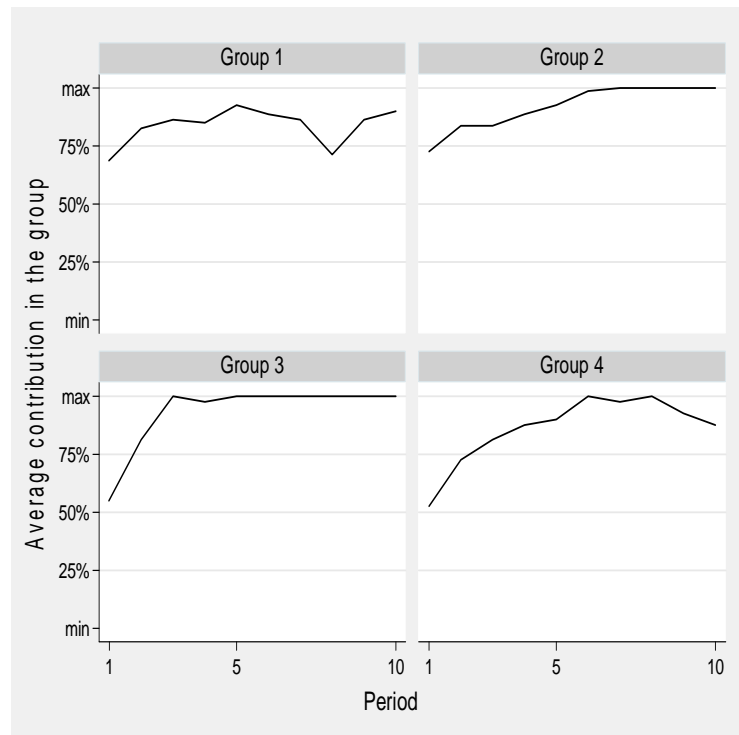
Appendix 2: Control Questionnaire

1. Each group member has an endowment of 20 ECUs. Nobody (including you) contributes any ECUs to the project. What is:
 - a. Your income at the end of the first stage?
 - b. The income of the other group members?.....
2. Each group member has an endowment of 20 ECUs. You contribute 20 ECUs to the project. All other group members contribute 20 ECUs each to the project. What is:
 - a. Your income at the end of the first stage?
 - b. The income of the other group members?.....
3. Each group member has an endowment of 20 ECUs. The other three group members contribute together a total of 30 ECUs to the project. What is:
 - a. Your income at the end of the first stage if you contribute 0 ECUs to the project?
 - b. Your income at the end of the first stage if you contribute 15 ECUs to the project?
4. Each group member has an endowment of 20 ECUs. You contribute 8 ECUs to the project. What is:
 - a. Your income at the end of the first stage if the other group members together contribute a further total of 7 ECUs to the project?.....
 - b. Your income at the end of the first stage if the other group members together contribute a further total of 22 ECUs to the project?.....
5. At the second stage you distribute the following points to your three other group members: 9, 5, 0. What are the total costs of your distributed points?....
6. What are your costs if you distribute 0 points?
7. By how many percent will your income from the first stage be reduced when you receive from the other group members a total of:
 - a. 0 points? ...
 - b. 4 points? ...
 - c. 15 points? ...
8. At the second stage you distribute the following points to your three other group members: 2, 2, 0. In the third stage you distribute the following points to your three other group members: 1, 1, 1. What are the total costs of your distributed points?....
9. By how many per cent is your second stage income reduced, if you have received the sum of 3 counter-points from the other group members in stage 3?

Appendix 3: Additional Sheet FAV Treatment

Additional information for today's experiment

This experiment has already been run at two laboratories in Bonn (Germany) and London (UK). The following figure shows you the contribution behaviour in four selected groups from both locations.



In these selected graphs you see how much the 4 group members contributed on average to the group project. In the selected groups 1 and 2 the contributions are high right from the beginning. In the selected groups 3 and 4, average contribution starts somewhat lower but rises over the course of the experiment.

Some additional numbers from the experiments in Bonn and London:

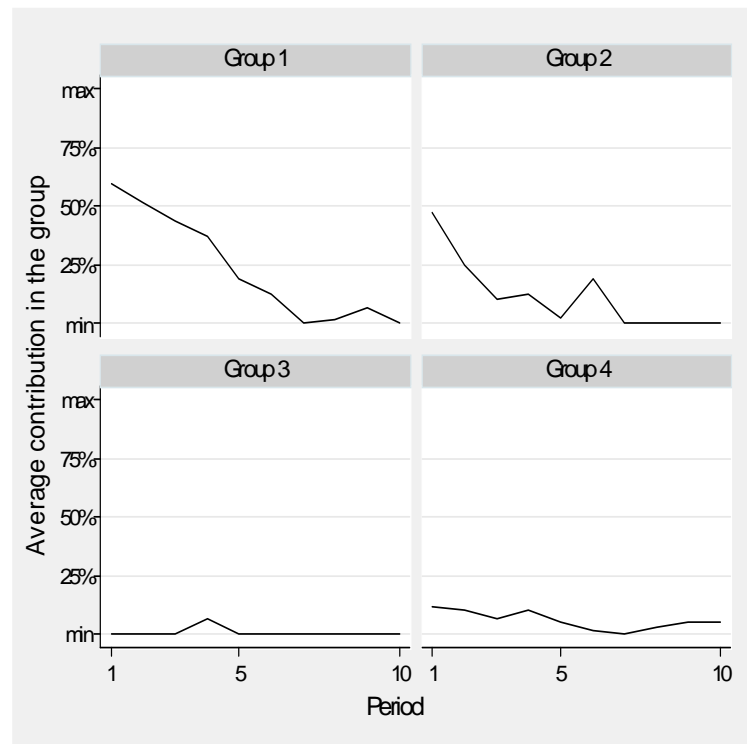
- In the first round of the two previous experiments, 58% to 74% of the contributions were between 10 and 20.
- In the last round, up to 53% of the contributions were equal to 20.
- Players, whose income was reduced in stage 2, only used counter-points in stage 3 in 29% to 41% of all possible cases.

All these graphs and numbers are meant to give you some orientation concerning the type of situation you might be in during the experiment and what you might possibly expect from the other members of your group.

Appendix 4: Additional Sheet UNFAV Treatment

Additional information for today's experiment

This experiment has already been run at two laboratories in Bonn (Germany) and London (UK). The following figure shows you the contribution behaviour in four selected groups from both locations.



In these selected graphs you see how much the 4 group members contributed on average to the group project. In the selected groups 1 and 2 the contributions are rather low and display a decrease over the course of the experiment. In the selected groups 3 and 4, average contributions are low right from the beginning

Some additional numbers from the experiments in Bonn and London:

- In the first round of the two previous experiments, 37% to 55% of the contributions were between 0 and 10.
- In the last round, up to 52% of the contributions were equal to 0.
- Players, whose income was reduced in stage 2, only used counter-points in stage 3 to reduce the income of those who had given them points in stage 2. Every time they used counter-points they reduced the income of those who had given them points in stage 2 by 18% to 25%.

All these graphs and numbers are meant to give you some orientation concerning the type of situation you might be in during the experiment and what you might possibly expect from the other members of your group.

Appendix 5: Belief Elicitation

As noted in the main text, subjects in the belief elicitation treatments (FAV-B, UNFAV-B, 1FAV-B, 1UNFAV-B) had to read the same instructions and answer the same control questions as the participants in the contribution treatments (FAV, UNFAV, 1FAV, 1UNFAV). However, once ztree started subjects were told the following on their screens:

You have just read the instructions and answered the control questions. This game has already been played in other laboratories. The participants of those experiments read the same instructions as you and answered the same control questions. You are not going to play this game today. Instead, your task is to answer the following question:

What do you think, how much did those previous participants contribute on average in the first round to the project?

Your guess.....

We have rounded the average down to the next whole number, i.e. the correct answer is a whole number between 0 and 20. If your guess is exactly correct, you will receive 4 Euros. If your guess is wrong, you will receive 0 Euros.

Preprints 2011

2011/04: Petersen N., The Role of Consent and Uncertainty in the Formation of Customary International Law

forthcoming in: Reexamining Customary International Law, Cambridge, Cambridge University Press, In Press.

2011/03: Petersen N., Antitrust Law and the Promotion of Democracy and Economic Growth

2011/02: Engel C., Besonderes Verwaltungsrecht und ökonomische Theorie

2011/01: Engel C., Zhurakhovska L., Oligopoly as a Socially Embedded Dilemma. An Experiment